



**HAL**  
open science

# Signature géographique et météorologique des variations spatiales et temporelles de la température de l'air au sein d'une zone urbaine

Jérémy Bernard

► **To cite this version:**

Jérémy Bernard. Signature géographique et météorologique des variations spatiales et temporelles de la température de l'air au sein d'une zone urbaine. Génie civil. Ecole Centrale de Nantes (ECN), 2017. Français. NNT: . tel-01449935v1

**HAL Id: tel-01449935**

**<https://hal.science/tel-01449935v1>**

Submitted on 14 Feb 2017 (v1), last revised 20 Mar 2020 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Thèse de Doctorat

**Jérémy BERNARD**

*Mémoire présenté en vue de l'obtention du  
**grade de Docteur de l'École centrale de Nantes**  
sous le label de l'Université de Nantes Angers Le Mans*

**École doctorale : Sciences Pour l'Ingénieur, Géosciences, Architecture (SPIGA)**

**Discipline : Génie civil**

**Unité de recherche : Ambiances Architectures Urbanités (AAU)**

**Institut de Recherche en Sciences et Techniques de la Ville (IRSTV)**

**Soutenue le 11 janvier 2017**

## **Signature géographique et météorologique des variations spatiales et temporelles de la température de l'air au sein d'une zone urbaine**

### **JURY**

Rapporteurs : **M. Valéry MASSON**, Ingénieur en chef des ponts et chaussées, Météo France / CNRM-GAME / GMME / TURBAU, Toulouse, France

**M<sup>me</sup> Sandrine ANQUETIN**, Directrice de recherche CNRS, Laboratoire LTHE - CNRS UMR 5564, Grenoble, France

Examineurs : **M<sup>me</sup> Christiane WEBER**, Directrice de recherche, UMR Tetis, Montpellier, France

**M. Pétrissans MATHIEU**, Professeur des universités, Université de Lorraine, Épinal, France

**M. Frédéric KUZNIK**, Professeur des universités, INSA de Lyon, Lyon, France

Invitée : **M<sup>me</sup> Céline PHILLIPS**, Ingénieur, ADEME, Valbonne, France

Directrice de thèse : **M<sup>me</sup> Marjorie MUSY**, Directrice de recherche, CEREMA, Nantes, France

Co-directrice de thèse : **M<sup>me</sup> Isabelle CALMET**, Maître de conférence, Laboratoire LHEEA - ECN, Nantes, France

Co-encadrant de thèse : **M. Erwan BOCHER**, Ingénieur de recherche, Laboratoire Lab-STICC - CNRS UMR 6285, Vannes, France



# Remerciements

Certaines données utilisées dans cette thèse ont été utilisées gratuitement. En contre-partie, j'ai le devoir de remercier les fournisseurs de ces données de la manière suivante : *Ce travail a bénéficié d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre du Programme Investissements d'Avenir pour le projet EQUIPEX GEOSUD portant la référence ANR-10-EQPX-20*. Je souhaite également remercier l'ensemble des personnes qui ont contribué à l'acquisition des données climatiques de l'ONEVU et à leur mise à disposition dans le cadre de ma thèse. Merci également aux différents services des collectivités de Nantes, Saint-Herblain, Angers, Beaucouzé et La Roche-sur-Yon d'avoir contribué à l'acquisition de nouvelles données de température et d'humidité de l'air sur leur territoire. Je remercie également l'ADEME et la région Pays de la Loire, qui ont financé mon travail de thèse.

Hemingway a écrit dans son livre *En avoir ou pas* quelques mots que j'aime beaucoup : « *de quelque façon qu'il s'y prenne, un homme seul est foutu d'avance* ». Je trouve que c'est une belle introduction pour remercier les personnes qui ont partagé ces trois années avec moi. Ces personnes avec qui j'ai passé de bons moments, que j'ai parfois dû supporter, mais qui ont dû me supporter également... La liste des personnes citées est loin d'être exhaustive. Aussi, je m'excuse d'avance à toutes les personnes qui ne trouveront pas leur nom dans ces remerciements.

Je voudrais tout d'abord remercier tous les membres du jury d'avoir participé à ma soutenance de thèse, et notamment les deux rapporteurs de mon manuscrit, Valéry Masson et Sandrine Anquetin. À mes encadrants, Erwan, Isabelle et Marjorie, merci beaucoup d'avoir accepté d'encadrer mon travail et d'avoir été si complémentaires pendant ces trois années de thèse : Marjorie pour toujours avoir répondu rapidement et pertinemment à mes interrogations, Isabelle pour les relectures attentives de mon manuscrit, Erwan pour m'avoir aidé à mieux définir ma problématique et pour m'avoir orienter vers l'utilisation des logiciels libres. Dans le même esprit, merci à l'ancienne équipe SIG pour leur aide précieuse lorsque j'avais besoin de support sous Linux ou sur la plate-forme OrbisGis : merci à Gwendall, Nicolas, Sylvain, Adam

et Gwenaël.

J'ai également eu la chance de partager quotidiennement les bureaux de l'IRSTV avec toute une bande de joyeux lurons. Auline, Benjamin, ne changez rien : bonne humeur ou humeur massacrant, pour des discussions légères ou un peu plus lourdes... Merci à tous les deux pour votre soutien pour les périodes les moins rigolottes de la thèse, mais aussi pour tout le reste. Bahtiyor, *my office mate*, je suis ravi d'avoir partagé mon bureau pendant trois années avec toi. Fabrice, merci pour ta bonne humeur lorsque tu daignais nous rendre visite au nord de la Loire ; Anne, pour ta gentillesse et les échanges que nous avons eu les deux premières années de ma thèse. Enfin, merci à tous les autres : Marie et son porte-voix, Chloé et son vélo magique, MH la reine des stagiaires, Noëlle, Patrice, Jean-Michel, Hervé, et tous ceux que je ne cite pas. Même si j'étais moins présent au CRENAU, je souhaite également remercier l'ensemble des doctorants pour leur accueil lors de mes passages épisodiques au laboratoire, mais aussi Pascal qui a cru en mes talents d'acousticien, Frédéric qui est arrivé tardivement mais qui reste plus qu'un bon compagnon de train.

L'ADEME, via sa cellule thèse et son association des doctorants, m'a permis d'échanger non-seulement avec une quantité et une diversité remarquable de jeunes chercheurs mais également avec des professionnels issus du monde opérationnel. En plus de tous les doctorants de l'ADEME avec qui nous avons partagé de très bons moments lors des colloques, je souhaite sincèrement remercier Nicolas, Maguy et Valérie, qui assurent l'animation de la cellule thèse de façon MA-GI-STRALE. Merci également à Jean-Marc pour le temps qu'il a consacré à l'association 2AD, et à Céline d'avoir soutenu mon travail de thèse du début à la fin.

Dans le cadre du projet URBIO dans lequel s'inscrivait ma thèse, j'ai également eu la chance de travailler avec Pascal et Dominique. Merci à ces deux as de la clef à molette, avec qui j'ai déployé les réseaux de capteurs de température sur les agglomérations de Nantes, d'Angers et de La Roche-sur-Yon. Plus particulièrement à Pascal, merci au temps que tu as pu consacrer aux nombreuses discussions que nous avons eu concernant le traitement des données météorologiques. URBIO n'aurait pas pu fonctionner sans Marie, pierre angulaire du projet. Merci à toi d'avoir animé la communauté et de m'avoir donné l'occasion de travailler avec nos collègues écologues. À ce titre, ma pensée vient tout particulièrement à Arnaud et à Hervé mais également à tous les autres membres du projet.

Et parce que le travail ne représente pas l'essentiel de ma vie, j'aimerais également remercier toutes les personnes présentes autour de moi ces trois dernières années. Tout d'abord les copains de la maison de quartier des Dervallières : Ahmed, Khalid, Marie, Gweltaz, Sa-

mira, Mous et une mention particulière aux collégiens qui ont littéralement réussi à m'épuiser certaines semaines. Une grosse pensée aux copains nantais également : Arnaud, Juju, Antoine, Aline, Christophe et Livia, Marion, Noémie, Robin. Merci au 4 avenue bonnet et surtout à ses habitants (Karim, Lucile, Régis, Caroline, Colette, Tim et Analia, Juliette, Isabelle) et aux trois lascars qui ont squatté à mon appartement et qui m'ont donc tenu compagnie pendant que ma belle s'envolait la semaine : Aymeric, Nicolas et bien sûr FX. Parce que je veux oublier le moins de monde possible, merci aussi aux copains qui n'ont pas daigné vivre avec moi à Nantes pendant ces trois années mais qui me sont cependant indispensables : Adri, Dom, Yaya et Franck, Aurel, Diane et Célian, Alexis et Laura, Coco, Vita, marin, Émilie et Guillaume, le 6 avenue Marx Dormoy (Anaëlle, Hugo, Maï et Yoyo), le 4 de l'Abbé Grégoire (Gygy et Coco) et j'en oublie... Parce que les copains c'est bien, mais que ma famille reste ma famille, je ne veux pas oublier de remercier ma sœur et mes parents pour leurs fréquents séjours dans le grand ouest, mes cousins et en particulier cousin et son frangin, mes grands-parents paternels et maternels, tonton et tata, bopap et fougou et tous les autres. Enfin, parce que c'est avant tout toi qui a dû me supporter pendant les périodes de rédaction les moins joyeuses, merci Sarah d'avoir été si compréhensive et si patiente.



# Table des matières

<b>Nomenclature</b>	<b>13</b>
<b>Liste des tableaux</b>	<b>23</b>
<b>Table des figures</b>	<b>25</b>
<b>Introduction</b>	<b>35</b>
<b>1 Le climat urbain : état de l'art</b>	<b>41</b>
1.1 Îlot de chaleur urbain : considérations générales	41
1.1.1 Définition	41
1.1.1.1 Îlot de Chaleur Urbain de surface, de sol et atmosphérique	41
1.1.1.2 Positionnement de la thèse	43
1.1.2 Échelles d'analyse	43
1.1.2.1 Causes	43
1.1.2.2 Échelles spatiales	44
1.1.2.3 Échelles temporelles	44
1.1.2.4 Positionnement de la thèse	46
1.1.3 Méthodes d'analyse	46
1.1.3.1 Simulation numérique basée sur des modèles de connaissances	47
1.1.3.2 Calcul empirique par régression géoclimatique	50
1.1.3.3 Classification par caractérisation géoclimatique	53
1.1.3.4 Calcul par interpolation géostatistique	55
1.1.3.5 Positionnement de la thèse	56
1.2 Îlot de chaleur urbain : modélisation par régression géoclimatique	58
1.2.1 Acquisition des données météorologiques	58
1.2.1.1 Caractéristiques du matériel de mesure	59



1.2.1.2	Positionnement des capteurs de température . . . . .	62
1.2.1.3	Méthodes de mesure . . . . .	64
1.2.2	Choix d'une station comme référence du climat rural . . . . .	66
1.2.3	Caractérisation des variations temporelles . . . . .	67
1.2.3.1	Variations journalières et saisonnières . . . . .	67
1.2.3.2	Autres échelles de variations . . . . .	70
1.2.4	Caractérisation des variations spatiales . . . . .	74
1.2.4.1	Donnée spatiale et indicateurs géographiques . . . . .	74
1.2.4.2	Forme et taille de la zone d'influence . . . . .	78
1.2.5	Calibration et vérification des modèles . . . . .	82
1.2.5.1	Calibration des modèles . . . . .	82
1.2.5.2	Vérification des modèles . . . . .	82
1.2.6	Positionnement de la thèse . . . . .	84
<b>2</b>	<b>Les données météorologiques mesurées</b>	<b>87</b>
2.1	Présentation des zones d'étude . . . . .	88
2.2	Présentation des réseaux ou ensembles de capteurs . . . . .	90
2.2.1	Réseaux de capteurs de température . . . . .	90
2.2.1.1	Positionnement des capteurs . . . . .	92
2.2.1.2	Description du matériel de mesure . . . . .	92
2.2.1.3	Méthode et matériel d'acquisition des données . . . . .	94
2.2.1.4	Maintenance et calibration du matériel . . . . .	96
2.2.1.5	Conclusion sur les réseaux de capteurs de température . . . . .	96
2.2.2	Ensembles de capteurs météorologiques . . . . .	97
2.3	Présentation des jeux de données . . . . .	101
2.3.1	Jeu de données de calibration . . . . .	104
2.3.2	Jeux de données de vérification provenant de la même zone d'étude . . . . .	108
2.3.2.1	Acquisition des données à une autre période . . . . .	108
2.3.2.2	Acquisition des données pour d'autres stations . . . . .	109
2.3.2.3	Acquisition des données à une autre période et pour d'autres stations . . . . .	111
2.3.3	Jeux de données de vérification provenant d'une autre zone d'étude . . . . .	114
2.3.3.1	Agglomération d'Angers . . . . .	115

2.3.3.2	Agglomération de la Roche-sur-Yon . . . . .	116
2.4	Méthode de correction automatique des températures erronées . . . . .	117
2.4.1	Mise en évidence de l'erreur . . . . .	117
2.4.2	Hypothèses explicatives de l'erreur . . . . .	122
2.4.2.1	Erreur de positionnement . . . . .	122
2.4.2.2	Erreur matériel . . . . .	125
2.4.3	Formulation d'un modèle de correction de l'erreur matérielle . . . . .	129
2.4.3.1	Hypothèses explicatives de l'erreur . . . . .	129
2.4.3.2	Mise en équation de l'erreur : modélisation . . . . .	129
2.4.3.3	L'abri ventilé mécaniquement : une référence ? . . . . .	135
2.4.3.4	Confrontation du modèle de l'erreur aux observations . . . . .	139
2.4.3.5	Calibration d'un modèle de correction . . . . .	141
2.4.4	Vérification du modèle de correction . . . . .	151
2.4.4.1	Étalonnage des capteurs utilisés . . . . .	152
2.4.4.2	Évaluation des performances du modèle . . . . .	153
2.4.5	Application du modèle : correction des jeux de données . . . . .	156
2.4.5.1	Périodes de correction : rayonnement solaire direct sur l'abri . . . . .	157
2.4.5.2	Méthode d'identification des périodes de correction . . . . .	157
2.4.5.3	Confrontation des périodes d'erreur identifiées aux périodes d'erreur observables . . . . .	158
2.4.5.4	Correction des jeux de données . . . . .	163
2.4.5.5	Évaluation de la méthode de correction . . . . .	166
2.5	Conclusion du chapitre . . . . .	167
<b>3</b>	<b>Îlot de chaleur urbain : calibration des modèles</b>	<b>169</b>
3.1	Méthodologie générale de résolution du problème spatial et temporel . . . . .	170
3.1.1	Îlot de chaleur urbain : calcul aux périodes clés . . . . .	170
3.1.1.1	Conditions préalables à l'identification d'une période clé . . . . .	170
3.1.1.2	Identification des périodes clés en journée adimensionnée . . . . .	173
3.1.1.3	Calcul d'indicateurs de température . . . . .	176
3.1.1.4	Estimation de la valeur des extremums journaliers . . . . .	176
3.1.2	Îlot de chaleur urbain : procédure de modélisation . . . . .	178

3.1.2.1	Observation de la distribution spatiale et temporelle des indicateurs de température . . . . .	179
3.1.2.2	Formalisation d'une méthode de résolution du problème spatial et temporel : notion d'ensemble de modèles . . . . .	180
3.1.2.3	Des alternatives à la méthode principale de résolution . . . . .	181
3.2	Modélisation temporelle par régression linéaire multiple . . . . .	192
3.2.1	Élaboration des variables explicatives . . . . .	193
3.2.1.1	Proposition de variables météorologiques . . . . .	193
3.2.1.2	Proposition de périodes d'intérêt . . . . .	195
3.2.1.3	Proposition de fonctions usuelles . . . . .	195
3.2.1.4	Procédure utilisée pour choisir les variables explicatives les plus pertinentes . . . . .	198
3.2.2	Procédure d'identification du modèle temporel le plus pertinent . . . . .	199
3.3	Modélisation spatiale par régression linéaire multiple . . . . .	200
3.3.1	Identification d'indicateurs géographiques pertinents . . . . .	201
3.3.1.1	Indicateurs thématiques . . . . .	201
3.3.1.2	Indicateurs spatiaux . . . . .	207
3.3.2	Procédure et données utilisées pour le calcul des indicateurs géographiques . . . . .	208
3.3.2.1	Indicateurs thématiques . . . . .	208
3.3.2.2	Indicateurs spatiaux . . . . .	211
3.3.3	Sites d'observation du microclimat urbain : quelle représentativité géographique des zones d'étude ? . . . . .	213
3.3.4	Procédure d'identification du modèle spatial le plus pertinent . . . . .	216
3.4	Conclusion du chapitre . . . . .	218
<b>4</b>	<b>Vérification et choix de l'ensemble de modèles le plus performant</b>	<b>221</b>
4.1	Indicateur retenu pour juger les performances d'un ensemble de modèles . . . . .	221
4.2	Performances générales des ensembles de modèles . . . . .	223
4.3	Choix du meilleur ensemble de modèles . . . . .	228
4.4	Discussion de la pertinence physique des modèles spatiaux et du modèle temporel retenus . . . . .	231
4.5	Analyse spatiale et temporelle des résultats . . . . .	240

4.5.1	Principes d'analyse des résultats . . . . .	240
4.5.2	Analyse spatiale des résultats . . . . .	243
4.5.3	Analyse temporelle des résultats . . . . .	249
4.6	Améliorations à apporter . . . . .	256
4.7	Conclusion du chapitre . . . . .	257
<b>5</b>	<b>Modélisation du climat urbain : exemples d'application</b>	<b>259</b>
5.1	Reconstitution du signal de température de n'importe quel point du territoire . .	259
5.1.1	Procédure de reconstitution . . . . .	260
5.1.2	Application : simulation numérique du microclimat d'un quartier . . .	265
5.2	Spatialisation des résultats : cartes climatiques . . . . .	268
5.2.1	Îlot de chaleur urbain et modification des besoins de chaleur et de froid	268
5.2.2	Îlot de chaleur urbain et projets d'alerte aux allergènes d'origine végétale	276
<b>6</b>	<b>Conclusions et discussions</b>	<b>281</b>
6.1	Synthèse des résultats obtenus . . . . .	281
6.2	Limites techniques du travail effectué et futures travaux . . . . .	287
6.2.1	Modification des données . . . . .	288
6.2.2	Modification des méthodes et des modèles . . . . .	289
6.3	Mise en oeuvre de la méthodologie pour d'autres zones urbaines . . . . .	292
<b>A</b>	<b>Contexte géographique de chaque site de mesure</b>	<b>305</b>
<b>B</b>	<b>Données climatiques de chaque site de mesure</b>	<b>313</b>



# Nomenclature

## Symboles romains

$\dot{T}$	Taux de réchauffement de l'air	$^{\circ}C.s^{-1}$
$A$	Surface de contact entre un solide et l'air environnant	$m^2$
$A_g$	Surface de territoire	$m^2$
$A_{g_{ref}}$	Surface de l'unité géographique de référence utilisée pour le calcul des indicateurs géographiques	$m^2$
$A_{sec}$	Surface d'une section	$m^2$
$AH$	Humidité absolue de l'air	$kg.m^{-3}$
$AIC$	Critère d'information d'Akaike	-
$AIC_c$	Critère d'information corrigé d'Akaike	-
$B$	Besoins de chaleur ou de froid d'un bâtiment	$W$
$B_{rel}$	Besoins de chaleur ou de froid d'un bâtiment relativement aux besoins du même bâtiment situé sur un site de référence <i>ref</i>	%
$C_p$	Capacité thermique massique	$J.kg^{-1}.K^{-1}$
$corr_T$	Correction à appliquer systématiquement aux données de température $T$	$^{\circ}C$
$D$	Indicateur géographique désignant une densité	-
$d_{dil}$	Distance de dilatation des bâtiments	$m$
$d_{mod/i}$	Distance entre un point <i>mod</i> où interpoler et une station <i>i</i> utilisée pour l'interpolation	$m$

$d_{mur}$	Distance entre un point de mesure et le mur le plus proche	$m$
$d_{periph}$	Indicateur géographique représentant la distance d'un point à la périphérie de la ville	$m$
$dir_{pr}$	Direction correspondante à une part $pr$ de la rose des vents	$rad$
$DJU$	Degrés-jours-unifiés de chaleur ou de froid	$^{\circ}C$
$E_T$	Erreur de mesure de la température en conditions stationnaires	$^{\circ}C$
$e_T$	Erreur de mesure de la température	$^{\circ}C$
$e_{res}$	Erreur de modélisation statistique correspondant aux résidus du modèle $y$	unité de $y$
$f()$	Relation mathématique fonctionnelle reliant plusieurs grandeurs	-
$f_L$	Fonction de vraisemblance	-
$H$	Hauteur moyenne de bâtiments	$m$
$h$	Coefficient de convection entre une surface et l'air	$J.m^{-2}.K^{-1}$
$HP$	Hauteur de précipitation	$m$
$JDD_{calib}$	Jeu de données utilisé pour <b>calibrer</b> les modèles <b>spatiaux et temporels</b> - acquis au sein de la zone d'étude de Nantes	-
$JDD_{extrapol}^{(Angers)}$	Jeu de données utilisé pour <b>vérifier</b> que les modèles <b>spatiaux</b> et les modèles <b>temporels</b> sont extrapolables à d'autres zones d'études - acquis au sein de la zone d'étude d' <b>Angers</b>	-
$JDD_{extrapol}^{(LRY)}$	Jeu de données utilisé pour <b>vérifier</b> que les modèles <b>spatiaux</b> et les modèles <b>temporels</b> sont extrapolables à d'autres zones d'études - acquis au sein de la zone d'étude de <b>LRY</b>	-

$JDD_{spat\&temp}$	Jeu de données utilisé pour <b>vérifier</b> les modèles <b>spatiaux</b> et les modèles <b>temporels</b> - acquis au sein de la même zone d'étude que le jeu de données de calibration (Nantes)	-
$JDD_{spat}$	Jeu de données utilisé pour <b>vérifier</b> les modèles <b>spatiaux</b> - acquis au sein de la même zone d'étude que le jeu de données de calibration (Nantes)	-
$JDD_{temp}$	Jeu de données utilisé pour <b>vérifier</b> les modèles <b>temporels</b> - acquis au sein de la même zone d'étude que le jeu de données de calibration (Nantes)	-
$K \downarrow$	Densité de flux radiatif courtes ongueurs d'onde (CLO) descendant	$W.m^{-2}$
$K \uparrow$	Densité de flux radiatif courtes ongueurs d'onde (CLO) montant	$W.m^{-2}$
$K_{\perp} \downarrow$	Densité de flux radiatif courtes ongueurs d'onde (CLO) descendant sur une surface perpendiculaire aux rayons incidents	$W.m^{-2}$
$L \downarrow$	Densité de flux radiatif grandes longueurs d'onde descendant	$W.m^{-2}$
$L \uparrow$	Rayonnement grandes longueurs d'onde montant	$W.m^{-2}$
$m$	Masse	$kg$
$N$	Nébulosité	$octa$
$nb$	Nombre (quantité)	-
$nb_{ag}$	Nombre d'agrégat de bâtiments dans la zone urbaine considérée	-
$nb_{ech}$	Nombre d'échantillons utilisé pour une régression linéaire	-
$nb_{EM}$	Nombre d'ensembles de modèles (EM)	-
$nb_{hab}$	Nombre d'habitants d'une agglomération	-
$nb_{JDD}$	Nombre de jeux de données (JDD)	-



$nb_{x_n}$	Nombre de variables explicatives utilisées dans une régression linéaire multiple	-
$NDVI$	Indicateur géographique désignant l'indice de végétation par différence normalisée ou indice de Tucker	-
$p_{val}$	Valeur p représentant la probabilité que l'hypothèse de non-significativité du coefficient d'une relation linéaire soit vraie	-
$PA$	Pression atmosphérique	$Pa$
$Q^*$	Densité de flux radiatif net	$W.m^{-2}$
$Q_E$	Densité de flux de chaleur latente	$W.m^{-2}$
$Q_F$	Densité de flux de chaleur anthropique	$W.m^{-2}$
$Q_H$	Densité de flux de chaleur sensible	$W.m^{-2}$
$r$	Résolution d'un capteur	unité de la grandeur physique mesurée
$\widehat{RMSE}$	Résultat du calcul de la fonction $rm\widehat{se}()$	unité de ()
$R^2$	Coefficient de détermination - mesure la qualité de la prédiction d'une régression linéaire	-
$R_{tampon}$	Rayon du cercle correspondant à l'unité de référence pour le calcul des indicateurs géographiques	$m$
$RE$	Réflectance d'une surface	-
$RH$	Humidité relative de l'air	%
$RMSE$	Résultat du calcul de la fonction $rmse$	unité de ()
$S$	Section	$m^2$
$SVF$	Facteur de vue du ciel	-
$T$	Température de l'air	$^{\circ}C$
$t$	Temps	$s$
$T_S$	Température de surface	$^{\circ}C$

$T_{base}$	Température de base d'une plante, représentant la température d'air minimale nécessaire pour assurer la croissance de la plante	$^{\circ}C$
$T_{IN}$	Température d'air en entrée d'un système	$^{\circ}C$
$T_{max}$	Température maximale journalière de l'air	$^{\circ}C$
$T_{min}$	Température minimale journalière de l'air	$^{\circ}C$
$T_{moy}$	Température moyenne journalière de l'air	$^{\circ}C$
$U$	Vitesse du vent	$m.s^{-1}$
$UA$	Conductance thermique	$W.K^{-1}$
$V$	Volume	$m^3$
$VM_m$	Notation générique désignant une variable météorologique $m$	unité de $VM_m$
$W$	Largeur moyenne de rues	$m$
$x_g$	Variable explicative des variations spatiales de l'ICU. Construite à partir d'un indicateur géographique $IG_g$	unité de $x_g$
$x_m$	Variable explicative des variations temporelles de l'ICU. Construite à partir d'un indicateur météorologique $IMR_m$	unité de $x_m$
$x_n$	Variable statistique $n$ explicative d'un modèle $y$	unité de $x_n$
$y$	Variable statistique à modéliser	unité de $y$
$Z$	Angle zénithal solaire	$rad$

### Symboles grecs

Les symboles grecs sont assimilables aux coefficients ou constantes, exceptés ceux qui utilisent l'opérateur de différence  $\Delta$ . La plupart des coefficients et des constantes ne sont pas décrits ici. Ils seront directement définis dans le manuscrit lorsqu'ils seront utilisés.

$\alpha_{smg}$	Coefficient de régression de la variable explicative $x_g$ dans le modèle de l'indicateur $ITLS_{moy}$	$^{\circ}C.[x_g]^{-1}$
----------------	--	------------------------

$\alpha_{ssg}$	Coefficient de régression de la variable explicative $x_g$ dans le modèle de l'indicateur $ITLS_{std}$	$^{\circ}C.[x_g]^{-1}$
$\alpha_{tm}$	Coefficient de régression de la variable explicative $x_m$ dans le modèle de l'indicateur $ITLT$	$[x_m]^{-1}$
$\Delta Q_A$	Terme d'advection horizontale	$W.m^{-2}$
$\Delta Q_S$	Terme de stockage dans les matériaux (sols, murs, toits)	$W.m^{-2}$
$\Delta T$	Différence de température entre la ville et sa périphérie (si aucun indice n'est précisé)	$^{\circ}C$
$\Delta t$	Pas de temps de mesure	$s$
$\delta t$	Intervalle de temps d'une durée inférieure à celle du pas de temps de mesure	$s$
$\Delta T_{1m/d_0}$	Différence de température entre un point situé à 1 m du mur et un point de référence situé à une distance $d_0$ du mur	$^{\circ}C$
$\Delta T_{i-ref}$	Différence de température entre une station $i$ et une station de référence $ref$	$^{\circ}C$
$\rho$	Masse volumique de l'air à 20°C et sous 1 atm	$1,204 kg.m^3$
$\tau$	Constante de temps d'un système	$s$

### Principales opérations statistiques

$\frac{q_2}{q_1}int()_v$	Calcule un <b>intercentile</b> d'un jeu de données (différence entre les centiles $q_1$ et $q_2$ calculés selon la variable $v$ )	unité de ()
$cen_q()_v$	Calcule le <b>centile</b> $q$ d'un jeu de données selon la variable $v$	unité de ()
$eam()_v$	Calcule l' <b>équivalent robuste de l'écart-type</b> d'un jeu de données (basé sur des calculs de médianes) selon la variable $v$	unité de ()
$max()_v$	Calcule la valeur <b>maximale</b> d'un jeu de données selon la variable $v$	unité de ()
$med()_v$	Calcule la valeur <b>médiane</b> d'un jeu de données selon la variable $v$	unité de ()

$min()_v$	Calcule la valeur <b>minimale</b> d'un jeu de données selon la variable $v$	unité de ()
$moy()_v$	Calcule la valeur <b>moyenne</b> d'un jeu de données selon la variable $v$	unité de ()
$\widehat{rmse}()$	Calcule une <b>version normalisée de la fonction rmse()</b> - Indicateur créé pour simplifier l'interprétation de RMSE	-
$rmse()$	Calcule la <b>racine carrée de l'erreur quadratique moyenne</b> d'un jeu de données par rapport à un jeu de données de référence	unité de ()
$std()_v$	Calcule l' <b>écart-type</b> d'un jeu de données selon la variable $v$	unité de ()
$var()_v$	Calcule la <b>variance</b> d'un jeu de données selon la variable $v$	unité de () <sup>2</sup>
$var_{mes}()_v$	Calcule la <b>variance</b> d'un jeu de données imputable à l'incertitude de la <b>mesure</b> (variance calculée selon la variable $v$ )	unité de () <sup>2</sup>
$var_{met}()_v$	Calcule la <b>variance</b> d'un jeu de données imputable aux variations des conditions <b>météorologiques</b> (variance calculée selon la variable $v$ )	unité de () <sup>2</sup>
$var_{tot}()_v$	Calcule la <b>variance totale</b> d'un jeu de données selon la variable $v$	unité de () <sup>2</sup>

### Notations pour désigner les variations spatiales et temporelles

Deux types de variations de la température sont traités dans ces travaux : les variations spatiales et les variations temporelles. Les variations spatiales sont décrites par l'indice ( $i$ ) (nom d'une station ou coordonnées spatiales d'un point de l'espace). Les variations temporelles sont décrites par l'exposant ( $S_s$ ) (saison), la variable ( $j$ ) (date du jour), l'exposant correspondant à la période diurne ( $^{jour}$ ) ou à la période nocturne ( $^{nuit}$ ), la variable  $P_p$  (autres périodes) et l'indice  $k$  (pas de temps au sein de  $j$ )

Exemples :  $X_{i,k}^{(pri)}(j)$  correspond à la valeur de la variable  $X$  mesurée à la station  $i$  au pas de temps  $k$  d'une journée  $j$  printanière

<i>(jour)</i>	Période diurne	-
<i>(nuit)</i>	Période nocturne	-
$(S_s)$	Notation générique désignant une saison $s$ ( <sup><i>hiv</i></sup> ) pour l'hiver, ( <sup><i>pri</i></sup> ) pour le printemps, ( <sup><i>ete</i></sup> ) pour l'été, ( <sup><i>aut</i></sup> ) pour l'automne	-
$i$	Nom d'une station ou coordonnées d'un point de l'espace	-
$k$	Pas de temps au sein d'une journée $j$	-
$ref$	Désigne la station utilisée pour qualifier les conditions météorologiques à l'échelle de la zone urbaine	-
$j$	Journée définie entre deux levers de soleil successifs	-
$j_{moy}$	Journée moyenne type obtenue en moyennant le profil temporel de plusieurs journées selon une ou plusieurs variables	-
$P_p$	Période $p$ d'intérêt	-
<b>Autres indices</b>		
$a$	Abri ventilé protégeant les capteurs du rayonnement	-
$c$	Capteur de température d'air	-
$\overline{bat}$	Tous les sols non recouverts par un bâtiment	-
$a+c$	Ensemble air de l'abri et capteur de température d'air	-
$a_0$	Abri ventilé de référence (un capteur placé sous cet abri mesure sans erreur la température de l'air extérieur)	-
$a_{meca}$	Abri ventilé mécaniquement	-
$a_{nat}$	Abri ventilé naturellement	-
$bat$	Bâtiment	-
$chaleur$	Désigne la période de chauffage	-
$eau$	Eau	-

<i>fac</i>	Façades	-
<i>froid</i>	Désigne la période de climatisation	-
<i>MF</i>	Propre à l'abri Météo-France	-
<i>pir</i>	Désigne une bande spectrale correspondant au proche-infrarouge	-
<i>pix</i>	Désigne un pixel d'une image	-
<i>rou</i>	Désigne une bande spectrale correspondant au rouge	-
<i>sat</i>	Désigne un satellite de télédétection	-
<i>sol</i>	Sol de la zone urbaine	-
<i>surf</i>	Ensemble des surfaces urbaines : sols, façades et toits de bâtiments	-
<i>TRH</i>	Propre à l'abri TRH	-
<i>URB</i>	Propre à l'abri URBIO	-
$X_{cor}$	Valeurs de X après correction	unité de X
$X_{est}$	Valeurs estimées de X	unité de X
$X_{h=10m}$	Valeur de X à la hauteur de 10 m)	unité de X
$X_{ini}$	Valeurs initiales de X	unité de X
$X_{obs}$	Valeurs observées de X	unité de X
$X_{res}$	Valeurs résiduelles (par exemple erreur résiduelle après correction des valeurs $X_{ini}$ )	unité de X
$X_{seuil}$	Valeur seuil de la variable X)	unité de X

### Notations génériques des indicateurs géographiques et météorologiques

<i>ITL</i>	Indicateur météorologique décrivant les conditions locales de <b>température</b>	°C
<i>ITLS</i>	Indicateur <b>climatique</b> décrivant les conditions locales de température (détenant une information purement <b>spatiale</b> des variations de l'ICU)	°C

$ITLS_{moy}$	Valeur <b>moyenne</b> de l'indicateur $ITL$ pour la station $i$ )	$^{\circ}C$
$ITLS_{std}$	<b>Écart-type</b> de l'indicateur $ITL$ pour la station $i$ )	$^{\circ}C$
$ITLT$	Indicateur météorologique décrivant les conditions locales de <b>température</b> (détenant une information purement <b>temporelle</b> des variations de l'ICU)	-
$IMR_m$	Indicateur <b>météorologique</b> décrivant les conditions d'une variables météorologique $m$ à l' <b>échelle d'une agglomération</b>	unité de $m$
$IG_g$	Notation générique désignant un indicateur <b>géographique</b> $g$	unité de $IG_g$

### Acronymes

BDD	Base de données
EM	Ensemble de modèles
ICU	Îlot de chaleur urbain
JDD	Jeux de données
LCZ	<i>Local climate zone</i>
MF	Météo-France
OMM	Organisation météorologique mondiale
RG	Régression géoclimatique

# Liste des tableaux

1.1	Principales caractéristiques des méthodes d’analyse de l’ICU . . . . .	57
2.1	Informations sur le positionnement des capteurs de température pour chaque réseau . . . . .	92
2.2	Caractéristiques du matériel de mesure de température d’air pour chaque réseau de capteurs . . . . .	93
2.3	Caractéristiques du matériel et des méthodes utilisés pour l’acquisition des données de température d’air de chaque réseau de capteurs . . . . .	94
2.4	Définition des principales grandeurs physiques météorologiques mesurées par Météo-France . . . . .	98
2.5	Caractéristiques du matériel utilisé et des méthodes employées relativement à 16 études issues de la littérature pour produire des données de température d’air	107
2.6	Valeurs des coefficients de l’équation 2.19 déterminées par régression linéaire .	139
2.7	Valeurs des coefficients des équations 2.21 et 2.22 déterminées par régression linéaire après centrage et réduction des variables explicatives . . . . .	151
2.8	$RMSE_{res}$ et $\widehat{RMSE}$ de l’erreur estimée à partir de chaque modèle : Nakamura, Cheng et celui proposé dans cette thèse . . . . .	156
2.9	Fréquence et valeurs moyennes saisonnières des corrections apportées aux valeurs de température mesurées par chaque station des réseaux TRH et FluxSAP	165
3.1	Indicateurs géographiques thématiques et bases de données utilisées pour leur calcul . . . . .	206
4.1	Variante utilisées par les cinq meilleurs ensembles de modèles pour la combinaison {saison : automne, jeu de données : calibration, indicateur $ITL$ : indicateur nocturne $ITL^{(nuit)}$ }. . . . .	226



4.2	Variantes utilisées par le modèle $ITL^{(nuit)}$ (retenu pour chaque saison) et performances ( $\widehat{RMSE}$ ) obtenues pour chaque jeu de données. . . . .	230
4.3	Valeurs <i>normalisées</i> des termes $\alpha_{t_{VM_m}} amp_{x_{VM_m}}$ issus des équations régissant le calcul de l'indicateur $ITLT^{(nuit)}$ . . . . .	238
4.4	Valeurs <i>normalisées</i> des termes $\alpha_{ss_{IG_g}} amp_{x_{IG_g}}$ issus des équations régissant le calcul de l'indicateur $ITLS_{std}^{(nuit)}$ . . . . .	238
4.5	Valeurs <i>normalisées</i> des termes $\alpha_{sm_{IG_g}} amp_{x_{IG_g}}$ issus des équations régissant le calcul de l'indicateur $ITLS_{moy}^{(nuit)}$ . . . . .	239
5.1	Nombre de jours séparant la floraison du lierre dans chacun des parcs à celle dans le Jardin des Plantes . . . . .	280

# Table des figures

1	Plan de la thèse . . . . .	39
1.1	Représentation des principales couches contenues dans la couche limite atmosphérique (CLA) . . . . .	42
1.2	Imbrication des échelles spatiales de la climatologie urbaine . . . . .	45
1.3	Représentation schématique du bilan d'énergie d'un volume d'air de la CLS . . . . .	47
1.4	Classification par caractéristiques des modèles numériques de simulation du climat urbain à l'échelle locale . . . . .	49
1.5	Relation entre le logarithme du nombre d'habitants d'une zone urbaine et son ICU lors de conditions météorologiques claires et calmes . . . . .	51
1.6	Relation entre la température moyenne diurne et nocturne et la densité de végétation . . . . .	53
1.7	Présentation des 17 LCZ existants . . . . .	54
1.8	Exemples d'abris ventilés utilisés pour protéger les capteurs de température du rayonnement. Deux catégories existent : les abris à ventilation naturelle comme (a) l'abri de Stevenson <i>modèle réduit</i> ou (b) l'abri de Stevenson <i>miniature</i> et (c) les abris à ventilation forcée . . . . .	61
1.9	Critères d'installation d'un capteur de température en classe 1 . . . . .	63
1.10	Échantillonnage de valeurs de température pour différents pas de temps et différents capteurs de mesure . . . . .	65
1.11	Chaîne de traitement classique utilisée pour traiter les variations temporelles de l'ICU . . . . .	67
1.12	Classification des différentes approches utilisées dans la littérature pour traiter les variations temporelles de l'ICU . . . . .	69
1.13	Profil typique du flux de chaleur anthropique généré pour (a) les jours de la semaine, et (b) les jours du week-end . . . . .	71

1.14	Méthodes classiquement utilisées dans la littérature pour traiter de l'influence des conditions météorologiques sur les variations temporelles de l'ICU. . . . .	73
1.15	Catégorisation des principaux indicateurs et bases de données géographiques qui sont utilisés dans la littérature traitant de modélisation de l'ICU par des méthodes de régression géoclimatique . . . . .	76
1.16	Représentation conceptuelle de la notion de footprint, attribuant la température mesurée par le capteur à différentes surfaces alentour . . . . .	79
1.17	Classification des solutions proposées dans la littérature pour identifier les surfaces qui contribuent à la température d'air mesurée par le capteur. . . . .	80
2.1	Situation géographique des trois zones d'études . . . . .	88
2.2	Disposition des sites de mesure pour chacune des zones d'études : (a) agglomération de Nantes, (b) agglomération d'Angers, (c) agglomération de la Roches-sur-Yon . . . . .	91
2.3	Méthode employée pour synchroniser les données issues de deux capteurs de température des réseaux TRH ou FluxSAP . . . . .	95
2.4	Interpolation linéaire quart-horaire appliquée à un signal de direction de vent .	100
2.5	Réseaux/ensembles de capteurs et jeux de données utilisés pour calibrer et vérifier la modélisation de l'ICU à l'échelle d'une zone urbaine . . . . .	101
2.6	Évolution de la quantité de données utilisables selon le nombre de stations de mesure utilisées ( $JDD_{calib}$ ) . . . . .	102
2.7	Répartition saisonnière des journées du jeu de données de calibration utilisées .	104
2.8	Répartition du type de temps selon la saison et la direction du vent pour le jeu de données de calibration . . . . .	106
2.9	Répartition saisonnière des journées du jeu de données de vérification $JDD_{temp}$ utilisées . . . . .	109
2.10	Répartition du type de temps selon la saison et la direction du vent pour le jeu de données de vérification $JDD_{temp}$ . . . . .	110
2.11	Répartition saisonnière des journées du jeu de données de vérification $JDD_{spat}$ utilisées . . . . .	111
2.12	Répartition du type de temps selon la saison et la direction du vent pour le jeu de données de vérification $JDD_{spat}$ . . . . .	112

2.13 Répartition saisonnière des journées du jeu de données de vérification $JDD_{spat\&temp}$ utilisées . . . . .	112
2.14 Répartition du type de temps selon la saison et la direction du vent pour le jeu de données de vérification $JDD_{spat\&temp}$ . . . . .	113
2.15 Répartition saisonnière des journées du jeu de données de vérification $JDD_{extrapol}^{(Angers)}$ utilisées . . . . .	115
2.16 Répartition du type de temps selon la saison et la direction du vent pour le jeu de données de vérification $JDD_{extrapol}^{(Angers)}$ . . . . .	116
2.17 Répartition saisonnière des journées du jeu de données de vérification $JDD_{extrapol}^{(LRY)}$ utilisées . . . . .	117
2.18 Répartition du type de temps selon la saison et la direction du vent pour le jeu de données de vérification $JDD_{extrapol}^{(LRY)}$ . . . . .	118
2.19 Station de mesure où une erreur de mesure de température est observée : (a) photographie de la station sur site, (b) illustration de l'évolution temporelle de la température observée un jour de ciel clair . . . . .	119
2.20 Méthodologie adoptée pour corriger un jeu de données dont certaines stations engendrent des erreurs de mesure de la température d'air . . . . .	121
2.21 Isothermes mesurés au sein d'une rue canyon orientée est-ouest . . . . .	123
2.22 Description du site de mesure et du positionnement du matériel utilisé pendant la campagne de mesure 1 . . . . .	124
2.23 Différence de température entre l'air au voisinage d'un mur soumis au rayonnement solaire et l'air mesuré à 1 m du mur . . . . .	125
2.24 Description du site de mesure et du positionnement du matériel utilisé pendant la campagne de mesure 2 . . . . .	126
2.25 Évolution temporelle de chacune des variables météorologiques mesurées localement lors de la campagne de mesure 2. . . . .	127
2.26 Représentation schématique des échanges de chaleur observables entre un abri ventilé renfermant un capteur de température d'air et leur environnement . . . . .	130
2.27 Évolution de la différence de température nocturne mesurée entre l'abri URBIO et l'abri ventilé mécaniquement en fonction du taux de réchauffement de l'air et après classification par intervalles de vitesse du vent . . . . .	132

2.28	Évolution de la différence de température nocturne mesurée entre l'abri TRH et l'abri ventilé mécaniquement en fonction de la vitesse du vent et après classification par intervalles de taux de réchauffement de l'air . . . . .	136
2.29	Évolution de la différence de température nocturne mesurée entre l'abri UR-BIO et l'abri ventilé mécaniquement en fonction de la vitesse du vent et après classification par intervalles de taux de réchauffement de l'air . . . . .	137
2.30	Évolution de la différence de température diurne mesurée entre l'abri UR-BIO et l'abri ventilé mécaniquement en fonction de la vitesse du vent et après classification par intervalles de rayonnement solaire montant et descendant . . . . .	138
2.31	Évolution de la différence de température diurne mesurée entre l'abri TRH et l'abri ventilé mécaniquement en fonction du rayonnement global descendant et montant et après classification par intervalles de vitesse du vent. . . . .	140
2.32	Erreur (a) diurne et (b) nocturne estimée à partir des trois modèles étudiés en fonction de l'erreur observée . . . . .	145
2.33	Résidus de l'erreur (a) diurne et (b) nocturne selon le modèle utilisé et en fonction de chaque variable météorologique $VM_m$ . . . . .	147
2.34	Erreur (a) diurne et (b) nocturne estimée à partir des trois modèles étudiés en fonction de l'erreur observée . . . . .	150
2.35	Description du site de mesure et du positionnement du matériel utilisé pendant la campagne de mesure 3 . . . . .	152
2.36	Erreur résiduelle $e_{T_{res}}$ après correction et erreur $e_{T_{ini}}$ si aucune correction n'est effectuée . . . . .	155
2.37	Erreur diurne estimée à partir des trois modèles étudiés en fonction de l'erreur observée durant la campagne de vérification 3 : (a) pour les abris propres, (b) pour les abris sales . . . . .	156
2.38	Confrontation des périodes d'ensoleillement du capteur TRH11 avec celles de forte variation du taux de réchauffement de l'air . . . . .	160
2.39	Maquette Sketchup du site TRH11 avec ombres portées à 12h (heure solaire) . . . . .	161
2.40	Maquette Sketchup du site TRH06 avec ombres portées à 12h (heure solaire) . . . . .	161
2.41	Confrontation des périodes d'ensoleillement du capteur TRH06 avec celles de forte variation du taux de réchauffement de l'air . . . . .	162
2.42	Performances de la méthode de correction proposée - Application à la station 11 du réseau TRH . . . . .	166

3.1	Évolution de la dispersion interstation de la température et de celle du taux de réchauffement au cours d'une journée moyenne pour le jeu de données de calibration $JDD_{calib}$ : (a) selon l'indicateur <i>écart-type</i> et (b) selon l'indicateur <i>intercentile</i> . . . . .	172
3.2	Évolution (en temps adimensionné) de la dispersion interstation de la température et de celle du taux de réchauffement au cours d'une journée moyenne pour le jeu de données de calibration $JDD_{calib}$ : (a) selon l'indicateur <i>écart-type</i> et (b) selon l'indicateur <i>intercentile</i> . . . . .	174
3.3	Évolution (en temps adimensionné) de la dispersion interstation de la température et de celle du taux de réchauffement au cours d'une journée moyenne pour le jeu de données de vérification $JDD_{spat\&temp}$ selon l'indicateur <i>écart-type</i> . . . . .	175
3.4	Procédure de calcul des indicateurs météorologiques <i>ITL</i> décrivant les températures locales . . . . .	177
3.5	Distribution saisonnière des valeurs de l'indicateur de température locale nocturne $ITL^{(nuit)}$ pour les stations TRH04 et TRH11 . . . . .	180
3.6	Distribution des valeurs de l'indicateur <i>ITL</i> selon le type de temps avant et après centrage et réduction . . . . .	182
3.7	Méthodologie générale utilisée pour modéliser les variations spatiales et temporelles de l'ICU . . . . .	183
3.8	Division microclimatique et géographique d'une station de mesure en plusieurs sites : exemple pour la direction de vent $dir_1 = [0 - 60]^\circ$ . . . . .	184
3.9	Centrage et réduction appliqués aux données de deux stations de mesure (a) lorsque les deux sites sont isotropes, (b) lorsqu'au moins un des deux sites est anisotrope . . . . .	185
3.10	$\Delta ITL_{i,dir_{pr}}$ et contexte géographique des stations TRH11, TRH10 et MF_Boug selon la direction du vent . . . . .	187
3.11	Distribution (histogramme) des valeurs et valeurs pour chaque station et chaque direction de vent (a) de l'indicateur $ITLS_{moy}$ et (b) de l'indicateur $ITLS_{std}$ lorsque la variante <i>tri directionnel</i> est utilisée . . . . .	188
3.12	Distribution des valeurs (a) de l'indicateur $ITLT^{(jour)}(j)$ et (b) de l'indicateur $ITLT^{(nuit)}(j)$ selon si la variante <i>tri saisonnier</i> est utilisée . . . . .	190
3.13	Périodes $P_m$ utilisées pour calculer les indicateurs météorologiques <i>IMR</i> associés à chacun des indicateurs $ITL^{(nuit)}$ et $ITL^{(jour)}$ . . . . .	196

3.14	Valeurs prises par l'indicateur nocturne $ITLT^{(nuit)}(j)$ en fonction de chaque indicateur météorologique $IMR_m$ calculé pour chacune des périodes $P_m$ . . . .	197
3.15	Procédure utilisée pour déterminer le terme $x_m$ (propre à chaque variable météorologique $VM_m$ ) le plus approprié pour expliquer les variations de l'indicateur $ITLT(j)$ . . . . .	198
3.16	Procédure utilisée pour déterminer la combinaison des termes $x_m$ qui explique le mieux les variations de $ITLT(j)$ . . . . .	199
3.17	Surfaces urbaines élémentaires utilisées pour calculer les indicateurs géographiques . . . . .	202
3.18	Représentation schématique de l'acquisition d'une image satellitaire multi-spectrale à partir d'un satellite de résolution $r_{sat} = 5m$ enregistrant cinq canaux . . . . .	205
3.19	Associations d'un ou de plusieurs indicateurs géographiques à chacun des termes du bilan d'énergie d'une surface urbaine . . . . .	207
3.20	Différenciation des surfaces considérées comme urbaines de celles considérées comme non-urbaines pour le centre-ville de Nantes . . . . .	213
3.21	Répartition des 200 points tirés aléatoirement sur le territoire nantais . . . . .	214
3.22	Distribution des valeurs de chaque indicateur géographique $IG_g$ calculées pour 200 points aléatoirement choisis sur la zone d'étude et pour les stations de chaque jeu de données utilisé. (a) agglomération de Nantes, (b) agglomération d'Angers ( $JDD_{extrapol}^{(Angers)}$ ), (c) agglomération de LRY ( $JDD_{extrapol}^{(LRY)}$ ) . . . . .	215
3.23	Évolution de l'indicateur $ITLS_{moy_i}$ en fonction de l'indicateur géographique $D_{surf}$ pour la saison hivernale lorsque la variante <i>tri directionnel</i> est utilisée . .	217
3.24	Représentation schématique des phases (a) de calibration et (b) de vérification (à partir d'un nouveau jeu de données) d'un modèle linéaire et d'un modèle non-linéaire . . . . .	218
3.25	Procédure utilisée pour déterminer la combinaison de termes $x_g$ qui explique le mieux les variations des indicateurs $ITLS_{moy}$ et $ITLS_{std}$ . . . . .	218
4.1	Procédure utilisée pour estimer les indicateurs $ITL^{(jour)}$ et $ITL^{(nuit)}$ . Ces estimations permettent ainsi de vérifier la performance des EM en comparant les valeurs estimées et les valeurs observées . . . . .	222
4.2	Distribution saisonnière des valeurs de $RMSE$ et de $\widehat{RMSE}$ pour chacun des jeux de données . . . . .	225

4.3	Fréquence d'utilisation de la variante <i>tri directionnel</i> par les cinq meilleurs EM obtenus pour chaque combinaison {jeu de données, saison} (a) lorsque l'indicateur nocturne est modélisé, (b) lorsque l'indicateur diurne est modélisé . . . . .	227
4.4	Fréquence d'utilisation de la variante <i>tri saisonnier</i> par les cinq meilleurs EM obtenus pour chaque combinaison {jeu de données, saison} lorsque l'indicateur nocturne est modélisé . . . . .	228
4.5	Procédure utilisée pour choisir l'EM le plus performant en considérant tous nos jeux de données . . . . .	229
4.6	Valeurs estimées de l'indicateur $ITL^{(pri)}$ en fonction des valeurs observées pour chaque station de mesure des jeux de données (a) $JDD_{spat\&temp}$ , (b) $JDD_{extrapol}^{(Angers)}$ et (c) $JDD_{extrapol}^{(LRY)}$ . . . . .	241
4.7	Types de dysfonctionnement observables lorsque l'on trace le nuage de point $ITL_{est} = f(ITL_{obs})$ pour une station de mesure donnée : (a) modèle temporel erroné (indicateur $ITLT$ ), (b) modèle spatial de la dispersion erroné (indicateur $ITLS_{std}$ ), (c) modèle spatial de la moyenne erroné (indicateur $ITLS_{moy}$ ) et (d) modèles spatiaux erronés pour une ou plusieurs directions de vent . . . . .	242
4.8	Distribution des valeurs de l'indicateur $NDVI$ calculées pour 200 points aléatoirement choisis sur l'agglomération de Nantes et pour les stations des jeux de données $JDD_{calib}$ et $JDD_{spat\&temp}$ . . . . .	244
4.9	Contexte géographique du site N11 . . . . .	244
4.10	$NDVI$ des stations de référence $MF_{Angers}$ et $MF_{LRY}$ en fonction du rayon de la zone tampon utilisée pour le calcul des indicateurs géographiques thématiques. . . . .	246
4.11	Contexte géographique des stations de référence (a) $MF_{LRY}$ et (b) $MF_{Angers}$ au sein d'unités géographiques de référence concentriques de rayon 50, 100, 200, 300, 400, 500, 750 et 1000 m. . . . .	247
4.12	Distribution des valeurs de l'indicateur $NDVI$ calculées pour 200 points aléatoirement choisis sur l'agglomération de La Roche-sur-Yon et pour les stations du jeu de données $JDD_{extrapol}^{(LRY)}$ . . . . .	247
4.13	Valeurs estimées de l'indicateur $ITL$ en fonction des valeurs observées pour le site N05 . . . . .	249



4.14	Résidus du modèle $ITL^{(pri)}$ (correspondant aux zones 1 et 2 de la figure 4.13) en fonction de chaque indicateur météorologique calculé pour chaque période $P_m$ .	251
4.15	Valeurs estimées de l'indicateur $ITL$ en fonction des valeurs observées pour le site A02	252
4.16	Résidus du modèle $ITL^{(pri)}$ (correspondant aux zones hautes 1 et 2 de la figure 4.15) en fonction de chaque indicateur météorologique calculé pour chaque période $P_m$ .	253
4.17	Résidus du modèle $ITL^{(pri)}$ (correspondant aux zones basses 1 et 2 de la figure 4.15) en fonction de chaque indicateur météorologique calculé pour chaque période $P_m$ .	254
4.18	Évolution de la fonction inverse $f(U) = \frac{1}{U+c}$ pour plusieurs valeurs de constante $c$ .	256
5.1	Évolution (en temps adimensionné) de la dispersion interstation de la température et de celle du taux de réchauffement au cours d'une journée moyenne pour le jeu de données de vérification $JDD_{spat\&temp}$	260
5.2	Représentation schématique de l'évolution au cours d'un cycle diurne de la température de l'air pour une station $i$ et d'une station de référence $ref$ .	261
5.3	Procédure utilisée pour reconstituer à un pas de temps $k$ et pour un ensemble de journées $j$ donnés le signal de température de n'importe quel point $i$ du territoire.	264
5.4	Contexte géographique du site TRH06	266
5.5	Évolution de la température d'air d'un quartier estimée à partir des solutions 1, 2 et 3.	267
5.6	Localisation des 855 points répartis sur le territoire de la ville de Nantes	271
5.7	(a) Triangulation de Delaunay et (b) interpolation linéaire appliquée à une série de points de mesure d'une grandeur physique $X$	272
5.8	Cartographie des besoins relatifs (a) de chauffage et (b) de climatisation pour la ville de Nantes	273
5.9	Évolution temporelle de la somme des degrés-jours de croissance pour différents parcs de la ville de Nantes	278
5.10	Cartographie de la date de floraison de chaque parc étudié par rapport à celle du jardin des plantes.	279

A.1	Contexte géographique des sites du jeu de données $JDD_{calib}$ . . . . .	306
A.2	Contexte géographique des sites du jeu de données $JDD_{temp}$ . . . . .	307
A.3	Contexte géographique des sites du jeu de données $JDD_{spat}$ . . . . .	308
A.4	Contexte géographique des sites du jeu de données $JDD_{spattemp}$ . . . . .	309
A.5	Contexte géographique des sites du jeu de données $JDD_{extrapol}^{(Angers)}$ . . . . .	310
A.6	Contexte géographique des sites du jeu de données $JDD_{extrapol}^{(Angers)}$ . . . . .	311
B.1	Données climatiques des sites du jeu de données $JDD_{calib}$ . . . . .	314
B.2	Données climatiques des sites du jeu de données $JDD_{temp}$ . . . . .	315
B.3	Données climatiques des sites du jeu de données $JDD_{spat}$ . . . . .	316
B.4	Données climatiques des sites du jeu de données $JDD_{spattemp}$ . . . . .	317
B.5	Données climatiques des sites du jeu de données $JDD_{extrapol}^{(LRY)}$ . . . . .	318
B.6	Données climatiques des sites du jeu de données $JDD_{extrapol}^{(LRY)}$ . . . . .	319



# Introduction

Selon le groupe d'experts intergouvernemental sur l'évolution du climat (GIEC), la température moyenne de surface de la Terre ainsi que la température moyenne de l'air ont augmenté de près de 1°C depuis le début des années 1900 (Watson et Albritton 2001). Selon leurs prévisions, cette augmentation devrait se poursuivre au cours du vingt-et-unième siècle. En Europe, les vagues de chaleur devraient également se produire plus intensément et plus fréquemment qu'au siècle dernier (Revel et al. 2012). Dans les zones urbaines, l'intensité de ces vagues de chaleur pourraient être amplifiée. Depuis plus de deux siècles, l'Homme a en effet mis en évidence que l'urbanisation, caractérisée par la densification des activités humaines et l'artificialisation des sols, était à l'origine de l'augmentation de la température de l'air en ville. En 1800, Howard (1818) mesurait déjà un écart de l'ordre de 1,5 °C entre le centre-ville de Londres et la campagne environnante. Cette différence de température est aujourd'hui qualifiée d'îlot de chaleur urbain. La température de l'air n'est pas la seule variable climatique modifiée par l'urbanisation : l'humidité de l'air, les conditions de vent, les régimes de précipitations sont également altérés (Oke 1987). Le développement de tels microclimats est dû à la morphologie particulière des espaces urbains (la taille et l'espacement des objets urbains - bâtiments, ponts, etc.) mais aussi de la nature particulière de leurs sols qui sont souvent des matériaux imperméables. Or, en Europe, les villes concentrent les trois quarts de la population et la plupart des activités socio-économiques. Les effets du changement climatique, conjointement aux modifications climatiques induites localement par l'urbanisation peuvent donc avoir dans le siècle à venir des impacts sur notre société : des impacts sanitaires lors d'évènements intenses ponctuels tels que les canicules (Champiat 2009 ; Scherer et al. 2014) mais également des impacts sur le confort extérieur (Blazejczyk et al. 2013 ; Höpfe 1999), sur les consommations d'énergie (Gros 2013 ; Malys 2012) et également sur la richesse, la densité et le comportement de la faune et de la flore (Jochner et Menzel 2015 ; Wilby et Perry 2006).

Les pouvoirs publics s'assurent à travers une approche multi-objectifs (objectifs qui peuvent être liés à la mobilité, à la mixité sociale, à l'accès aux soins ou à la culture, etc.) de

la bonne organisation du territoire et du bien-être de leurs citoyens. À l'échelle d'une ville ou d'un groupement de communes, les collectivités locales ont par exemple besoin d'identifier et de limiter les conséquences potentiellement dommageables du changement climatique : on parlera de politiques d'adaptation au changement climatique. Des outils d'aide à la décision sont mis en oeuvre pour diagnostiquer la vulnérabilité d'un territoire au changement climatique et éventuellement proposer des leviers d'action pour diminuer la vulnérabilité de certains quartiers. La géographie est une discipline clé pour appréhender le territoire de façon systémique, interdisciplinaire. L'enjeu des recherches menées sur la ville est donc de fournir aux collectivités des méthodologies qui permettent d'appréhender des phénomènes complexes à travers une représentation spatiale simplifiée du territoire. Il peut s'agir de cartes mettant en évidence les différences spatiales de périodes d'émission des pollens, de températures maximales atteintes lors d'une vague de chaleur, ou encore d'économies ou de surconsommations d'énergie engendrées par l'îlot de chaleur urbain. Ces cartes peuvent ensuite être croisées avec des données socio-économiques et donner lieu à un diagnostic et à un plan d'actions de la collectivité. Par exemple : si un quartier où la population est particulièrement âgée est diagnostiqué comme très chaud en été du fait d'une forte proportion de surfaces minéralisées, il pourra faire l'objet d'un projet de végétalisation afin d'y limiter le réchauffement estival.

Ce travail de recherche s'intéresse à la distribution spatiale et temporelle de la température de l'air dans une aire urbaine dans l'optique de fournir de tels éléments d'aide à la décision. Il s'inscrit dans le cadre du projet de recherche URBIO <sup>1</sup>, s'intéressant entre autres aux interactions existantes entre biodiversité et climat urbain. La température de l'air est fortement liée à la quantité d'énergie solaire reçue à la surface terrestre, qui, en un point donné, varie au cours d'une journée, des saisons et des années. Il est ainsi nécessaire de relever la température de l'air suffisamment fréquemment afin de suivre ses variations au cours d'une journée. Spatialement, la température n'est également pas constante : sa mesure en un point donne donc une information spatiale ponctuelle. Certaines zones de l'espace sont néanmoins relativement homogènes d'un point de vue climatique. Une mesure réalisée au sein d'un tel espace est donc qualifiée de représentative de celui-ci : la valeur mesurée peut être allouée à l'ensemble de la zone et perd donc son caractère ponctuel.

Depuis quelques années, les objets connectés ont permis à des plates-formes de science participative de voir le jour. Dans le domaine de la microclimatologie urbaine, les capteurs de température extérieure, utilisés pour la régulation des équipements de chauffage des bâtiments

---

<sup>1</sup> <http://www.irstv.fr/fr/recherche/contrats-de-recherche/187-urbio> [Accédé le 02/11/2016]

(par exemple la plateforme NetAtmo<sup>2</sup>), peuvent ainsi être utilisés pour étudier l'îlot de chaleur urbain (Meier et al. 2015). Aucune métadonnée sur la position du capteur n'est cependant associée à ces données, ou alors celles-ci ne sont pas mises à disposition du public. Il est donc impossible d'évaluer la taille de la zone représentative de la valeur mesurée par le capteur. Les capteurs de température peuvent en effet être disposés à l'intérieur de zones climatiquement très hétérogènes (proche d'un arbre, d'un mur, d'une sortie de ventilation d'un bâtiment, etc.) et donc être représentatives de conditions très particulières. Un volume important de données, associé à des méthodes de pré-traitements, doivent donc être utilisés afin de faire ressortir du sens de tels jeux de données. Des organismes institutionnels (collectivités territoriales, observatoires de recherche) disposent également de réseaux de capteurs afin d'observer (entre autres) le climat urbain sur de longues périodes de temps. Les capteurs sont installés sur l'espace public ou dans des espaces privés. Le financement et la maintenance de tels réseaux sont lourds, limitant souvent à une vingtaine le nombre de capteurs utilisés. Ces réseaux ne couvrent donc jamais entièrement le territoire des villes et l'un des enjeux est de choisir des emplacements qui représentent l'espace urbain dans toute sa diversité. Dans ce cas, la qualité des données est appréciée au détriment de la quantité : le matériel est connu, son emplacement également.

Des outils de modélisation du climat urbain permettent de simuler plutôt que de mesurer les variations spatiales et temporelles de températures à l'échelle urbaine. Ces outils utilisent des méthodes numériques pour résoudre des équations de la physique, des informations géographiques (telles que la densité moyenne de bâtiments par quartier, la hauteur moyenne, etc.), et également des données météorologiques. Les données de mesures (campagnes de mesures ponctuelles ou jeux de données précédemment décrits) sont utilisées pour valider leur bon fonctionnement. Une fois validés, ces modèles présentent l'avantage par rapport aux mesures de pouvoir prévoir l'effet de modifications urbaines sur le microclimat. Ils peuvent donc être utilisés tant pour diagnostiquer climatiquement un territoire que pour juger de la pertinence d'un scénario de rénovation urbaine plutôt qu'un autre. Leur mise en oeuvre est cependant complexe pour un utilisateur non-averti : des compétences solides relatives à la physique de l'atmosphère et aux méthodes numériques sont nécessaires pour faire fonctionner correctement une simulation, compétences qui ne sont pas toujours disponibles dans les services des collectivités.

Les interrogations des dirigeants sont souvent étroitement liées à des indicateurs géographiques du territoire. Par exemple, quel serait l'effet sur le microclimat du remplacement de plusieurs immeubles vieillissants par un terrain de football enherbé ? Les objets manipu-

---

<sup>2</sup> <https://www.netatmo.com/fr-FR/weathermap> [Accédé le 16/10/2016]

lés ici sont donc les bâtiments que l'on souhaite retirer, ainsi que la nature des sols que l'on souhaite modifier. Les informations qui leur sont propres (la taille des bâtiments, la nature ou le type d'utilisation des sols) sont stockées dans des bases de données géographiques souvent gérées par les collectivités. Relier directement la température d'un quartier à des indicateurs géographiques descriptifs des formes urbaines et de la nature des sols pourrait donc s'avérer intéressant pour les décideurs. Ce sera donc un des enjeux majeurs de cette thèse : établir des relations empiriques entre des données météorologiques mesurées sur un territoire et des informations géographiques caractérisant ce territoire. Ces relations devront permettre d'estimer les variations spatiales et temporelles de la température au sein d'une zone urbaine.

Le premier chapitre fait l'état des connaissances concernant les causes du développement d'un microclimat urbain. Les différentes méthodes qui sont utilisées pour analyser l'îlot de chaleur urbain sont également présentées. Le deuxième chapitre traite des données de mesure météorologique, notamment des emplacements utilisés pour les capteurs, qui sont répartis sur les territoires des agglomérations de Nantes, Angers et La Roche-sur-Yon. Une comparaison du type d'équipement utilisé pour mesurer la température est réalisée afin d'identifier si des équipements pourraient être à l'origine d'erreurs de mesure. Des méthodes de prévision de ces erreurs sont finalement établies afin de corriger les données considérées erronées. Le troisième chapitre est consacré à la présentation de la méthodologie utilisée pour estimer les variations spatiales et temporelles de la température de l'air au sein d'une agglomération. Un ensemble de modèles spatiaux et temporels sont calibrés à partir d'un jeu de données acquis sur le territoire de Nantes Métropole. Des variantes de la méthodologie générale sont proposées et permettent, à partir du même jeu de données, d'établir plusieurs autres ensembles de modèles. Dans le chapitre quatre, le choix de l'ensemble de modèles le plus performant pour estimer les variations spatiales et temporelles de l'ICU est identifié. Pour cela, les estimations de température réalisées à partir de chaque ensemble de modèles sont confrontées aux observations des jeux de données nantais, angevins et yonnais. Les résultats obtenus avec l'ensemble de modèles le plus convaincant sont analysés en détail. Enfin, au chapitre cinq, la méthodologie développée est illustrée dans le cadre d'exemples d'applications qui ont un lien avec la microclimatologie urbaine. Les liens qui existent entre l'ensemble des chapitres 2, 3, 4 et 5 sont synthétisés figure 1.

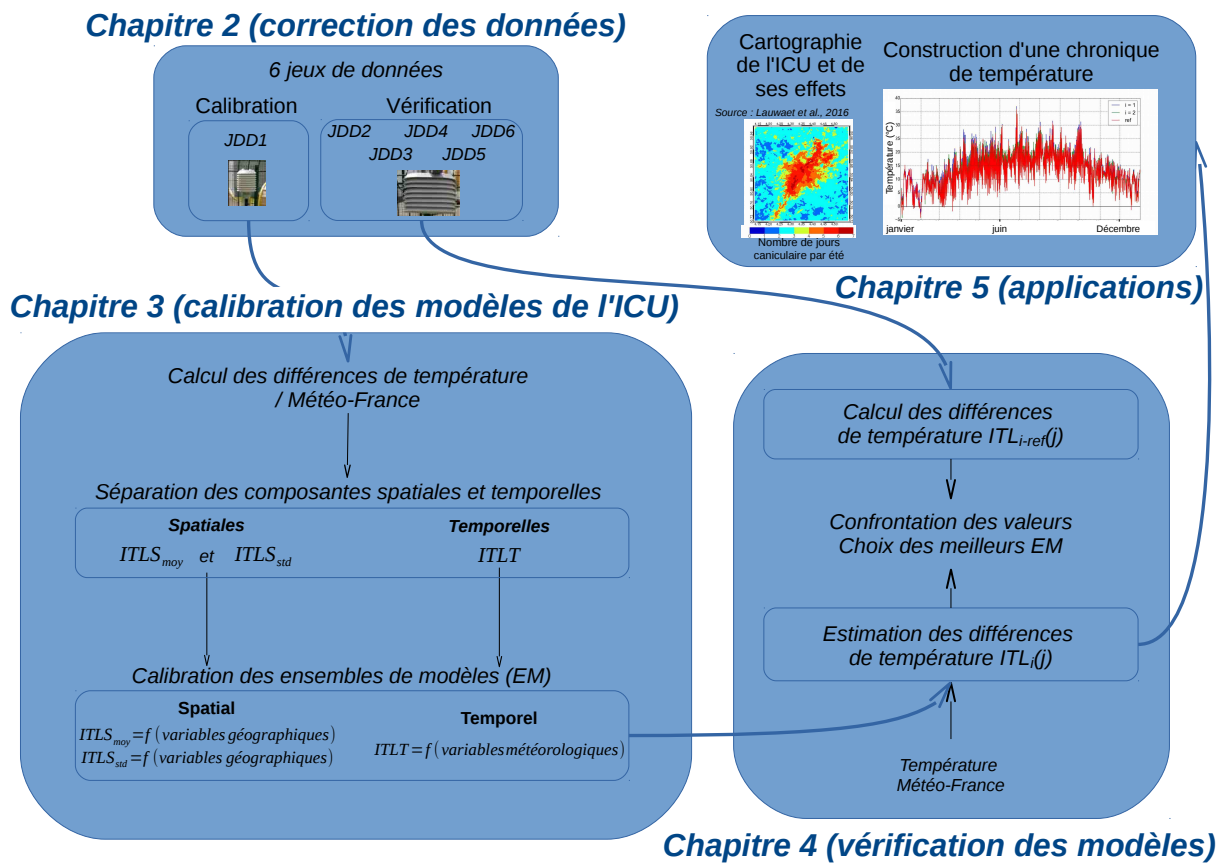


FIGURE 1 – Plan de la thèse : synthèse des interactions qui existent entre les chapitres 2, 3, 4 et 5





## Le climat urbain : état de l'art

### 1.1 Îlot de chaleur urbain : considérations générales

#### 1.1.1 Définition

##### 1.1.1.1 Îlot de Chaleur Urbain de surface, de sol et atmosphérique

Les premiers écrits qui font état de différences de température entre la ville et sa périphérie datent du XVII<sup>ème</sup> siècle. L'intérêt pour ces températures était principalement d'ordre sanitaire. L'air réputé sale et chaud des villes était accusé d'accélérer la propagation de maladies (Stewart 2011). Howard (1818) fut ensuite le premier scientifique à laisser des travaux décrivant et analysant ces différences de température. Il mesura des différences moyennes de température d'air de l'ordre de 1,5°C entre le centre-ville de Londres et sa périphérie proche (moins d'une dizaine de kilomètres). Le terme d'îlot de chaleur urbain (ICU) a été utilisé par la suite pour définir la différence de température observable entre une zone urbaine et les zones rurales environnantes.

Stewart (2011), dans le cadre de sa thèse de doctorat, a réalisé une revue chronologique de l'intérêt de l'Homme pour l'ICU depuis Howard. Il est proposé de le décliner en trois types : l'ICU atmosphérique, l'ICU de surface et l'ICU souterrain (Leconte 2014 ; Stewart 2011).

L'ICU atmosphérique est lui-même dissocié en deux sous-catégories par Oke (1987) selon la couche étudiée (ces couches sont représentées figure 1.1).

- Couche limite de surface (CLS) : c'est la couche la plus proche du sol. Elle inclut la canopée urbaine, où est mesuré l'ICU décrit par Howard (1818). C'est l'ICU le plus documenté et les causes de sa formation seront décrites par la suite dans le paragraphe 1.1.2.1.
- Couche de mélange : en l'absence de vent, l'ICU se présente sous la forme d'un dôme de chaleur au-dessus de la ville. Selon Oke (1987), les principales hypothèses expliquant sa formation sont la présence de l'ICU de CLS, les émissions de chaleur anthropiques par les toits (cheminées), des échanges de chaleur sensible plus importants avec la canopée urbaine par phénomènes de turbulence et une absorption importante du rayonnement courte longueur d'onde par les particules fines résultantes de la pollution atmosphérique.

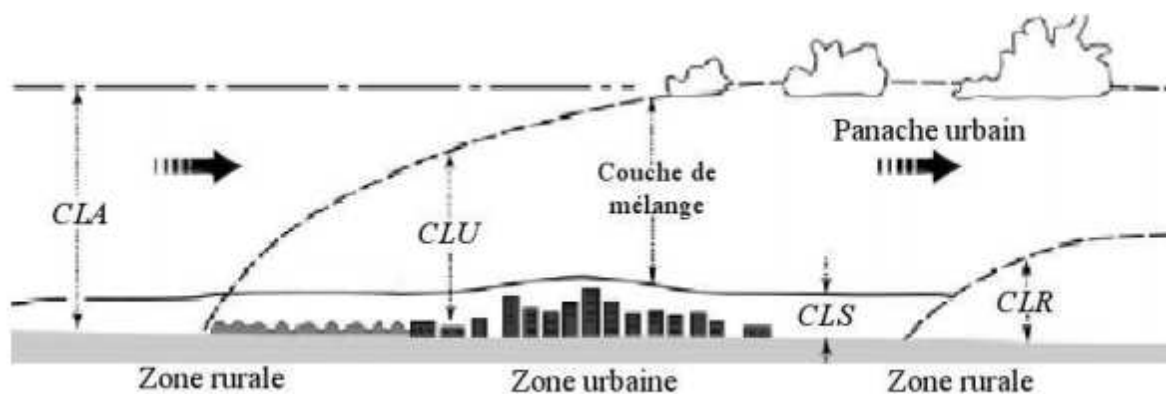


FIGURE 1.1 – Représentation des principales couches contenues dans la couche limite atmosphérique (CLA) : couche limite rurale (CLR) et couche limite urbaine (CLU), elle-même composée d'une couche limite de surface (CLS), d'une couche limite de canopée urbaine (CU) et d'une couche de mélange. Source : adaptation de Oke (2004) par Bouyer (2009)

L'ICU de surface (principalement des routes et des toits pour la zone urbaine et des feuilles d'arbres ou d'herbacés pour les surfaces rurales) est attribué aux différences de caractéristiques thermiques de matériaux (albédo, effusivité, etc.).

L'ICU souterrain est sans doute le moins étudié de tous les types d'ICU (Taniguchi et al. 2008). Il altère pourtant la composition des eaux souterraines en accélérant les réactions géochimiques et géobiologiques (Knorr et al. 2005). Sa formation découle en fait principalement de l'ICU de CLS.

### 1.1.1.2 Positionnement de la thèse

L'objectif de cette thèse est de modéliser les différences de température présentes entre les différents quartiers d'une zone urbaine dans la CLS. Cet espace, comprenant la canopée urbaine, est le lieu de vie des citoyens et le principal lieu de développement de la faune et de la flore. Le terme ICU fera donc référence dans la suite de la thèse seulement à l'ICU de CLS, seule couche à laquelle nous nous intéresserons par la suite.

L'échelle verticale d'analyse étant adoptée, il s'agit maintenant de fixer l'échelle horizontale et l'échelle temporelle.

## 1.1.2 Échelles d'analyse

### 1.1.2.1 Causes

Connaître les causes du développement de l'ICU permet de mieux comprendre les raisons de son analyse à diverses échelles spatiales et temporelles. Son origine est attribuée aux particularités morphologiques des zones urbaines, mais aussi aux types de surfaces et aux activités humaines (Grimmond et al. 2009 ; Oke 1987). Celles-ci modifient les différentes contributions au bilan d'énergie.

- La géométrie des rues, que l'on assimile souvent à des canyons, augmente la surface exposée au rayonnement solaire, diminue le facteur de vue du ciel, et réduit les vitesses de vent. Les conséquences sont respectivement une augmentation de l'absorption de rayonnement courtes longueurs d'ondes, une diminution des émissions de rayonnement grandes longueurs d'ondes vers le ciel et une diminution également de la quantité de chaleur évacuée par turbulence.
- La modification des sols en ville a également une influence sur leur bilan énergétique et hydrique. Les matériaux utilisés pour la construction sont souvent imperméables et l'eau est évacuée plutôt qu'évaporée. Pour une même quantité d'énergie incidente, un sol urbain aura donc par rapport à un sol naturel un déficit de chaleur latente qu'il compensera par un surplus de chaleur sensible. Ces mêmes matériaux ont souvent également une plus grande inertie thermique, notamment une effusivité thermique importante. Ceci leur permet d'emmagasiner au cours d'une journée plus de chaleur que la plupart des sols non-urbains.

- Les activités humaines sont souvent plus denses en ville qu'en périphérie. L'énergie dissipée par les transports mais également par les bâtiments contribue à augmenter le flux de chaleur anthropique rejeté dans l'atmosphère.

### 1.1.2.2 Échelles spatiales

L'ensemble des modifications du bilan énergétique précédemment décrites engendrent des différences horizontales de température d'air qui peuvent être considérées selon plusieurs échelles spatiales. Oke (1987) en définit trois : la méso-échelle, l'échelle locale et la micro-échelle, qui sont illustrées figure 1.2.

La méso-échelle correspond à des phénomènes de taille caractéristique comprise entre 10 et 200 km. Elle est donc particulièrement adaptée pour comparer la différence de température entre une zone urbaine et sa campagne environnante comme l'a par exemple fait Howard (1818).

L'échelle locale correspond à des phénomènes de taille caractéristique allant de 100 m à 50 km. Les agglomérations sont constituées de différents types de tissus urbains, chacun ayant des caractéristiques morphologiques, des types de sols et des activités humaines qui lui sont propres. Ces différences se répercutent sur le bilan énergétique et font donc apparaître des différences de température au sein d'une même zone urbaine (Stewart 2011). Dans la partie basse de cette échelle, il est donc nécessaire de ne plus considérer la tâche urbaine comme une seule entité homogène mais comme plusieurs entités altérant différemment le climat local.

La micro-échelle ou échelle microclimatique, concerne les phénomènes allant d'une taille de 1 cm à 1 km. Plus un volume d'air et une surface sont proches, plus leur température sont corrélées. Deux masses d'air situées à quelques dizaines de mètres peuvent ainsi avoir des températures très différentes selon le type de surface qui entoure chacune d'elles. Des différences de température apparaissent donc au sein même d'un quartier ou d'une rue, impactant directement les besoins de chaleur ou de froid d'un bâtiment selon son positionnement dans le quartier (Gros 2013 ; Malys 2012).

### 1.1.2.3 Échelles temporelles

De même que pour les variations spatiales, plusieurs échelles temporelles sont utilisées pour analyser le climat urbain. L'ICU d'un quartier évolue au fil d'une journée, entre deux journées consécutives, au cours des saisons et également au cours des années.

Au fil d'une journée et entre deux journées, la variation de la valeur de l'ICU enregistrée dans un même quartier est attribuée aux conditions météorologiques ou aux activités

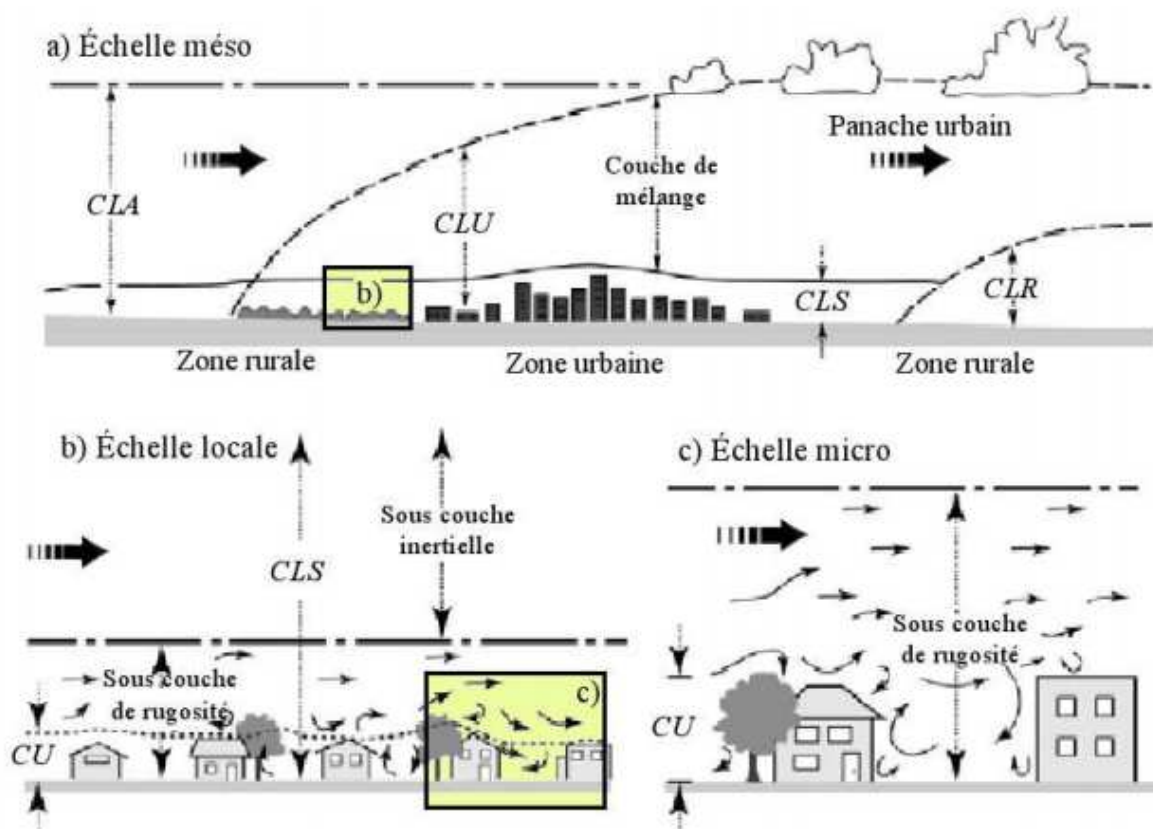


FIGURE 1.2 – Imbrication des échelles spatiales de la climatologie urbaine (couches contenues dans la couche limite atmosphérique (CLA) : couche limite rurale (CLR) et couche limite urbaine (CLU), elle-même composée d’une couche limite de surface (CLS), d’une couche limite de canopée urbaine (CU) et d’une couche de mélange). Source : adaptation de Oke (2004) par Bouyer (2009)

humaines, variables qui évoluent au cours du temps. En effet, une augmentation de la vitesse du vent et une diminution du rayonnement solaire peuvent sous certaines conditions réduire l’ICU à néant en quelques heures. De plus, aux heures où la circulation est la plus dense, certains quartiers situés sur des tronçons majeurs du réseau routiers peuvent voir leur température augmenter notablement (Leconte 2014). Le démarrage simultané de plusieurs systèmes de climatisation peut également être à l’origine des variations temporelles de l’ICU d’un quartier.

Les caractéristiques thermiques des sols végétalisés, le piégeage radiatif et la chaleur émise par les bâtiments évoluent au cours des saisons, impactant directement l’ICU.

- Le rôle rafraîchissant de la végétation est limité aux périodes où la végétation peut transpirer, c’est-à-dire pour de nombreuses espèces le printemps et l’été.
- Le piégeage radiatif engendré par les formes urbaines dépend entre autres de l’inclinaison des rayons du soleil. Cette inclinaison évolue au cours de l’année, ainsi que la quantité d’énergie emmagasinée par les surfaces terrestres pendant la période diurne.

- Les bâtiments situés dans des zones climatiques dont l'hiver est rude émettront souvent plus de chaleur en hiver que pendant les autres saisons.

Chacune de ces modifications de flux d'énergie est donc à l'origine des variations de l'ICU au cours de l'année.

La majorité des zones urbaines évoluent au fil des années. Le type d'activité évolue, la population augmente ou diminue, des changements d'affectation des sols sont opérés, des quartiers entiers sont rénovés voire reconstruits, etc. Ces modifications altèrent la valeur de l'ICU des quartiers où elles ont lieu, mais également celle des quartiers proches. La modification du climat à l'échelle de la planète impacte les conditions climatiques de la méso-échelle telle que définie par Oke (1987). Par exemple, les événements caniculaires seront plus fréquents et plus intenses dans les décennies à venir. L'ICU étant souvent maximum pendant ce type d'évènement, on peut s'attendre à ce que les valeurs moyennes et maximales de l'ICU augmentent de manière tendancielle. Au final, un quartier verra son ICU évoluer au cours des années selon les modifications urbaines qu'il subira mais également selon l'intensité du changement climatique.

#### **1.1.2.4 Positionnement de la thèse**

Les approches utilisées pour modéliser la température de l'air au sein d'une agglomération diffèrent selon l'échelle considérée. Notre travail se concentrera sur l'échelle locale : ce sont les différences de température d'air entre quartiers d'une zone urbaine qui nous intéressent. Les deux autres échelles seront néanmoins considérées, notamment la micro-échelle au chapitre 2. Les effets du changement climatique sur les valeurs de l'ICU ne seront pas directement pris en considération dans ce travail. Cependant, d'une certaine manière, toutes les échelles temporelles pourront être traitées à partir du travail effectué.

L'échelle spatiale qui nous intéresse est maintenant fixée verticalement (air situé dans la canopée) et horizontalement (différences de température entre quartiers d'une agglomération). Il s'agit maintenant de discuter des méthodes qui existent pour analyser les écarts de température à cette échelle, en considérant l'ensemble des échelles temporelles.

#### **1.1.3 Méthodes d'analyse**

Il existe plusieurs méthodes pour mettre en évidence des différences spatiales de température au sein d'une zone urbaine (à l'échelle locale). Elles sont ici regroupées en quatre types : les méthodes basées sur des modèles de connaissance (MC), celles basées sur des modèles de ré-

gression géoclimatique (RG), les méthodes de classification géoclimatique (CG) et celles d'interpolation géostatistique (IG). Des revues étoffées de la littérature font état de l'avancement de la recherche pour les deux premières méthodes (Grimmond et al. (2009) se concentre sur la première tandis que McKendry (2003) détaille les deux). La revue présentée ici, plus synthétique, est basée entre autres sur le travail bibliographique de ces deux auteurs.

### 1.1.3.1 Simulation numérique basée sur des modèles de connaissances

Les modèles de simulation numérique du climat urbain à l'échelle locale reposent sur le bilan d'énergie de volumes urbains de la CLS, mailles élémentaires de ces modèles et pouvant contenir une à plusieurs rues. L'équation 1.1 formalise ce bilan, présenté sous forme schématique sur la figure 1.3.

$$Q^* + Q_F = Q_H + Q_E + \Delta Q_S + \Delta Q_A \quad (1.1)$$

avec  $Q^*$  le flux radiatif net,  $Q_F$  le flux de chaleur anthropique,  $Q_H$  le flux de chaleur sensible,  $Q_E$  le flux de chaleur latente,  $\Delta Q_S$  le terme de stockage dans les matériaux (sols, murs, toits),  $\Delta Q_A$  le terme d'advection horizontale.

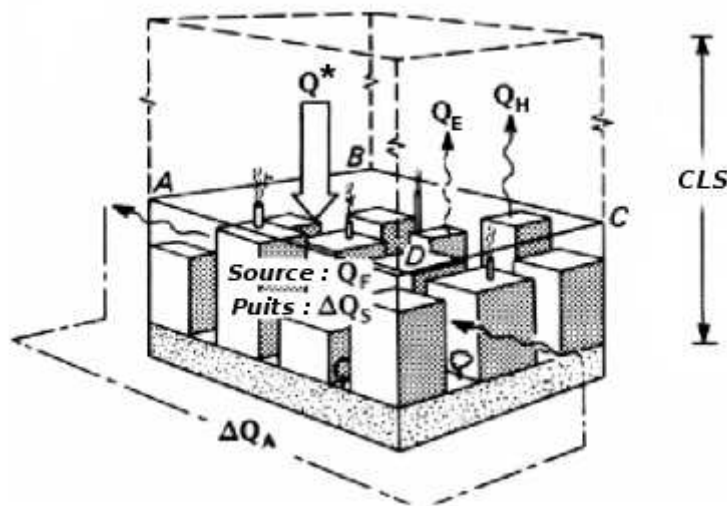


FIGURE 1.3 – Représentation schématique du bilan d'énergie d'un volume d'air de la couche limite de surface. Source : adapté de Oke (1987)

Pour réaliser ce bilan, il est nécessaire de connaître les conditions météorologiques extérieures à ce volume, notamment la température, l'humidité de l'air et la vitesse de vent dans la sous-couche inertielle (représentée figure 1.2). Ces données sont souvent issues d'une station d'observation rurale (proche de la ville) et utilisées en entrée du modèle. Des informations sur



la forme et les caractéristiques des matériaux qui constituent la ville sont également nécessaires. La représentation de la ville ainsi que le calcul des flux à l'intérieur des volumes élémentaires peuvent être traités différemment selon les modèles. Grimmond et al. (2010) en ont étudiés 33 : ils ont identifié les modes de représentation et les modes de calcul qui les différencient et ainsi établi 11 caractéristiques permettant de les classer par catégories. Ces caractéristiques de modèles, représentées figure 1.4, sont entre autres que :

- tous les flux sont modélisés ou certains sont négligés comme le flux de chaleur latente ou le flux anthropique (figure 1.4 - 1) ;
- toutes les surfaces sont considérées de la même manière ou sont différenciées selon qu'elles soient des façades, des toitures ou des sols (figure 1.4 - 5 et 5) ;
- la géométrie des tissus urbains peut être plus ou moins simplifiée : aucune prise en compte, représentée par des rues canyons sans orientation, représentée par des rues canyons orientées, représentée par des rues détaillées (prise en compte des intersections) (figure 1.4 - 6) ;
- le nombre de multi-réflexions considérées au sein de la canopée peut varier entre zéro et l'infini (figure 1.4 - 7) ;
- le terme de stockage peut être calculé de différentes manières : comme un terme résiduel, en considérant la conduction au sein des matériaux ou par des cycles d'hystérésis (figure 1.4 - 9) ;
- des bilans sont effectués pour une ou plusieurs couches, permettant potentiellement d'obtenir, pour plusieurs hauteurs, des valeurs de température d'air, de température de surface, d'humidité d'air et de vitesse de vent (figure 1.4 - 11 et 12).

L'étude des 33 modèles avait pour objectif, outre cette classification, de comparer leur précision en termes de modélisation des flux de chaleur mis en jeu lors du bilan énergétique. Pour cela, un jeu de données collectées sur un site industriel de Vancouver a été utilisé (Grimmond et Oke 2002 ; Voogt et Grimmond 2000). Ces données ont été choisies car elles avaient été utilisées de nombreuses fois pour valider les modèles de plusieurs équipes de recherche, notamment Best et al. (2006), Grimmond et Oke (2002) et Masson et al. (2002).

Un des résultats de cette comparaison est que les modèles les plus détaillés ne modélisent pas forcément mieux l'ensemble des flux mis en jeu dans le bilan énergétique d'un volume urbain. D'autres conclusions ressortent de l'analyse :

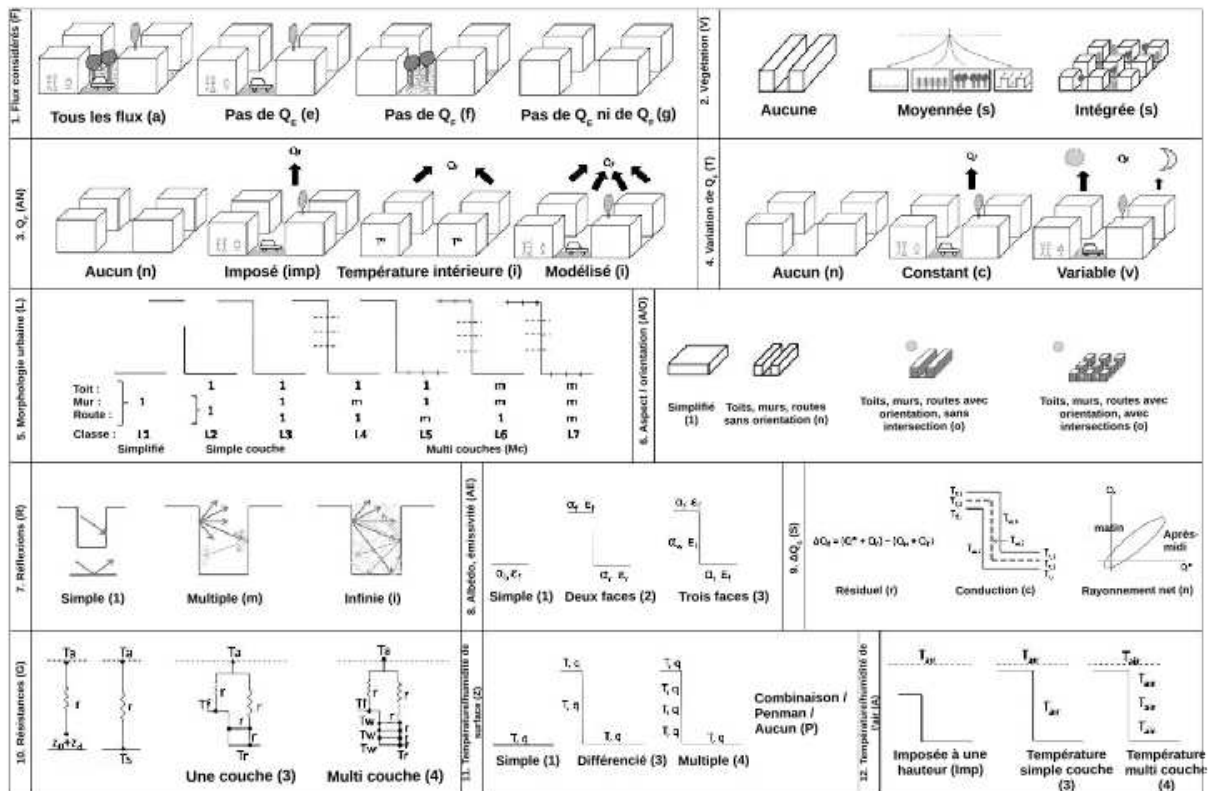


FIGURE 1.4 – Classification par caractéristiques des modèles numériques de simulation du climat urbain à l'échelle locale. Source : adapté de Grimmond et al. (2010)

- les modèles qui simulent correctement les flux la journée les modélisent souvent moins bien la nuit ;
- ne pas modéliser les flux de chaleur latente conduit de nuit comme de jour à une mauvaise estimation de tous les flux ;
- ne pas modéliser les flux de chaleur anthropique n'a que très peu d'influence la journée mais une influence plus marquée la nuit ;
- considérer l'orientation des rues (et dans une moindre mesure les intersections) permet presque toujours d'obtenir des résultats plus précis notamment la journée ;
- en journée, considérer de manière simple les caractéristiques d'albédo et d'émissivité (figure 1.4 - 8) permet de mieux estimer tous les flux.
- la méthode des résidus pour l'estimation du terme de stockage (figure 1.4 - 9) simule toujours mieux l'ensemble des flux, notamment en journée.

### 1.1.3.2 Calcul empirique par régression géoclimatique

Une deuxième manière d'étudier les variations spatiales de température d'air est d'expliquer les écarts observables entre plusieurs points d'une agglomération à partir d'une ou plusieurs variables. Des modèles statistiques de régression linéaires ou non-linéaires sont ainsi établis entre une variable à modéliser  $y$  et une ou plusieurs variables explicatives  $x_n$ .

$$y = f(x_1, x_2, \dots, x_n, e_{res}) \quad (1.2)$$

avec  $e_{res}$  les résidus, représentatifs du terme qui n'a pas pu être expliqué par le modèle.

Ce type de modèle, comparé aux MC vus en 1.1.3.1, est simple à appréhender. Pour le mettre en oeuvre, il faut un jeu de données de mesure de température et des données qui permettent de calculer les variables  $x_n$  explicatives des différences de température d'air  $y$  entre quartiers. Bien que les phénomènes physiques soient considérés lors de l'identification des variables explicatives pertinentes, ce type de modèle n'a pas vocation à calculer des flux d'énergie entre les surfaces urbaines et l'atmosphère. Il permet donc seulement d'estimer des températures ou des différences de température entre quartiers d'une agglomération.

Un des objectifs souvent recherché par les modélisateurs est d'identifier une variable explicative qui, quelle que soit la ville étudiée, permette d'estimer l'écart de température  $\Delta T$  entre le centre urbain et sa périphérie. Karl et al. (1988) ont par exemple établi une relation non-linéaire entre le nombre d'habitants  $nb_{hab}$  et la différence de température mesurée entre le centre urbain et sa campagne environnante (équation 1.3). Cette relation a été établie à partir des minimas journaliers  $min(T_k(j))_k$  relevés pour un échantillon de 305 villes et villages. La différence de température  $\Delta min(T_k(j))_k$  entre chaque centre urbain et sa campagne périphérique a été calculée pendant une année de mesure quelles que soient les conditions météorologiques. Au final, c'est la moyenne de ces différences qui est estimée à partir de la population.

$$moy(\Delta min(T_k(j))_k)_j = 0,08045 \times nb_{hab}^{0,45075} \quad (1.3)$$

Dans le même registre, Oke (1973) a tracé, pour plusieurs zones urbaines d'Amérique du nord et d'Europe, les différences de température  $\Delta T$  observées par temps clair et calme (peu ou pas de nuages ni de vent) en fonction du logarithme de la population. Il a montré que la relation qui lie ces deux variables est linéaire mais que la pente de la droite est plus forte pour les villes américaines que pour les villes européennes (figure 1.5). Ceci met en évidence le fait

que le nombre d'habitants ne peut pas à lui seul expliquer l'intensité de l'ICU.

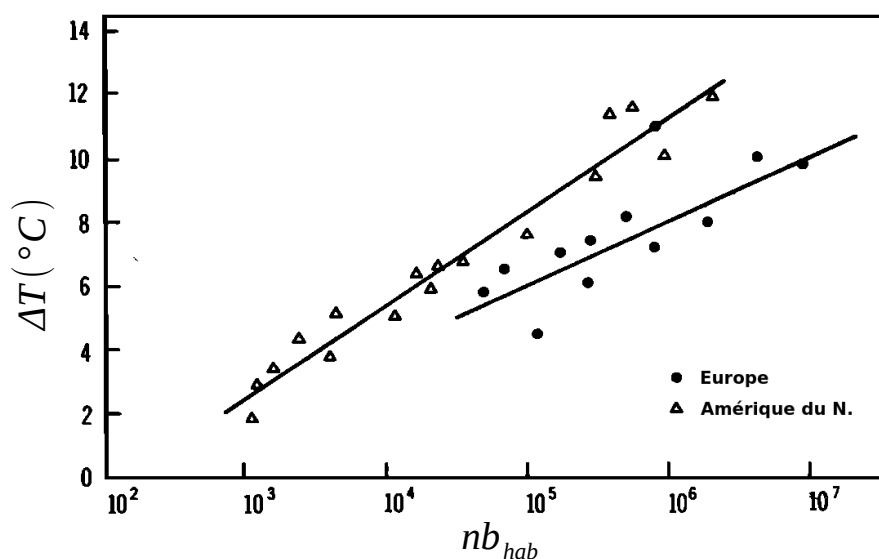


FIGURE 1.5 – Relation entre le logarithme du nombre d'habitants d'une zone urbaine et son ICU lors de conditions météorologiques claires et calmes. Source : adapté de Garstang et al. (1975)

Böhm (1998) conforte cette hypothèse en utilisant l'exemple de la ville de Vienne. Il montre que la population de Vienne est restée constante entre 1951 et 1995, et que pourtant l'ICU n'a pas cessé de croître. Dans ce cas, cette augmentation s'explique selon lui par une modification de l'utilisation des sols, une augmentation considérable des consommations d'énergie (2,5 fois plus élevées en 1995 qu'en 1951) et une augmentation de la surface habitable par personne. Ce constat laisse entrevoir une explication plus physique des relations qui existent entre variables explicatives et variable à modéliser.

Dans cet esprit, d'autres chercheurs construisent les variables explicatives de leur modèle à partir d'indicateurs physiquement liés à l'ICU. Ces indicateurs, identifiés notamment par Oke (1987) et également par Grimmond et al. (2010) et décrits en 1.1.2.1, sont principalement liés à la morphologie et à la typologie des sols urbains mais également aux activités humaines. Oke (1987) a par exemple proposé le rapport d'aspect moyen  $\frac{H}{W}$  des rues d'une agglomération (un indicateur de forme urbaine) comme variable explicative de l'ICU (équation 1.4).

$$\Delta T = 7,54 + 3,97 \times \ln\left(\frac{H}{W}\right) \quad (1.4)$$

avec  $\Delta T$  l'écart de température entre le centre urbain et sa périphérie,  $H$  la hauteur moyenne des bâtiments,  $W$  la largeur moyenne des rues.

Les modèles empiriques proposés jusqu'ici sont simples mais pauvres en informa-

tions spatiales et temporelles. Ils décrivent en effet une différence de température en distinguant seulement deux espaces (la ville et sa périphérie), alors que les MC permettent de distinguer une multitude d'espaces (mailles) au sein d'une zone urbaine. Dans le cas de RG, la différence de température est également calculée à un instant précis de la journée et est moyennée par le nombre de jours utilisés pour élaborer le modèle. La température ne peut donc pas dans ces conditions être suivie heure par heure au cours de la journée. Enfin, les jours utilisés pour élaborer le modèle peuvent être sélectionnés pour leurs conditions météorologiques particulières (calmes et claires). Il n'est donc pas possible d'estimer l'ICU quelles que soient les conditions météorologiques.

Pour pallier ces manques, d'autres modèles ont été conçus. En termes spatiaux, ils se distinguent des modèles précédents par un nombre de sites de mesure souvent supérieur à une dizaine. En termes temporels, une combinaison de plusieurs modèles est souvent utilisée pour différencier l'influence des variables explicatives de l'ICU au fil des saisons. Svensson et al. (2002) établissent ainsi quatre modèles pour représenter une année (un par saison), tandis que Suomi et al. (2012) en utilisent 12 (un par mois). Les différences de température sont parfois calculées à plusieurs instants de la journée, par exemple aux instants les plus chauds et les plus froids comme le font Petralli et al. (2014). Les conditions météorologiques régionales sont également prises en compte dans l'estimation de l'ICU. Alonso et al. (2007) considèrent par exemple le vent comme un facteur de pondération de l'ICU :

$$\Delta T = 0.122 \times U_{h=10m}^2 - 1,107 \times U_{h=10m} + 2,987 \quad (1.5)$$

avec  $\Delta T$  la différence de température de 22h à 6h entre une rue canyon et un site péri-urbain et  $U_{h=10m}$  la vitesse du vent mesurée à 10 m sur le site péri-urbain.

Comme pour les modèles simples explicités auparavant, ces nouveaux modèles utilisent en règle générale comme variables explicatives de l'ICU des indicateurs de forme urbaine ou de typologie de sols. Yokobori et Ohta (2009) relient par exemple la température moyenne mesurée pour 11 sites de mesure à leur densité de végétation à travers deux modèles (l'un diurne, l'autre nocturne). Ils mettent en évidence que plus la densité de végétation présente sur un site est élevée, plus sa température moyenne est faible (cette influence étant plus marquée la nuit que le jour - figure 1.6).

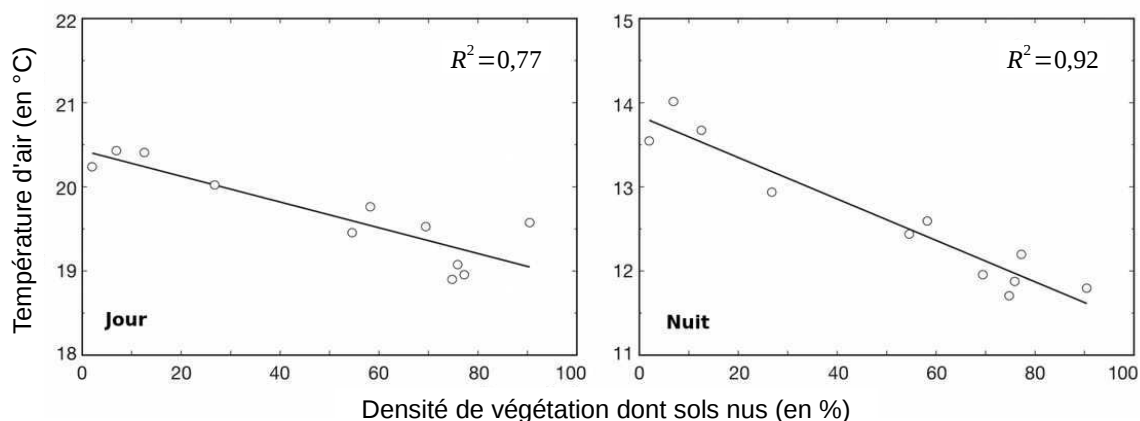


FIGURE 1.6 – Relation entre la température moyenne diurne et nocturne et la densité de végétation. Source : adapté de Yokobori et Ohta (2009)

### 1.1.3.3 Classification par caractérisation géoclimatique

Cette méthode est proche des méthodes de régression géoclimatique en ce sens qu'elle est basée sur la caractérisation géographique des différentes zones climatiques qui constituent une zone urbaine. C'est néanmoins une méthode de classification qui n'a pas pour objectif de quantifier les différences de température entre ces espaces.

Elle fait l'objet de nombreuses publications ces dernières années, suite à la proposition de Stewart (2011) de partitionner le territoire selon ses caractéristiques climatiques (concept de *local climate zones* - LCZ). Ce concept a été appliqué sur la ville de Nancy par Leconte (2014) qui décrit de la façon suivante l'approche utilisée par Stewart :

«[Elle] consiste à classer les territoires en différentes sous-classes en utilisant plusieurs paramètres. Quatre paramètres influençant le comportement thermique dans la couche limite de canopée urbaine ont été choisis afin de générer des sous-classes. Ces paramètres concernent la morphologie urbaine et l'occupation du sol :

- le type d'objet présent dans la zone (bâtiments ou plantes) ;
- la hauteur de ces objets ;
- la densité spatiale de ces objets dans la zone ;
- le caractère perméable ou imperméable des surfaces situées entre les objets de la zone.»

Une analyse de l'influence de ces quatre paramètres sur la température de l'air a permis de créer 17 types de LCZ, représentées figure 1.7.

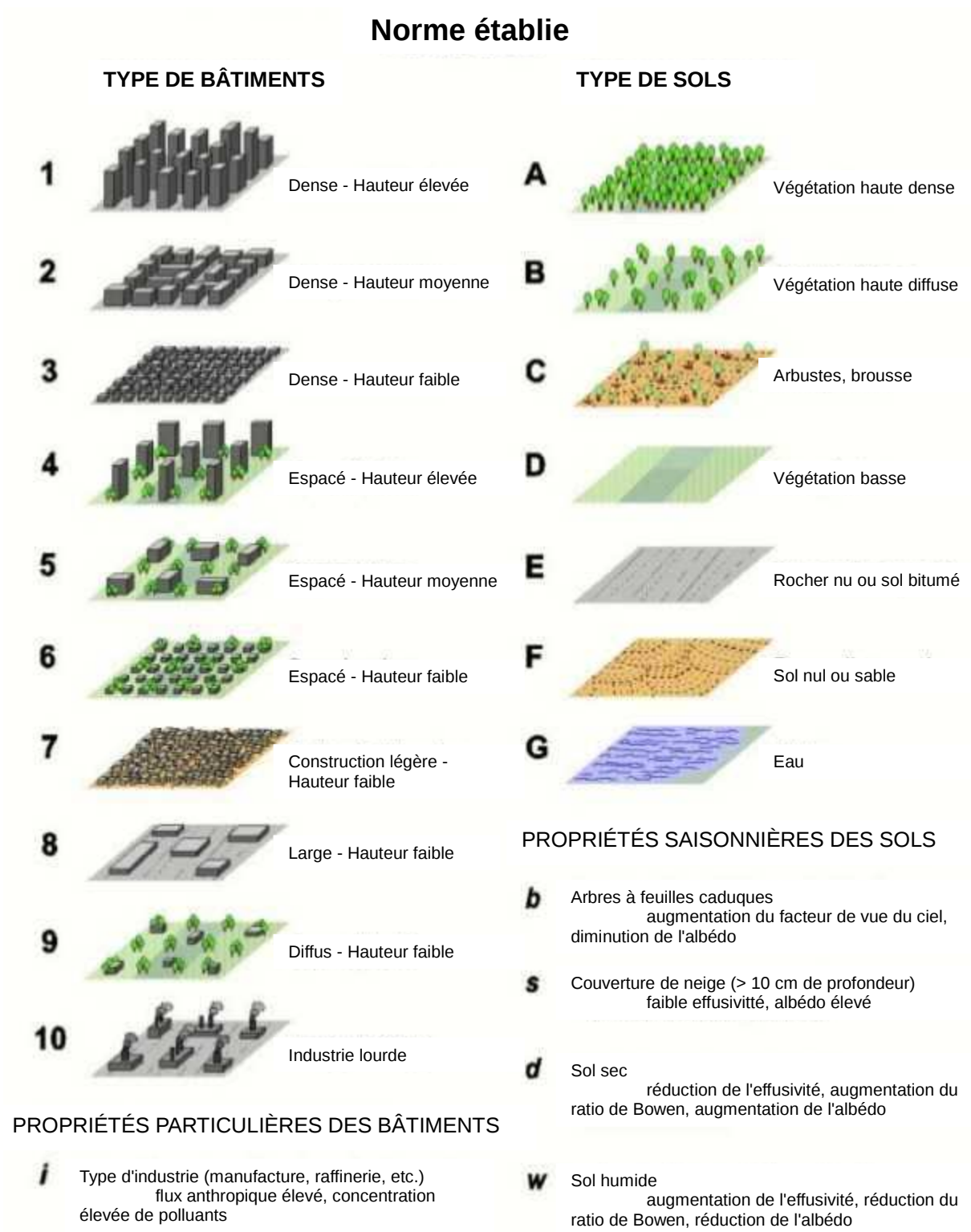


FIGURE 1.7 – Présentation des 17 LCZ existants. Il existe 10 types de LCZ urbanisés et 7 non urbanisés. Lors de l'identification, il est possible d'ajouter un indice pour préciser une propriété concernant les bâtiments ou l'occupation du sol. Source : Stewart (2011)

L'intérêt de cette classification est clair : identifier au sein d'un territoire des zones climatiquement homogènes à partir d'un jeu d'indicateurs donné. Elle permet ainsi de sortir

de la dichotomie simpliste urbain/rural en comparant les LCZ entre elles. Cependant, la mise en oeuvre opérationnelle de cette classification pour des villes européennes est une tâche complexe. En effet, les tissus urbains sont plus hétérogènes que ceux des villes nord-américaines, le maillage des rues est organisé de façon moins géométrique. L'automatisation de la classification est encore moins simple. Ainsi, Leconte (2014) a réalisé la classification LCZ de Nancy manuellement. Lelovics et al. (2014) ont proposé une solution automatique de classification pour la ville de Szeged (Hongrie), dans laquelle des choix arbitraires ont dû être effectués pour assigner une LCZ à certaines zones qui étaient trop petites pour être considérées comme des LCZ à part entière.

Cette méthode de classification peut être utilisée pour comparer plusieurs territoires puisqu'elle est extrapolable. Elle constitue également un bon support pour distinguer des zones a priori homogènes climatiquement, en vue de mettre en place un réseau de mesure (Lelovics et al. 2014). Cependant, elle ne permet pas de quantifier directement l'ICU.

#### 1.1.3.4 Calcul par interpolation géostatistique

Ce dernier moyen de cartographier les écarts de température consiste à interpoler des mesures relevées en différents points d'un territoire en utilisant des méthodes géostatistiques. Différentes techniques d'interpolation peuvent être mises en oeuvre mais toutes utilisent des variables communes :

- $nb_i$  le nombre de mesures de température  $T_i$  réparties sur le territoire où ces températures doivent être interpolées ;
- $d_i$  la distance entre un point de mesure  $i$  et un point à modéliser ;
- $\lambda_i$  le poids de la température  $T_i$  dans l'estimation de la température à modéliser  $T_m$  de telle sorte que  $T_m = \sum_{i=1}^{nb_i} \lambda_i T_i$ .

Certaines de ces techniques sont simples comme celle de la *distance inverse*. Le poids de pondération est dans ce cas calculé à partir de la relation 1.6

$$\lambda_i = \frac{\frac{1}{d_i^\gamma}}{\sum_i \frac{1}{d_i^\gamma}} \quad (1.6)$$

avec  $\gamma$  un paramètre de puissance fixé par l'utilisateur.

D'autres techniques, comme le krigeage, sont plus complexes et tiennent également compte de la distance entre les points de mesure deux-à-deux. Cette technique, bien que cou-



rante en météorologie depuis plusieurs décennies (Delhomme 1978), est assez peu utilisée en météorologie urbaine (Hinkel et al. (2003) l'utilisent pour identifier la forme de l'ICU moyen journalier dans une ville de l'Alaska). La mise en oeuvre de la technique étant gourmande en temps humain mais aussi en temps de calcul (Anderson 2002), elle est très rarement utilisée pour suivre l'évolution temporelle des écarts de température au cours d'une journée. Un second inconvénient par rapport aux modèles de régression réside dans sa mise en oeuvre. Le krigeage nécessite un réseau dense de stations de mesure de température au sein d'une zone urbaine. En effet, les variables géographiques telles que le type de sols ou la morphologie de la ville ne sont pas considérées dans les calculs, ne permettant pas d'estimer l'ICU pour n'importe quelle zone de la ville à partir de ces informations comme le permettent les modèles de régression. Szymanowski et Kryza (2009) considèrent néanmoins que le krigeage explique parfois mieux la variabilité spatiale que les méthodes basées sur des régressions géoclimatiques. Hjort et al. (2011) réfutent cette thèse en pointant du doigt que seul le nombre de sites de mesure plus important utilisé par Szymanowski et Kryza (2009) permet d'obtenir de meilleures performances. En effet, ils utilisent 206 sites, contre une vingtaine pour la plupart des études utilisant des méthodes de RG. Les méthodes géostatistiques peuvent également être utilisées en complément de modèles de régression, comme le recommandent Lelovics et al. (2014). C'est par exemple ce qu'ont fait Szymanowski et Kryza (2011), en établissant un modèle de régression multiple (avec des variables explicatives telles que la densité de surfaces artificielles), qu'ils ont couplé à une technique de krigeage des résidus (le terme  $e_{res}$  de l'équation 1.2).

### 1.1.3.5 Positionnement de la thèse

Les principales caractéristiques de chacune des quatre méthodes décrites précédemment sont rassemblées dans le tableau 1.1.

La méthode d'interpolation géostatistique nécessite entre autres un réseau de capteurs dense pour cartographier spatialement l'ICU. La méthode de modélisation par modèle de connaissance permet de connaître finement (tant spatialement que temporellement) les différences de température au sein d'une agglomération. C'est une méthode qui ne dépend pas de la présence d'un réseau de capteurs pour fonctionner mais qui est de nature assez complexe (basée sur des modèles physiques détaillés) et qui nécessite donc souvent une expertise pour faire fonctionner correctement les modèles. La méthode de modélisation par modèle de régression géoclimatique permet d'estimer plus simplement des différences de température au sein d'un territoire car elles sont directement reliées à des indicateurs géographiques (notés

Nom de la méthode	Description	Besoins en termes de mesure de température	Résolution spatiale horizontale	Résolution temporelle infra journalière
Modèle de connaissance MC	Modèle reposant sur le bilan énergétique d'un volume urbain	Un site représentatif de la température de la couche de mélange	Mailles carrées de taille variable	Souvent horaire
Régression géoclimatique RG	Modélisation basée sur des régressions entre des mesures de température et des indicateurs géographiques	Un site extérieur à la ville - plusieurs sites pour établir et vérifier les modèles	Mailles et formes variables	Température minimale ou maximale journalière moyennée par mois ou par année
Classification géoclimatique CG	Classification basée sur la caractérisation des tissus urbains par des indicateurs géographiques et des caractéristiques climatiques	Aucun site de mesure n'est nécessaire (la classification peut être seulement qualitative)	Mailles et formes variables	pas nécessairement de modélisation de la température au cours du temps
Interpolation géostatistique IG	Cartographie basée sur l'interpolation géostatistique de mesures de température	Plusieurs sites à chaque instant $t$ à cartographier	Modélisation spatiale continue des températures	Souvent température minimale ou maximale journalière moyennée par mois ou par année

TABLE 1.1 – Principales caractéristiques des méthodes d'analyse de l'ICU

$IG_g$ ). Elle est également potentiellement extrapolable à d'autres agglomérations sans réaliser de nouvelles campagnes de mesure, même si peu d'études ont vérifié les résultats de cette extrapolation (Balázs et al. 2009 ; Chen et al. 2012 ; Hjort et al. 2011). Cette méthode présente néanmoins l'inconvénient d'avoir une résolution temporelle bien moindre que celle des mo-

dèles de connaissance. La méthode de classification géoclimatique, quant à elle, a connu un réel essor ces dernières années. C'est une approche intéressante qui peut être extrapolée à plusieurs territoires en vue de leur comparaison. Cependant, la quantification de l'ICU reste toujours dépendante de la mise en place d'un réseau de capteurs de mesure.

L'objet de ce travail est d'utiliser un réseau de capteurs de mesure de température pour développer des modèles simples entre températures et indicateurs géographiques. L'objectif sous-jacent est que cette méthode soit utilisable pour n'importe quelle agglomération sans avoir à déployer un réseau de capteurs de mesure de température sur le territoire. Nous utiliserons donc la méthode de modélisation par régression géoclimatique. La mise en oeuvre de cette méthode sera décrite plus en détail en [1.2](#).

## **1.2 Îlot de chaleur urbain : modélisation par régression géoclimatique**

La méthode de modélisation de l'ICU par des modèles de régression géoclimatique a été brièvement décrite en [1.1.3.2](#). Cette nouvelle section est consacrée à décrire plus finement les enjeux liés à l'élaboration de tels modèles, notamment l'importance de la qualité des mesures et la prise en considération des variations spatiales et temporelles de l'ICU. Cette section sera également consacrée aux méthodes statistiques utilisées pour élaborer des modèles reliant géographie et microclimat et également pour les vérifier. Elle repose sur l'analyse de 17 études utilisant des méthodes de RG, que l'on considère représentatives de la production dans ce champ de recherche.

### **1.2.1 Acquisition des données météorologiques**

Les modèles établis par la méthode de RG sont calibrés à partir de données de mesure de température d'air. Il est donc tout à fait légitime de s'interroger sur la qualité des mesures avant d'utiliser ces données. Celle-ci repose sur le respect de trois principes fondamentaux :

- les capteurs utilisés mesurent bien la température de l'air de l'endroit où ils sont placés ;
- l'endroit où est placé le capteur est bien représentatif thermiquement de l'air d'un site dans sa globalité, et non d'un endroit du site en particulier ;

- l'ensemble des capteurs et donc des sites de mesure répartis sur la zone urbaine permet bien de mettre en évidence les particularités thermiques spatiales et temporelles de l'ICU.

Chacun de ces points sera traité respectivement dans les sous-sections [1.2.1.1](#), [1.2.1.2](#) et [1.2.1.3](#).

### 1.2.1.1 Caractéristiques du matériel de mesure

Pour mesurer une température d'air, deux types de capteurs sont majoritairement utilisés, reposant sur deux principes physiques différents. Le premier, appelé thermocouple, repose sur l'utilisation de l'effet Seebeck : une différence de potentiel apparaît à la jonction de deux matériaux soumis à une différence de température. Le second, thermomètre à résistance, repose sur la variation de la résistance électrique d'un matériau en fonction de la température. Il existe deux types de thermomètre à résistance.

- Les thermo-résistances qui utilisent des métaux purs dont la résistivité varie linéairement avec la variation de température (platine, titane, cuivre, etc.) : c'est le cas notamment des PT100 qui sont des thermomètres à résistance de platine.
- Les thermistances, utilisant d'autres matériaux dont la résistance varie non-linéairement avec la variation de température (oxydes métalliques, composites). Les thermistances mesurent dans une gamme de température souvent plus réduite que les thermo-résistances, mais bénéficient d'une meilleure précision.

La principale caractéristique qui différencie les thermocouples des thermomètres à résistance est leur temps de réponse. Lorsque l'air qui entoure le capteur change brutalement de température, le capteur ne détecte pas instantanément cette modification. Le déphasage entre l'instant où la température change et celui où ce changement est détecté est en fait représentatif du temps nécessaire au capteur pour se mettre à température. Pour synthétiser, à propriétés thermiques équivalentes, plus la taille et donc l'inertie thermique du capteur sont importantes, plus son temps de réponse est grand. La vitesse de l'air en contact avec le capteur agit de façon notable sur ce temps de réponse. En général, on considère que le temps de réponse d'un thermocouple est de l'ordre de la seconde tandis que celui d'un thermomètre à résistance est de l'ordre de la minute <sup>1</sup>. Cette différence a une influence sur la méthode de mesure employée et sera discutée en [1.2.1.3](#).

---

<sup>1</sup> [http://meas-spec.com.cn/downloads/Temperature\\_Sensor\\_Advantages.pdf](http://meas-spec.com.cn/downloads/Temperature_Sensor_Advantages.pdf) [accédé le 11/12/2015]

La précision des thermocouples et des thermomètres à résistance est variable et diffère selon les technologies utilisées. Bien qu'importante pour de nombreuses applications industrielles, cette information est reléguée au second plan dans le cas de mesures de température d'air extérieur. En effet, la première des préoccupations est de s'assurer que le capteur ne subit pas d'interférence d'extérieure lorsqu'il mesure la température de l'air. Si tel est le cas, la température du capteur est égale à celle de l'air. La variation de température d'un capteur au cours du temps dépend en fait de son bilan énergétique, qui se traduit par l'équation 1.7 (la température sera supposée homogène dans le capteur pour n'importe quel instant  $t$ ).

$$m_c \cdot C_{pc} \cdot \frac{dT_c}{dt} = Q_c^* + Q_{H_c} \quad (1.7)$$

avec  $m_c$  la masse du capteur,  $C_{pc}$  la capacité thermique massique du capteur,  $T_c$  la température du capteur,  $Q_c^*$  le flux radiatif net du capteur et  $Q_{H_c}$  le flux de chaleur sensible échangé entre la surface du capteur et l'air.

En considérant un régime établi où la température du capteur ne change pas ( $\frac{dT_c}{dt} = 0$ ) et en définissant  $Q_{H_c} = h \cdot (T - T_c)$ , l'équation se réduit à :

$$T - T_c = \frac{Q_c^*}{h_c} \quad (1.8)$$

avec  $h_c$  le coefficient de convection entre la surface du capteur et l'air (en  $W.K^{-1}$ ) et  $T$  la température de l'air.

On note  $e_T = T - T_c$  l'erreur de mesure de la température de l'air. Cette erreur est bien souvent supérieure à la précision des capteurs (Barnett et al. 1998a), notamment dans des conditions de fort ensoleillement (Lacombe et al. 2011). En effet, elle peut parfois atteindre quelques degrés Celsius alors que la résolution des capteurs de température est généralement de l'ordre du dixième de degré celsius. Pour minimiser cette erreur, il faut donc minimiser le rapport  $\frac{Q_c^*}{h_c}$ . Ce travail peut être réalisé en minimisant  $Q_c^*$  ou en maximisant  $h_c$  (il est bien sûr possible d'agir sur les deux termes).

- Minimiser  $Q_c^*$  peut se faire de deux façons. La première consiste à protéger les surfaces du capteur du rayonnement extérieur en utilisant un abri. Cet abri doit empêcher le soleil de passer mais ne doit pas empêcher l'air de se renouveler sous risque de surchauffe. Des abris à ailettes sont donc souvent utilisés (figures 1.8a et 1.8b). La deuxième consiste à diminuer la taille du capteur afin de limiter le rayonnement qu'il reçoit (souvent en utilisant des thermocouples de petite taille - discuté au chapitre 2).

- Maximiser  $h_c$  peut également se faire de deux façons. La première est d'augmenter la convection naturelle entre l'air et la surface du capteur. Pour cela il est hélas nécessaire d'augmenter la différence de température entre la surface du capteur et l'air, différence que l'on cherche justement à minimiser. La seconde est d'augmenter la convection forcée, ce qui peut être réalisé en augmentant la vitesse de l'air qui entre en contact avec le capteur. Cette solution est souvent utilisée couplée à un abri. L'abri est dit à *ventilation mécanique*, à opposer à l'abri dit à *ventilation naturelle* (Figure 1.8).

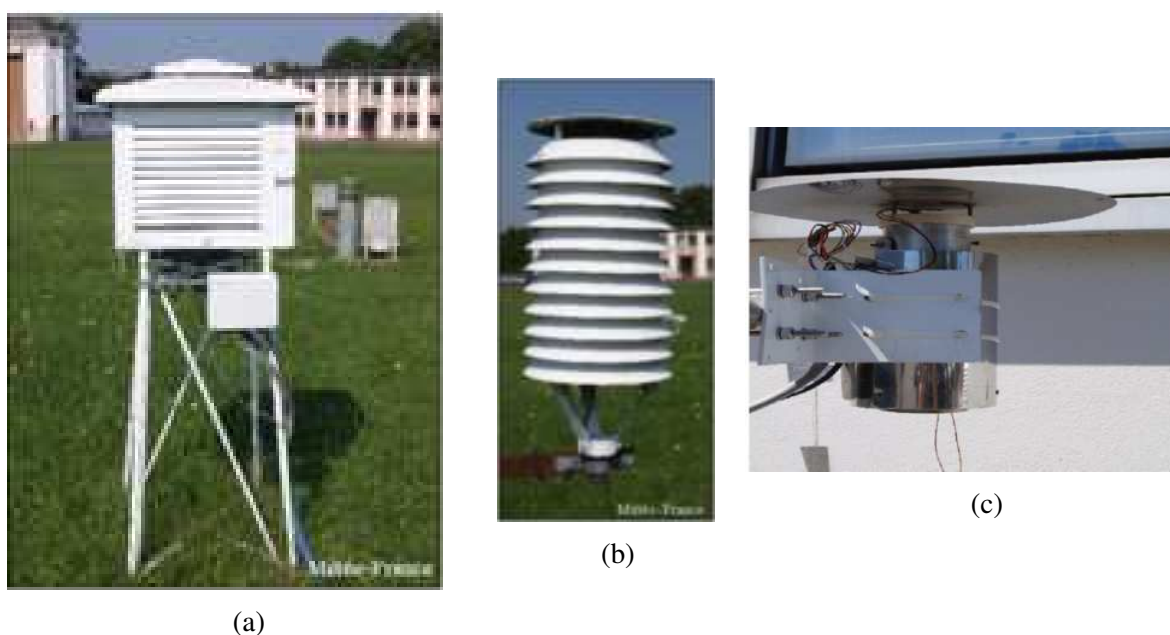


FIGURE 1.8 – Exemples d'abris ventilés utilisés pour protéger les capteurs de température du rayonnement. Deux catégories existent : les abris à ventilation naturelle comme (a) l'abri de Stevenson *modèle réduit* ou (b) l'abri de Stevenson *miniature* et (c) les abris à ventilation forcée. Source : Météo-France

Ces procédés (réduction de la taille du capteur, utilisation d'abris ventilés naturellement ou mécaniquement) ne permettent néanmoins pas toujours de réduire l'erreur  $e_T$  en-dessous de l'ordre de grandeur de la précision du capteur et peuvent dans certains cas être qualifiés d'*intrusifs* :

- un abri à ventilation naturelle de grande taille peut entraîner un déphasage dans le signal de température ;
- un abri à ventilation forcée peut sous certaines conditions aspirer de l'air chaud ou froid provenant de sous-couches inférieures et donc augmenter ou diminuer artificiellement la température mesurée.

Il n'y a donc pas de procédé toujours meilleur qu'un autre. Une analyse plus détaillée sera réalisée au chapitre 2, lorsque le matériel utilisé dans le cadre de cette thèse sera présenté.

Si le choix du matériel est fondamental pour mesurer la température d'air en extérieur, il est également fondamental que ce matériel soit homogène à tous les sites d'un réseau de capteurs (Oke 2004). En effet, l'utilisation d'un matériel différent pour deux sites de mesure augmente les incertitudes et complique l'intercomparaison de leur température.

### 1.2.1.2 Positionnement des capteurs de température

La température de l'air dépend des caractéristiques thermiques des surfaces qui l'entourent. Plus l'air est proche d'une surface, plus sa température est sensible à la température de cette surface (cf. section 1.1.2.2). Afin de réaliser des mesures représentatives d'un lieu, certaines précautions doivent être prises. Météo-France définit par exemple des classes permettant de qualifier la représentativité d'un site en fonction du respect de certaines contraintes de positionnement. Pour prétendre à la classe 1, un site de mesure de température d'air doit par exemple respecter les contraintes suivantes (illustration figure 1.9) :

- assurer qu'aucune source de chaleur ou aucun point d'eau ne soit situé à moins de 100 mètres de l'appareil de mesure ;
- assurer qu'aucun masque ne soit présent sur le parcours du soleil au-dessus de 3° d'angle par rapport à l'horizontale ;
- assurer de la présence d'une couverture de végétation basse inférieure à 10 centimètres dans un rayon inférieure à 100 mètres.

Si cette classification a du sens et peut être appliquée pour comparer des températures à l'échelle régionale ou nationale, ce n'est pas le cas à l'échelle locale, d'autant plus en zone urbaine. En ville, la diversité des surfaces, des sources de chaleur et des obstacles au soleil est telle qu'aucun site ne pourrait respecter ces recommandations. C'est justement l'influence de ces surfaces et de ces formes contrastées sur la température de l'air que l'on souhaite mesurer. Même si les recommandations précédentes ne peuvent pas être appliquées, d'autres plus appropriées au milieu urbain ont été proposées (Oke 2004). Le concept de représentativité existe également mais il est défini autrement. Un capteur situé dans l'espace urbain est considéré comme représentatif des conditions locales de température sous deux conditions.

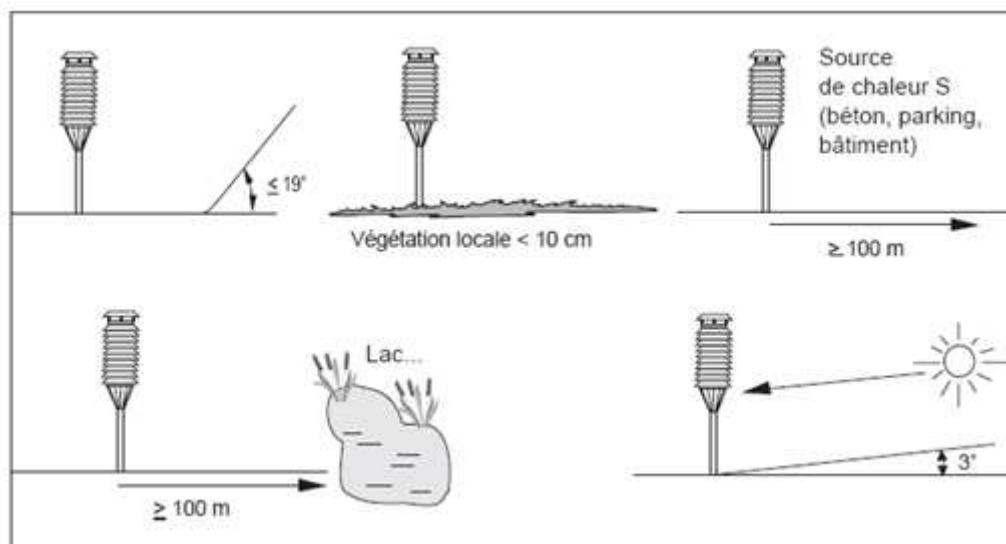


FIGURE 1.9 – Critères d’installation d’un capteur de température en classe 1. Source : Météo-France

- Première condition : la rue dans laquelle le capteur est placé doit être représentative des caractéristiques de forme et de surface des rues alentours. La hauteur des bâtiments, la largeur des rues, la quantité de végétation haute et basse seront par exemple des paramètres à considérer. Mesurer un quartier très minéralisé et à fort rapport d’aspect  $\frac{H}{W}$  ne doit pas être réalisé en positionnant le capteur dans le seul petit espace vert du quartier. Le risque est que les surfaces les plus proches (qui sont dans ce cas les surfaces enherbées) soient celles qui contribuent le plus à la valeur de la température mesurée. Ces notions de surfaces d’influence et de représentativité seront abordées plus en détail au chapitre 3.
- Seconde condition : le capteur ne doit pas être placé à proximité immédiate des surfaces. En effet, l’objectif est de mesurer les variations de température à l’échelle locale tout en s’affranchissant des variations à l’échelle microclimatique (cf. section 1.1.2.2 pour la définition des échelles). Pour cela, Oke (2004) recommande d’éloigner l’instrument de mesure à une distance supérieure à un mètre de toute surface (mur, sol, arbre, etc.). Si dans des espaces très ouverts, comme en campagne, la hauteur recommandée est de 1.25 à 2 m, placer les capteurs à des hauteurs comprises entre trois et cinq mètres ne pose pas de problème en milieu urbain (affirmation d’autant plus vraie lorsque la densité de bâtiments est élevée). Armstrong (1974), Nakamura et Oke (1988) et Taesler (1980) ont en effet montré qu’au-delà de deux mètres du sol, l’air situé dans la canopée est bien mélangé et par conséquent relativement homogène en température. Les différences maximales de température entre deux points de la rue sont souvent inférieures à 0.5°C et



toujours inférieures à 1°C. Dans le cas où les capteurs sont situés à proximité d'un axe routier, fixer les capteurs légèrement plus haut que trois mètres permet donc d'obtenir des températures tout-à-fait représentatives des conditions dans la canopée tout en limitant l'influence des gaz d'échappement.

### 1.2.1.3 Méthodes de mesure

Pour mesurer les variations spatiales et temporelles de la température au sein d'une zone urbaine, deux méthodes sont utilisées : la méthode de mesure *fixe* et la méthode de mesure *mobile*.

La méthode de mesure fixe, comme son nom l'indique, consiste à fixer des capteurs pendant une période donnée sur des supports immobiles (poteaux d'éclairage public, arbres, bâtiments). Cette méthode est souvent utilisée pour acquérir des jeux de données durant une à plusieurs années et à périodes d'échantillonnage de l'ordre de quelques minutes à une heure. Elle permet donc une caractérisation fine des variations temporelles de l'ICU. En termes matériel, les thermomètres à résistance et les abris ventilés naturellement sont le plus souvent utilisés.

- La raison de l'utilisation de thermomètres à résistance est qu'ils permettent de relever une température d'air moins fréquemment qu'il ne le faudrait avec un thermocouple. Ceci s'explique par leur temps de réponse plus élevé, qui a pour effet de moyennner naturellement le signal de température en coupant les variations hautes fréquences. Ces variations temporelles brutales et très ponctuelles de température sont dues à des phénomènes de turbulence présents au sein de la canopée, mais ne sont pas intéressantes à mesurer dans notre cas. Les thermocouples, quant à eux, sont sensibles à ces variations brusques de température. Afin de se débarrasser de cette information haute fréquence peu utile dans notre cas, il est donc nécessaire de sur-échantillonner puis de moyennner le signal obtenu (représentation schématique figure 1.10).
- La raison de l'utilisation d'abris ventilés naturellement est que les thermomètres à résistance sont très sensibles au rayonnement. Il convient donc de les protéger. L'utilisation d'abris ventilés mécaniquement (donnant globalement de meilleures performances que les abris ventilés naturellement) est possible mais nécessite continuellement un apport d'énergie électrique. L'équipement doit dans ce cas bien souvent être raccordé au réseau électrique, ce qui engendre des complications supplémentaires à l'installation du matériel. Sauf s'ils sont de taille très faible, l'utilisation d'abris est également recommandée

dans le cas où les thermomètres utilisés sont des thermocouples (problématique abordée au chapitre 2).

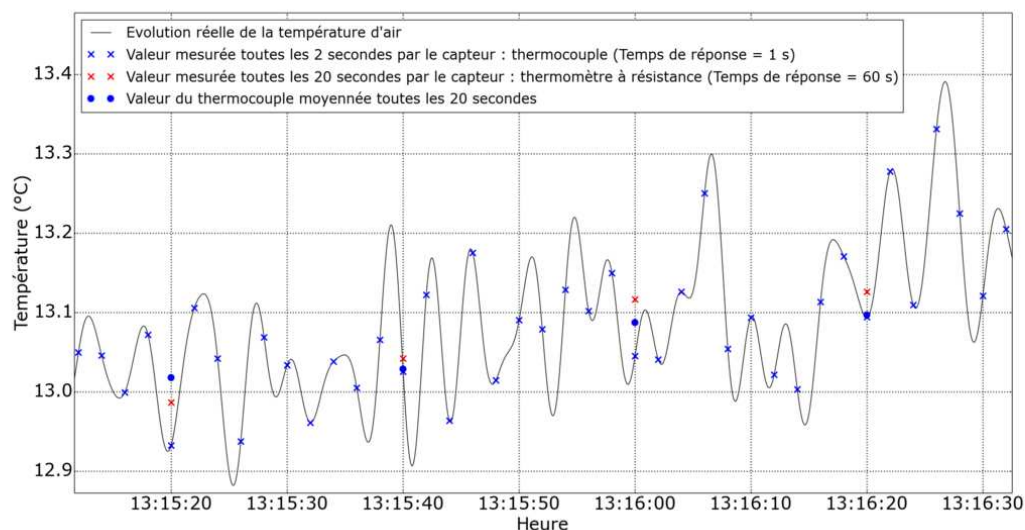


FIGURE 1.10 – Échantillonnage de valeurs de température pour différents pas de temps et différents capteurs de mesure

Le positionnement de ces capteurs doit également correspondre le plus fidèlement possible au positionnement recommandé par Oke (2004), décrit en 1.2.1.2. La distance aux surfaces urbaines doit être le point de vigilance principal. Elle doit être suffisamment importante afin que la mesure soit représentative de la température de la rue et également afin d'assurer une bonne ventilation de l'abri (limitation de l'échauffement du capteur en maximisant le coefficient de convection  $h_c$  - voir 1.2.1.1). L'inconvénient de la méthode de mesure fixe est que ses résultats sont très dépendants de la qualité des sites choisis au départ, mais également de leur nombre. En effet, la caractérisation spatiale de l'ICU est plus difficile à réaliser lorsque le nombre de sites est faible.

Contrairement à la méthode de mesure fixe, la méthode de mesure mobile ne dispose pas de cette contrainte d'emplacement, puisqu'elle consiste à fixer un thermomètre sur un bras de déport et à le déplacer le long d'un parcours défini au sein de l'agglomération. Les moyens de déplacements sont divers : la marche peut être utilisée pour des distances courtes (Chen et al. 2012 ; Le Bras 2015), tandis que le vélo (Yan et al. 2014), le deux roues motorisé (Yokobori et Ohta 2009) et la voiture (Balázs et al. 2009 ; Hart et Sailor 2009) sont utilisés pour des circuits plus longs. La couverture spatiale de cette méthode est donc potentiellement beaucoup plus importante. En termes matériel, un thermomètre, un abri ventilé et un GPS sont utilisés dans la plupart des cas.

- Le thermomètre peut être indifféremment un thermomètre à résistance ou un thermocouple. En effet, le temps de réponse du thermomètre à résistance peut être réduit à quelques secondes lorsqu'il est déplacé à une vitesse de 20 à 30 km/h (augmentation du coefficient  $h_c$  du capteur avec l'air). Cela permet donc de retranscrire les différences de température tout le long du parcours indifféremment avec un thermomètre à résistance ou avec un thermocouple.
- L'abri à ventilation naturelle est souvent privilégié, mais sa forme diffère de celle des abris utilisés habituellement dans le cadre de mesures fixes. Des tubes couchés dans le sens du déplacement leurs sont préférés afin d'assurer une bonne ventilation du capteur et limiter la prise au vent.
- Le GPS embarqué dans le véhicule permet de faire le lien entre l'instant de la mesure et la position géographique où elle a été effectuée.

Ces mesures sont coûteuses en temps humain car les trajets durent souvent plus d'une heure et sollicitent simultanément une à plusieurs personnes. Du fait du changement de température qui a pu se produire pendant la durée du parcours, il est parfois nécessaire de corriger les mesures qui ont été effectuées en fin de parcours. Le nombre de traversées est souvent faible, les mesures mobiles ne permettent pas une caractérisation temporelle de l'ICU aussi fine que les mesures fixes.

Les deux méthodes sont finalement assez complémentaires. La première permet d'identifier la composante temporelle de l'ICU et les instants clés pour lesquels les différences de température sont avérées. La seconde, mise en oeuvre à ces moments clés, permet d'identifier les principaux facteurs à l'origine de ces différences de température en multipliant les points de mesure.

### 1.2.2 Choix d'une station comme référence du climat rural

Une fois les mesures acquises pour chacun des sites, elles peuvent être :

- utilisées directement : un signal de température  $T_i(t)$  est associé à chaque station  $i$  ;
- plus rarement, dérivées temporellement : un signal  $\dot{T}_i(t) = \frac{dT_i}{dt}(t)$  qualifié de taux de rafraîchissement est associé à chaque station  $i$  (Oke et Maxwell 1975) ;
- très souvent rapportées à la température d'une station de référence : un signal  $\Delta T_{i-ref}(t) = T_i(t) - T_{ref}(t)$  qualifiant l'amplitude de l'ICU est associé à chaque station  $i$ .

La température de référence utilisée est dans la plupart des cas relevée sur un site extérieur à la ville, considéré comme représentatif de la température de la périphérie. La température peut être très différente entre deux points de la périphérie, selon le type d'activité alentour (arrosage des sols), la distance à la tâche urbaine, le type de sols, etc. Le choix du site n'est donc pas arbitraire, il doit être fait afin de ne pas faire ressortir des effets trop locaux et non représentatifs de la périphérie (Runnalls et Oke 2000). Dans la seule optique de comparer des sites entre eux, Fenner et al. (2014) utilisent la température moyenne de toutes leurs stations  $i$  comme référence  $ref$  ( $T_{ref} = moy(T_i(t))_i$ ). Cette méthode est néanmoins peu appropriée pour modéliser et calculer des différences de température à partir de régressions géoclimatiques puisqu'il faudrait associer un environnement géographique à cette station virtuelle, association qui serait elle-même source d'incertitudes.

### 1.2.3 Caractérisation des variations temporelles

L'ICU d'une station  $i$  quelconque n'est pas constant au cours du temps. Afin d'étudier ses variations temporelles, il s'agit d'identifier :

- les cycles principaux à l'origine des variations de température ;
- les éléments qui peuvent perturber la température à l'intérieur de ces cycles.

Le traitement des aspects temporels peut se résumer par la chaîne de traitement présentée figure 1.11.

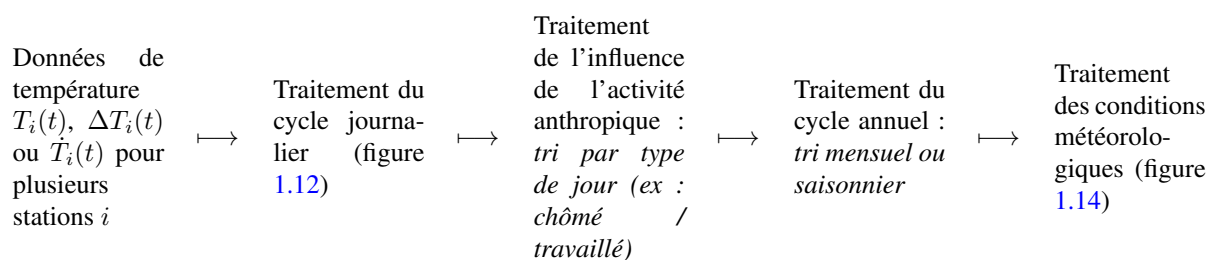


FIGURE 1.11 – Chaîne de traitement classique utilisée pour traiter les variations temporelles de l'ICU

#### 1.2.3.1 Variations journalières et saisonnières

Le rayonnement solaire global net  $Q^*$  est (avec  $Q_F$ ) l'un des deux termes source de l'équation 1.1). Le rayonnement solaire  $K \downarrow$  contenu dans le terme  $Q^*$ , vu d'un point de la surface terrestre,

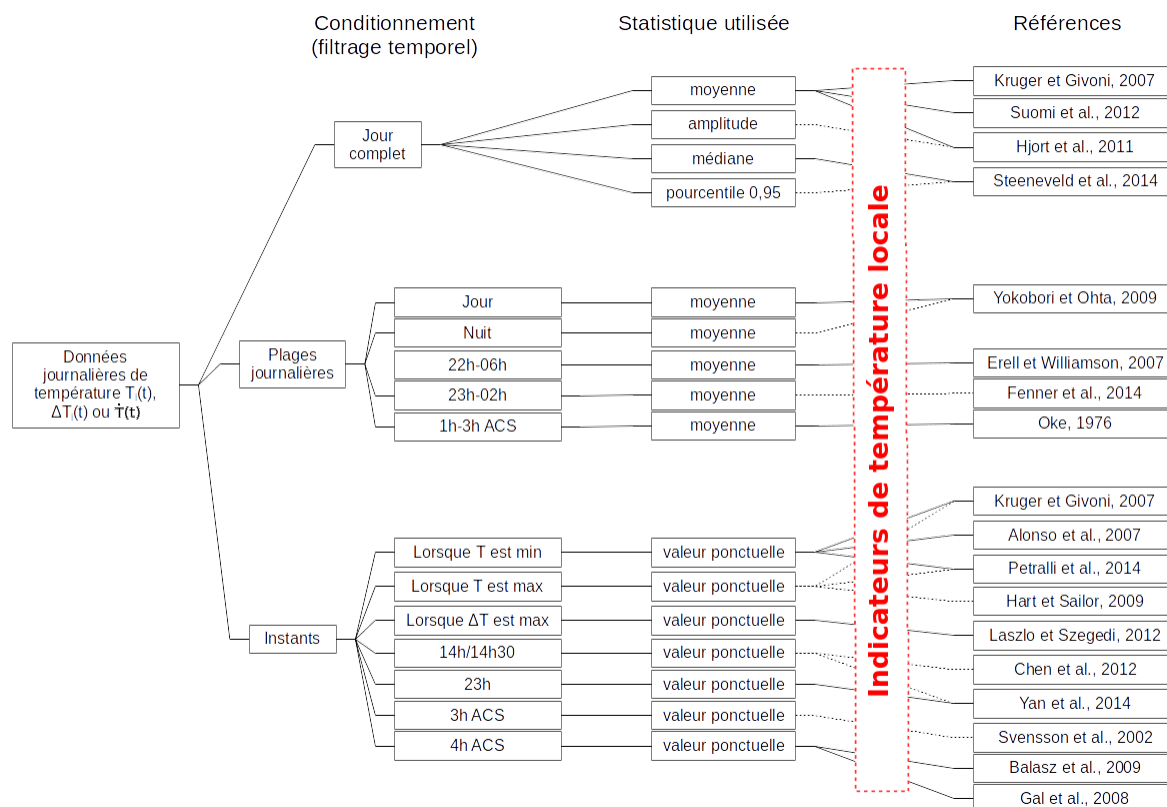
varie principalement selon deux cycles :

- le cycle annuel correspondant au cycle de rotation de la Terre autour du soleil (alternance des mois ou des saisons) ;
- le cycle journalier correspondant au cycle de rotation de la Terre sur elle-même (alternance jour / nuit).

Les variations des autres termes sources ( $Q_F$ ,  $K \uparrow$  et  $L \uparrow$  - le rayonnement grandes longueurs d'ondes montant) sont elles-mêmes dictées par ces deux cycles. L'angle d'incidence des rayons du soleil sur une surface varie au cours d'une journée mais également d'une année, modifiant la puissance reçue par une surface horizontale ou verticale. Selon les formes urbaines et les types de surfaces recevant ce rayonnement, la température de l'air varie ainsi spatialement et temporellement. Lorsque des mesures de température sont effectuées au sein d'une zone urbaine en vue d'étudier l'ICU, elles sont souvent triées selon ces deux cycles (figure 1.11).

Les études issues de la littérature traitent le cycle diurne de façons très diverses. Les principales différences apparaissent dans le choix de la période d'intérêt du cycle, mais également dans le choix du type d'indicateur statistique à utiliser. Ces différences sont mises en évidence figure 1.12.

Le choix de la période d'intérêt est souvent lié à la méthode de mesure employée : les études qui utilisent des mesures fixes utilisent des jours complets ou des plages horaires tandis que celles basées sur des mesures mobiles considèrent des instants particuliers de la journée. Des exceptions apparaissent : Svensson et al. (2002) ou encore Krüger et Givoni (2007) utilisent des données issues de stations fixes mais calculent des indicateurs de température locales ( $ITL$  à des instants précis. D'une manière générale, lorsque des plages ou des instants précis sont étudiés, le choix des horaires est rarement arbitrairement choisi. Souvent, il est motivé par des résultats issus de la littérature, mettant par exemple en évidence que l'ICU est maximal quelques heures après le coucher du soleil (environ 4h selon Oke (1987)). Ces éléments de la littérature sont parfois vérifiés à partir des données de l'étude : Balázs et al. (2009) confirment ainsi la valeur de 4h trouvée par Oke (1987) tandis que Svensson et al. (2002) ramènent finalement cette valeur à 3h. Concernant le type de statistique utilisé, la technique classique consiste à moyenniser les températures ou différences de température sur la période concernée. Quelques alternatives apparaissent lorsque la journée entière est considérée. Hjort et al. (2011) calculent ainsi l'amplitude de température au cours de la journée tandis que Steeneveld et al. (2014) calculent des valeurs particulières de centile de la série  $\Delta T(t)$  (par exemple le centile  $q =$



ACS : après le coucher du soleil

FIGURE 1.12 – Classification des différentes approches utilisées dans la littérature pour traiter les variations temporelles de l'ICU

0,95<sup>2</sup>) (Oke 1976).

Le traitement du cycle annuel se fait généralement en triant puis en moyennant les indicateurs de température précédemment obtenus (figure 1.12) par mois ou par saison. Si la définition du mois est la même pour tous les auteurs, celle de la saison ne fait pas toujours consensus. Le découpage par saison réalisé par Petralli et al. (2014) et également par Yan et al. (2014) est basé sur la définition de l'Organisation Météorologique Mondiale (OMM) appliquée à l'hémisphère nord (hiver : 1er décembre au 31 février, printemps : du 1er mars au 31 mai, été : du 1er juin au 31 août, automne : du 1er septembre au 31 novembre). Yokobori et Ohta (2009) ou encore Svensson et al. (2002) découpent eux-aussi leurs données par saison mais sans donner d'informations précises sur le découpage. Hjort et al. (2011) considèrent qu'un seul mois peut être utilisé pour représenter chaque saison : le mois central. Le découpage mensuel est moins fréquent (Suomi et al. 2012), peut-être notamment car il demande un jeu de données plus conséquent afin de consolider les analyses statistiques. Il permet en tous cas des analyses plus fines que le découpage saisonnier. Suomi et al. (2012) expliquent par exemple que pour

<sup>2</sup> pour rappel le centile 0,95 d'une série de donnée consiste à prendre son maximum en éliminant préalablement les 5% des valeurs les plus élevées.

une ville comme Turku (Finlande), l'influence de l'océan sur les stations les plus proches des côtes varie au cours des mois. Sa masse thermique est telle que les extremum de sa température annuelle sont déphasés dans le temps. Ainsi, il contribue au réchauffement des stations côtières en septembre, à leur refroidissement en mars mais il a moins d'influence le reste de l'année.

### 1.2.3.2 Autres échelles de variations

Les cycles solaires permettent un découpage rationnel du temps mais ne sont pas suffisants pour saisir toutes les échelles de variabilité temporelle de l'ICU. En effet, le flux de chaleur anthropique  $Q_F$  apparaît lui aussi comme un terme de premier ordre à considérer parfois au même titre que le terme de rayonnement global net  $Q^*$ . De plus, les conditions météorologiques régionales modifient la valeur de tous les termes de l'équation 1.1, notamment celle du terme source  $Q^*$  (plus particulièrement  $K \downarrow$ , à l'origine des cycles solaires).

### Modification du microclimat urbain par les activités humaines

Le terme  $Q_F$ , représentatif des flux générés par l'activité humaine, apparaît comme non-négligeable lorsque l'on réalise le bilan énergétique d'un volume urbain. Ceci est d'autant plus vrai dans les centres urbains densément bâtis et lorsque l'hiver est froid. Oke (1987) montre ainsi que pour une ville comme Montréal (Canada), le flux journalier moyen annuel de chaleur anthropique (estimé à partir des données de consommations de combustibles et d'électricité de la ville à  $99 \text{ W.m}^{-2}$  pour l'année 1961) est supérieur au flux journalier moyen annuel de rayonnement global net (estimé à  $52 \text{ W.m}^{-2}$ ). Celui de villes d'Europe de l'ouest est souvent moins important, représentant tout de même le tiers du flux de rayonnement global annuel net (à Berlin ouest,  $Q_F$  vaut sur l'année 1967 en moyenne  $21 \text{ W.m}^{-2}$  contre  $57 \text{ W.m}^{-2}$  pour  $Q^*$ ). Les valeurs de flux anthropiques sont bien entendu anciennes et ont toutes les chances d'avoir augmenté depuis<sup>3</sup>. Elles sont également moyennées ce qui leur fait perdre tout caractère dynamique. En effet, le flux anthropique est plus élevé à certaines périodes de l'année (qui correspondent notamment aux périodes de chauffage ou de climatisation), certains jours (les jours travaillés) et à certains moments de la journée (pics matinaux et vespéraux dus notamment aux consommations d'électricité spécifique, aux consommations de chaleur et au trafic routier - figure 1.13).

Peu de modèles de régression géoclimatique tentent de mettre en évidence l'effet du

<sup>3</sup> En effet, d'après l'Agence Internationale pour l'Énergie (AIE), plus de deux tiers de l'énergie primaire mondiale était consommée par les villes en 2009 (Sims et al. 2009) alors que l'énergie primaire mondiale a vu sa consommation annuelle doubler entre 1971 et 2012 (l'énergie 2014)

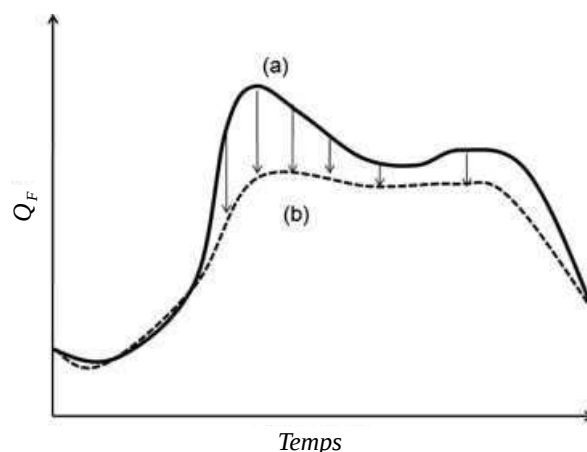


FIGURE 1.13 – Profil typique du flux de chaleur anthropique généré pour (a) les jours de la semaine, et (b) les jours du week-end. Source : Sailor (2011)

flux anthropique sur les températures d'air mesurées. Seuls Hart et Sailor (2009) l'intègrent en dissociant deux modèles : l'un pour les jours de la semaine et l'autre pour les jours du week-end. Ils mettent en évidence que durant l'après-midi, l'ICU est plus marqué en semaine que le week-end et qu'il est situé à proximité des axes majeurs de communication et des zones commerciales. Ils attribuent cette chaleur principalement au trafic routier qui est souvent plus dense dans l'après-midi en semaine que le week-end, constat cohérent avec celui fait par Sailor (2011). D'autres manières de modéliser l'effet de  $Q_F$  sur l'ICU existent mais elles sont propres aux méthodes de MC et sont donc insérées dans les équations de bilan de chaleur d'un volume urbain (voir la revue de la littérature de Sailor (2011) pour plus de détails sur ces méthodes). Elles ne peuvent donc pas conduire à modéliser l'ICU par des méthodes de régression géoclimatique. La seule proposition pouvant s'appliquer aux méthodes de RG est donc celle faite par Hart et Sailor (2009), qui consiste à trier les données selon les jours travaillés ou chômés. Elle a pour effet d'ajouter une composante temporelle due à l'activité humaine, qui se superpose aux composantes journalières et annuelles dues aux cycles solaires. On note d'ailleurs sur la figure 1.11 qu'elle s'intègre entre ces deux composantes.

### Modification du microclimat urbain selon les conditions météorologiques

Les conditions météorologiques régionales impactent tous les termes de l'équation 1.1 :

- la vitesse du vent, lorsqu'elle augmente, contribue à l'augmentation des termes  $Q_H$ ,  $Q_E$  et  $\Delta Q_A$  au détriment de  $\Delta Q_S$  ;



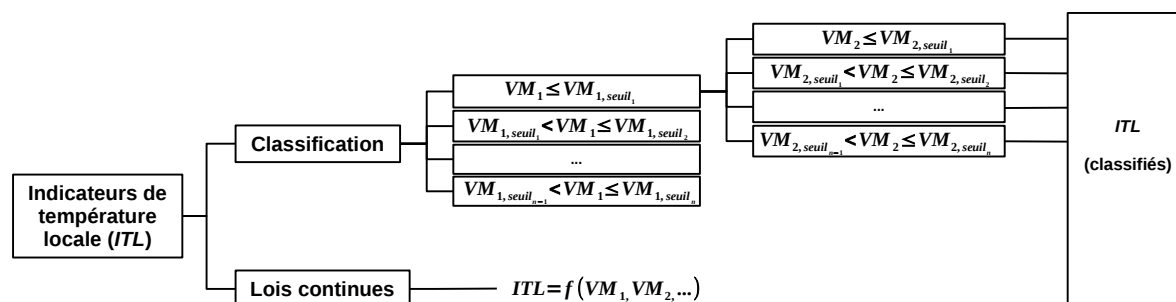
- la couverture nuageuse (ou nébulosité), lorsqu'elle augmente, diminue le rayonnement solaire descendant  $K \downarrow$  et augmente le rayonnement grandes longueurs d'onde descendant  $L \downarrow$ ;
- la température, lorsqu'elle diminue fortement en hiver, augmente le terme  $Q_F$  dû aux besoins de chauffage (c'est l'inverse pour les besoins de climatisation en été) ;
- l'humidité relative, lorsqu'elle augmente, contribue à la réduction du terme  $Q_E$  au détriment des termes  $Q_H$  et  $\Delta Q_S$  ;
- la présence de pluie modifie principalement les termes  $Q_H$ ,  $Q_E$  et  $\Delta Q_S$ .

Pour deux sites qui présentent une morphologie et des types de sols distincts, l'impact des conditions météorologiques sur les flux est potentiellement différent. La valeur de l'ICU varie donc temporellement et spatialement selon les variations temporelles de ces conditions météorologiques.

Les différences de température les plus élevées qui apparaissent au sein d'une ville sont relevées de nuit lorsque les conditions sont décrites comme claires et calmes (respectivement faible couverture nuageuse et faibles vitesses de vent) (Oke 1987). L'influence de la vitesse du vent et de la nébulosité sur l'ICU sont les plus décrites dans la littérature, même si d'autres variables météorologiques ont également une influence sur la valeur de l'ICU. Krüger et Givoni (2007) évaluent par exemple l'influence de la température régionale et du rayonnement solaire descendant sur l'ICU.

L'influence des conditions météorologiques sur la valeur de l'ICU est prise en compte de deux manières distinctes lorsque l'ICU est modélisé par des méthodes de RG. Ces deux méthodes, présentées schématiquement figure 1.14, utilisent toutes les deux les indicateurs de température locale issus du traitement schématisé figure 1.12.

La première consiste à classer les données par type de conditions météorologiques. Les données recueillies lorsque la vitesse du vent est supérieure à 3 m/s sont par exemple séparées du reste des données. Les variables météorologiques  $VM_m$  permettant la classification (dans notre exemple la vitesse du vent) ainsi que les seuils  $VM_{m_{seuil}}$  fixés pour ces variables (dans notre exemple  $U_{seuil_1} = 3 \text{ m/s}$ ) sont souvent bien différents d'une étude à l'autre. Svensson et al. (2002) ne conservent par exemple leurs données pour l'analyse que lorsque la nébulosité est inférieure ou égale à 2 octa tandis que Alonso et al. (2007) les classent par intervalle de 1 octa ( $N_{seuil_1} = 0 \text{ octa}$ ,  $N_{seuil_2} = 1 \text{ octa}$ ,  $N_{seuil_3} = 2 \text{ octa}$  et  $N_{seuil_4} = 3 \text{ octa}$ ). Dans



$VM_m$  valeur de la variable météorologique régionale  $m$  (à l'instant  $t$  auquel  $ITL$  a été calculé)  
 $VM_{m,seuil}$  valeur seuil de  $VM_m$  permettant de trier par classe les valeurs de l' $ITL$

FIGURE 1.14 – Méthodes classiquement utilisées dans la littérature pour traiter de l'influence des conditions météorologiques sur les variations temporelles de l'ICU.

chacune de ces deux études, les données pré-classées sont ensuite triées par vitesse de vent, en utilisant des seuils de vitesse  $U_{seuil}$  différents. D'autres études décrivent seulement de manière qualitative les conditions météorologiques (en les définissant par exemple comme *claires* (Hart et Sailor 2009)), de sorte qu'il est encore plus difficile de comparer leurs résultats à d'autres études.

La seconde façon de prendre en compte l'influence des conditions météorologiques sur l'ICU consiste à créer des modèles de régression entre l'ICU et les variables météorologiques. C'est ce que font par exemple Erell et Williamson (2007) qui utilisent un polynôme du deuxième ordre pour évaluer l'influence de la vitesse du vent  $U$  sur la valeur de l'ICU (équation 1.5). Cette relation illustre le fait que l'ICU décroît lorsque le vent augmente mais non linéairement. Ces résultats sont cohérents avec ceux obtenus par Oke (1976) ou encore Svensson et al. (2002) qui modélisent respectivement l'ICU comme une fonction de  $\frac{1}{U^{0,46}}$  et de  $\log(U)$ . Ces relations sont néanmoins limitées à un seul facteur explicatif de la variation de la valeur de l'ICU, la vitesse du vent, alors qu'elle est dépendante de plusieurs variables. Des régressions linéaires multiples prenant en compte la température, la vitesse de vent et également le rayonnement solaire régionaux sont ainsi utilisées par Krüger et Givoni (2007). Le problème est que l'action d'une variable sur l'ICU est différente selon le site étudié. Krüger et Givoni (2007) utilisent donc un modèle de régression différent par station. Même si ces modèles sont utiles à la compréhension des paramètres influents sur l'ICU de chaque site, ils ne peuvent pas être utilisés pour estimer l'ICU de n'importe quel point de l'agglomération.

L'ICU est en partie expliqué par un déphasage temporel de température dû à un terme de stockage  $\Delta Q_S$  important. Les conditions météorologiques à un instant  $t - 1$  ont donc potentiellement une influence sur la valeur de l'ICU à un instant  $t$ . Le Bras (2015) considère ce

potentiel déphasage, en calculant entre autres le rayonnement solaire moyen dans les 12 heures qui précèdent la mesure de l'ICU dans le but de voir si cet indicateur est une variable explicative de l'ICU. Pour expliquer les variations de l'ICU à un instant  $t$ , toutes les autres études citées dans cette thèse utilisent les conditions météorologiques seulement à ce même instant  $t$ .

## 1.2.4 Caractérisation des variations spatiales

La section précédente ne traite que de la variabilité de l'ICU au cours du temps mais ne permet pas de comprendre les différences d'ICU spatiales, c'est-à-dire entre les stations. Cette section est donc consacrée aux variations spatiales. Elle aborde les approches existantes pour qualifier et quantifier les éléments du territoire à l'origine de l'ICU. Une première partie est consacrée aux indicateurs géographiques et aux bases de données permettant de les créer. La seconde partie traite de la taille et la forme de la zone alentour à chaque capteur qui doit être considérée pour calculer ces indicateurs.

### 1.2.4.1 Donnée spatiale et indicateurs géographiques

Robitaille et Laguë (2009) donnent une définition assez fine du concept d'indicateur géographique en s'appuyant sur la littérature existante :

*«Le concept d'indicateur, étymologiquement, est utilisé afin de montrer (Joerin et Rondier 2007). «[Il] traduit des données et des statistiques en une information succincte et claire. Il favorise une meilleure compréhension des phénomènes complexes et une utilisation par différents groupes de personnes ayant des préoccupations diverses» (Desthieux 2004). L'indicateur géographique simplifie donc une information sur un phénomène ayant une référence spatiale. Desthieux (2004) et Maby (2003) distinguent deux principaux types d'indicateurs géographiques : les indicateurs thématiques et les indicateurs spatiaux. Les indicateurs thématiques sont définis par l'association d'un ensemble de valeurs à un objet géographique (ex. : proportion des logements construits avant 1946 à l'échelle des aires de diffusion). Ce sont les indicateurs les plus connus et les plus utilisés dans la production de cartes thématiques. [...] Les indicateurs spatiaux nous informent sur l'organisation de l'objet dans l'espace. Il est utilisé comme un élément de base dans l'analyse spatiale (ex. : distance aux commerces d'alimentation le plus proche).»*

Les indicateurs géographiques, dans les études qui utilisent des approches de modélisation par RG, sont utilisés pour qualifier le territoire proche d'une station de mesure. Ils

peuvent être triés selon les deux catégories d' $IG_g$  définis précédemment. La première catégorie, celle des indicateurs thématiques, est elle-même découpée en deux sous-catégories : les indicateurs décrivant la morphologie (des objets ancrés au sol mais aussi du sol lui-même) et ceux qui décrivent l'occupation des sols (par leur nature et par leur type d'occupation). La deuxième catégorie contient les indicateurs spatiaux de distance à un objet géographique. La figure 1.15 dresse une liste non-exhaustive des indicateurs géographiques utilisés dans les études de RG. Ces indicateurs sont triés selon les catégories définies précédemment et le type de bases de données qui permettent de les calculer est indiqué.

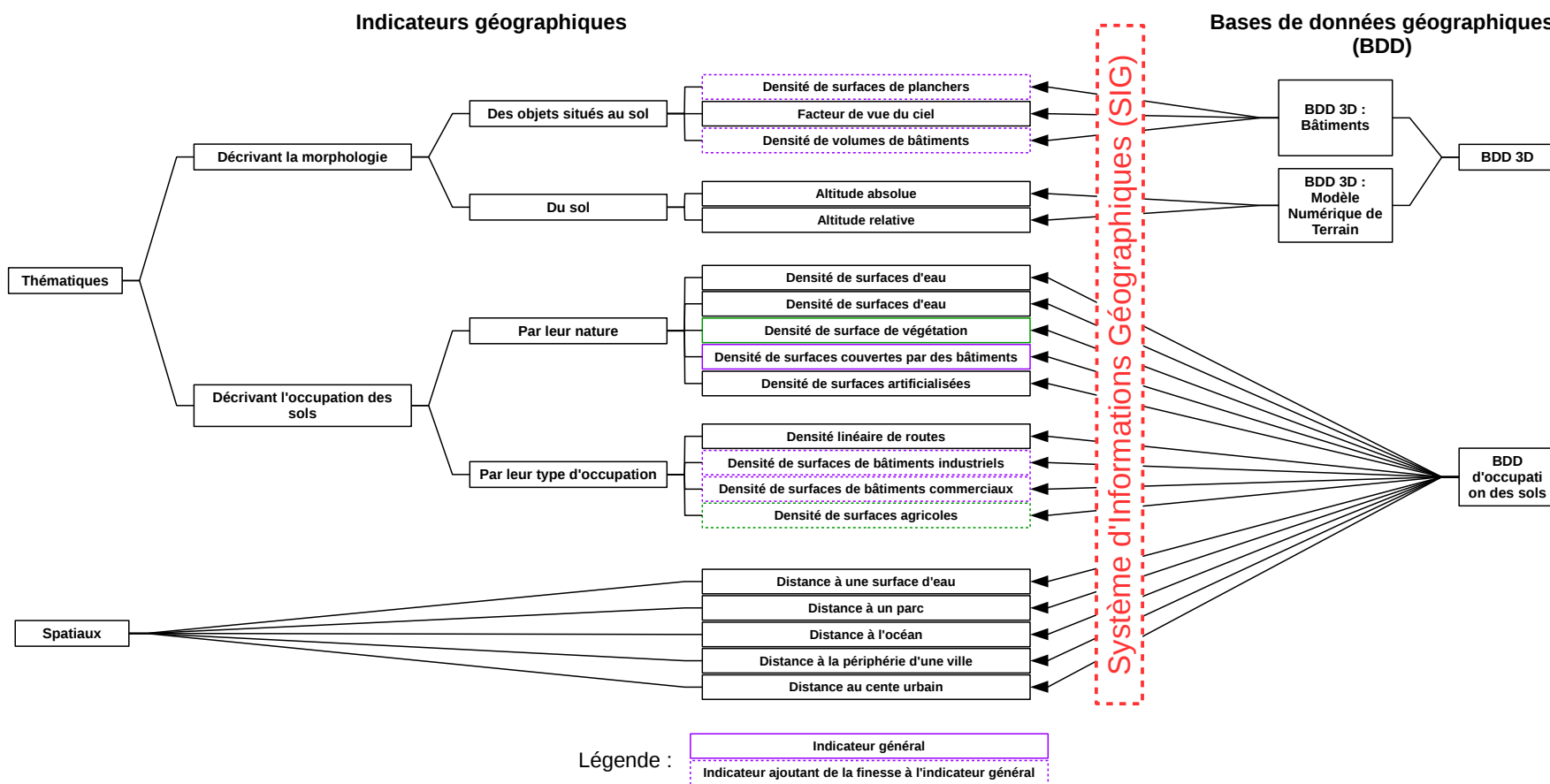


FIGURE 1.15 – Catégorisation des principaux indicateurs et bases de données géographiques qui sont utilisés dans la littérature traitant de modélisation de l'ICU par des méthodes de régression géoclimatique

Cette méthode de classification permet d'établir des liens clairs entre indicateurs géographiques et indicateurs de température locale. Les formes de la ville, décrites par les indicateurs morphologiques, influent par exemple fortement sur les échanges radiatifs  $Q^*$  et le flux de chaleur anthropique  $Q_F$  dégagé par les bâtiments. Les indicateurs relatifs à la nature des surfaces permettent d'identifier si des surfaces sont capables d'évaporer l'eau contenue dans les sols mais également de leur associer des caractéristiques thermiques particulières (effusivité, albédo, etc.). Le type d'occupation des sols permet d'identifier des zones à fort dégagement de chaleur anthropique  $Q_F$  (zones industrielles par exemple) et d'identifier également si des surfaces végétales auront un fort potentiel d'évaporation dû à un arrosage régulier (ex : terres cultivées). Tous ces indicateurs sont calculés au sein d'une zone géographique que l'on qualifiera de contributive à la température de l'air mesurée par le capteur (concept qui sera décrit plus en détail à la section 1.2.4.2). Cette première approche consiste donc à identifier les surfaces contributives du point de vue du capteur. Une seconde approche existe et est potentiellement complémentaire de la première. Elle consiste à identifier les principales sources de chaleur ou de fraîcheur de la ville et à calculer la distance du capteur à ces sources. Les indicateurs associés à cette approche sont les indicateurs spatiaux qui définissent par exemple la distance à un espace vert.

Certains indicateurs permettent d'affiner d'autres plus globaux. C'est par exemple le cas de la densité de surface agricole utilisée par Suomi et al. (2012). Cet indicateur dérive de la densité de surface de végétation, en qualifiant plus précisément l'activité se déroulant sur certaines surfaces couvertes de végétation.

Les bases de données utilisées pour calculer un indicateur sont souvent propres à chaque étude. Pour connaître l'emplacement et la hauteur des bâtiments, les bases de données utilisées sont souvent des bases de données municipales. L'information sur la nature des surfaces est issue d'un traitement d'images satellitaires (ou de photographies aériennes) qui permet de distinguer par exemple les surfaces artificialisées des surfaces de végétation. Ce traitement, réalisé par photo-interprétation, peut être effectué visuellement pour des surfaces restreintes (Krüger et Givoni 2007 ; Yan et al. 2014 ; Yokobori et Ohta 2009). Il peut également être automatisé à partir d'algorithmes de classification pour des surfaces plus larges. Les données produites par cette méthode ont l'avantage d'être homogènes et continues à l'échelle du territoire étudié (lorsqu'elles sont produites dans le cadre de l'étude (Balázs et al. 2009 ; Petralli et al. 2014) ou produites par les collectivités (Fenner et al. 2014 ; Hart et Sailor 2009)) et parfois même à l'échelle d'un pays (lorsqu'elles ont été produites par des institutions étatiques).

Le type de bases de données utilisé est souvent lié aux objectifs de l'étude.

- L'objectif peut être de comprendre les interactions qui existent entre la température mesurée à une station et les surfaces alentours. Dans ce cas les surfaces en jeu sont limitées et il est tout-à-fait envisageable de qualifier manuellement la donnée à partir d'images satellitaires ou de relevés de terrain (Krüger et Givoni 2007 ; Yan et al. 2014) ;
- Il peut être de cartographier les différences de température à l'échelle d'une zone urbaine. Ces dernières sont dans ce cas calculées pour des points disposés sur l'ensemble du territoire à partir des modèles de RG préalablement établis. Les indicateurs géographiques doivent donc être calculés sur un territoire beaucoup plus vaste et des méthodes de traitement automatique d'images s'imposent.
- L'objectif peut être d'appliquer les modèles de RG établis à partir de quelques sites d'une zone urbaine à d'autres zones urbaines. Dans ce cas il est nécessaire que les bases de données soient homogènes entre toutes les zones urbaines (tant pour les bases de données morphologiques que les bases de données d'occupation des sols).

#### 1.2.4.2 Forme et taille de la zone d'influence

Les indicateurs géographiques thématiques qui ont fait l'objet de la section précédente sont tous calculés au sein d'une empreinte au sol de forme et de taille donnée. Les objets géographiques contenus au sein de cette empreinte contribuent à expliquer la température de l'air mesurée par le capteur. Ceci peut être conceptualisé par la notion de *footprint*, illustrée par la figure 1.16 : la surface contributive (A) explique par exemple 50% de la température de l'air mesurée par le capteur. Sa taille, sa forme et sa position dépendent de la direction et de la vitesse du vent, de la hauteur à laquelle est placé le capteur ainsi que de la présence d'obstacles à l'écoulement de l'air.

La figure 1.17 présente les différentes façons d'intégrer cette notion de footprint. Par simplification, de nombreuses études ne différencient pas la contribution des surfaces selon leur distance au capteur. Ainsi, les indicateurs géographiques thématiques sont calculés au sein d'une empreinte au sol supposée expliquer 100% de la température. La taille de l'empreinte est choisie soit au vu des résultats de la littérature, soit en testant plusieurs tailles dans l'étude et en conservant celle qui fonctionne le mieux. Le choix final repose souvent sur une maximisation de la relation entre les indicateurs géographiques et les indicateurs de température locale (relation détaillée en section 1.2.5).

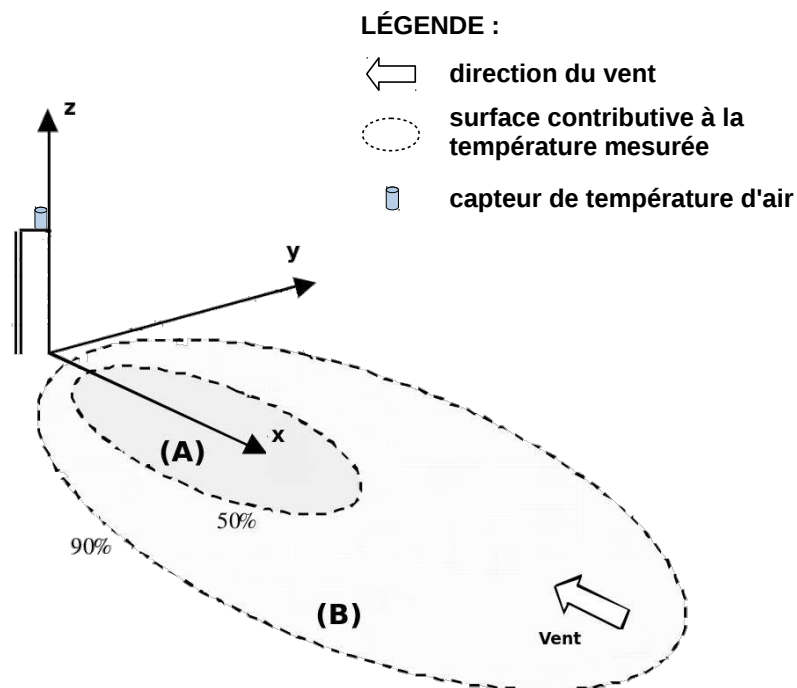


FIGURE 1.16 – Représentation conceptuelle de la notion de footprint, attribuant la température mesurée par le capteur à différentes surfaces alentour. La surface (A) explique par exemple 50% de la température, la surface (B) en explique 90% et la surface (B) - (A) en explique par conséquent 40%. La taille, forme et position de ces surfaces contributives à la température dépendent de la vitesse et de la direction du vent. Source : adapté de Oke (2004)

Quelques études considèrent la contribution des surfaces selon leur distance au capteur. On distingue pour cela deux techniques. La première utilise des indicateurs spatiaux indiquant la distance du capteur à des objets géographiques particuliers tels qu'une surface d'eau (Suomi et al. 2012), un parc (Yan et al. 2014) ou encore le centre (László et Szegedi 2012) ou la périphérie (Balázs et al. 2009) d'une ville. Ces indicateurs sont souvent utilisés en complément de l'approche décrite au paragraphe précédent. Cependant, ils posent souvent problème car les objets qu'ils utilisent sont pour la plupart difficiles à définir. La définition d'un parc est par exemple vague et sa distance d'action peut varier selon sa taille. Jauregui (1991) et Spronken-Smith et Oke (1998) ont estimé que la distance jusqu'à laquelle le refroidissement d'un parc pouvait être mesuré était de l'ordre de sa largeur (ou de son diamètre inscrit), mais sans décrire dans quelles proportions, à une distance donnée, le parc agissait sur la température de l'air. Le potentiel de rafraîchissement d'un parc dépend également des éléments présents dans son enceinte, mais aussi dans ses alentours proches. Le nombre, la forme, la taille et la répartition des arbres peut influencer sur la dynamique de la température d'air au sein du parc :



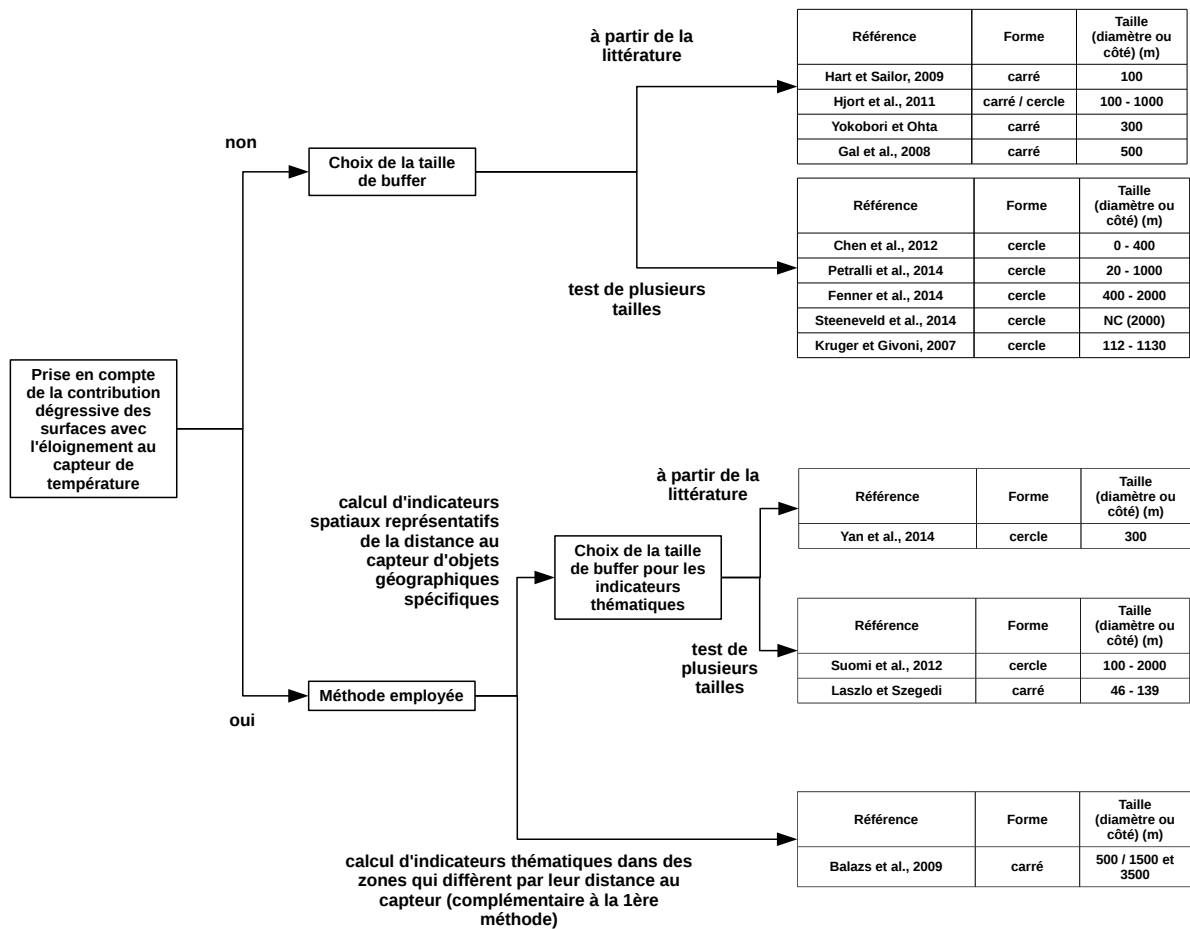


FIGURE 1.17 – Classification des solutions proposées dans la littérature pour identifier les surfaces qui contribuent à la température d'air mesurée par le capteur.

- un espace vert densément boisé permettra ainsi une réduction de la température dans le courant de l'après-midi (Hart et Sailor 2009) tandis que s'il est dépourvu d'arbres son pouvoir de rafraîchissement sera surtout nocturne (Spronken-Smith et Oke 1999) ;
- de hautes palissades ou une densité frontale élevée de bâtiment autour d'un espace vert diminueront quant à elles nettement le pouvoir de rafraîchissement nocturne d'un parc en limitant le flux  $L \uparrow$  (Spronken-Smith et Oke 1999).

Les définitions du centre urbain ou de la limite à la périphérie sont eux aussi difficiles à définir d'un point de vue géoclimatique. Balázs et al. (2009) découpent ainsi le territoire en mailles carrées et calculent la densité de sols artificialisés (bâtiments et sols imperméables) dans chacune d'elles. Ils considèrent qu'une maille est en périphérie de la ville lorsque la densité de sols artificialisés est inférieure à 0.05. Les surfaces en eau ne sont également pas toutes identiques en termes d'emprise au sol, de profondeur et de type. Si les rivières sont souvent considérées comme rafraîchissantes le jour comme la nuit (Hathway et Sharples 2012), la mer

(Suomi et al. 2012) ainsi que les canaux ou petits points d'eau contribuent à abaisser l'ICU estival la journée mais à l'augmenter la nuit (Steenefeld et al. 2014).

Une autre méthode pour prendre en compte la contribution des surfaces selon leur distance est de calculer des indicateurs thématiques au sein de plusieurs empreintes au sol plus ou moins distantes du capteur de température. Sur la figure 1.16, cela revient par exemple à calculer les indicateurs pour l'empreinte au sol (A) et l'empreinte au sol (B) - (A). C'est par exemple ce que font Balázs et al. (2009) qui définissent trois différentes surfaces : un carré de 500 m de côté, un carré de 1500 m de côté extrudé du premier et un dernier carré de 3500 m de côté extrudé du deuxième. Ils montrent que le carré extrudé de 1500 m de côté est celui qui explique le plus l'ICU dans leur cas.

Le choix de la forme des zones explicatives est souvent contraint par la méthode choisie pour faire les mesures. Les mesures mobiles ne sont par exemple pas effectuées en un seul point mais le long d'un axe, sans forcément avoir prédéfini de sites. Ainsi le territoire est souvent découpé en un réseau de carrés, dans lequel s'insère le parcours de mesure. Toutes les mesures réalisées dans un carré sont supposées représentatives de la température à l'intérieur du carré.

À partir de cette analyse de la littérature existante, deux éléments intéressants sont à mettre en évidence.

- Aucune étude n'a utilisé d'empreintes au sol de taille variable selon le contexte géographique du capteur. Il est pourtant bien connu que cette taille est souvent plus faible dans les espaces fermés de centre-ville que dans les espaces ouverts de périphérie (Oke 2004).
- Si un capteur est situé à l'interface de deux zones très hétérogènes, les températures mesurées par le capteur pourront être très différentes selon la direction du vent. C'est notamment le cas si un capteur est placé en frontière d'un parc situé dans un contexte urbain densément bâti. La température sera différente selon si l'air vient du parc ou s'il vient de la zone urbaine dense. Il paraît donc primordial de vérifier que la température d'air ne varie pas trop avec la direction du vent. Si c'est le cas, Stewart (2011) propose d'attribuer une zone explicative de la température différente en fonction de la direction du vent (en quelques sortes, séparer un site en deux).

## 1.2.5 Calibration et vérification des modèles

### 1.2.5.1 Calibration des modèles

Les indicateurs géographiques calculés doivent permettre de décrire les variations spatiales de l'ICU au sein d'une zone urbaine. Pour cela, des relations sont établies entre les indicateurs géographiques et les indicateurs de température locale. Trois méthodes sont majoritairement utilisées dans la littérature, reposant toutes sur des modèles de régression.

- La plus courante consiste à relier chaque  $ITL$  à un (ou plusieurs)  $IG_g$  par une relation linéaire simple (ou multiple).
- La deuxième méthode, la Méthode de Régression Généralisée (MRG) est en fait une généralisation de la première méthode à des variables explicatives ayant une distribution qui ne suit pas une loi normale. Cette méthode nécessite souvent un échantillon plus large pour fonctionner, même si Hjort et al. (2011) l'utilisent par exemple avec seulement 26 sites.
- La troisième méthode repose sur l'apprentissage par arbre de décision (Hart et Sailor 2009 ; Hjort et al. 2011). Hjort et Marmion (2008), qui utilisent cette méthode, ont montré qu'afin d'atteindre un optimum de prédiction, plus de 100 sites de mesure devaient être utilisés. Elle nécessite donc elle aussi un nombre plus important de sites de mesure qu'une méthode de régression linéaire simple ou multiple. Même si ce constat est relativement vrai pour de nombreuses méthodes statistiques, les méthodes reposant sur les arbres de décision semblent moins efficaces que la méthode des MRG lorsque l'échantillon est de taille plus faible (Hjort et al. 2011).

D'une manière générale, les relations qui utilisent de nombreuses variables explicatives expliquent mieux les variations spatiales de l'ICU que celles qui en utilisent peu. Cependant, ces relations sont très souvent statistiquement peu significatives (sauf si l'échantillon est très grand). Elles sont donc difficilement utilisables pour n'importe quel point de la ville et sont encore moins extrapolables à d'autres villes.

### 1.2.5.2 Vérification des modèles

Les relations établies à partir des indicateurs géographiques et de température locale peuvent être utilisées à l'échelle de la zone urbaine, mais également pour d'autres zones urbaines.

### Estimation de l'ICU pour la même zone urbaine

Les modèles permettent d'estimer la température de n'importe quelle zone de l'agglomération. Pour cela, il est nécessaire de calculer les variables explicatives (indicateurs géographiques) dans chaque zone d'intérêt. La forme et la taille de ces zones varient fortement d'une étude à l'autre mais cet aspect n'est pas le sujet de cette section. Il sera donc traité plus en détail au chapitre 3.

Pour vérifier si les relations établies pour un territoire sont satisfaisantes à l'échelle du territoire considéré, il convient d'utiliser un jeu de données de mesure qui n'a pas été utilisé pour calibrer le modèle, c'est-à-dire :

- pour les modèles temporels : utiliser les mêmes stations de mesure que pour la calibration mais pendant une période différente ;
- pour les modèles spatiaux : utiliser la même période que pour la calibration mais des sites de mesure différents ;
- pour vérifier simultanément les modèles spatiaux et temporels : utiliser un réseau de mesure différent de celui utilisé pour la calibration tant dans le choix de la période de mesure que dans le choix des sites.

Dans tous les cas, le matériel ainsi que les méthodes de mesure mis en place pour la campagne de calibration doivent être les mêmes que pour la campagne de vérification.

Peu d'études publient la vérification de leurs modèles. Lorsque c'est le cas, les mêmes méthodes et le même matériel de mesure sont utilisés pour la calibration et pour la vérification, ce qui est plutôt jugé comme positif. Aucune vérification des modèles temporels seuls n'est réalisée. Toutes les vérifications sont appliquées en utilisant d'autres jeux de données (vérification des modèles spatiaux) et pour une nouvelle période de mesure (sauf Hjort et al. (2011) qui utilisent la même période de mesure pour la calibration que pour la vérification). En termes d'indicateurs de performances des modèles, Hjort et al. (2011) calculent le coefficient de détermination entre les valeurs observées et les valeurs modélisées et obtiennent des valeurs proches de 0,6 au printemps et en hiver. Svensson et al. (2002) tracent les valeurs estimées en fonction des valeurs observées. Les valeurs hautes de l'ICU semblent légèrement sur-évaluées tandis que les valeurs basses semblent être sous-évaluées. Ils calculent le coefficient de détermination de leur nuage de point par rapport à la droite d'équation  $y = x$ . Sa valeur est de 0,41. À noter qu'aucune étude ne calcule l'erreur quadratique moyenne entre les valeurs observées et celles modélisées pour qualifier les performances de son modèle.

### Estimation de l'ICU pour d'autres zones urbaines

Les relations établies entre indicateurs géographiques et indicateurs de température locale peuvent être extrapolées à d'autres zones urbaines. Cela permet de connaître la variation spatiale de leur ICU à partir d'une station de mesure seulement.

Pour évaluer les performances d'extrapolation de modèles géoclimatiques, il est nécessaire d'utiliser un jeu de données de température provenant d'une autre zone urbaine. Les températures sont modélisées à partir des relations issues de la calibration et de la température d'une station de référence.

Trois articles décrivent l'extrapolation de modèles spatiaux. Hjort et al. (2011) extrapolent leurs modèles mais ne réalisent pas de vérification. Balázs et al. (2009) ainsi que Chen et al. (2012) utilisent un jeu de données acquis sur le territoire d'une autre agglomération mais basé sur les mêmes méthodes d'acquisition que pour la calibration (méthodes de mesure mobile). Chen et al. (2012) calculent une erreur quadratique moyenne de 0,48°C entre leur modèle et les données. Balázs et al. (2009) établissent une relation linéaire entre leur modèle et les valeurs observées. Le coefficient de détermination de cette relation est de l'ordre de 0,8. La pente de la droite est inférieure à un, ce qui implique que les valeurs les plus hautes sont sous-estimées par le modèle et les valeurs basses sur-estimées. L'ordonnée à l'origine est légèrement supérieure à 0, ce qui implique que les valeurs moyennes seront légèrement sur-estimées.

### 1.2.6 Positionnement de la thèse

Des données issues d'un réseau de capteurs existant mis en place par l'IRSTV<sup>4</sup> seront directement exploitées. Ces données de température ont été obtenues à partir de mesures fixes. Le matériel, les emplacements et les périodes de mesure seront décrits plus en détail au chapitre 2.

Un site de référence sera utilisé afin de modéliser par régression géoclimatique les variations spatiales de l'ICU. Ce site sera également présenté au chapitre 2.

La chaîne de traitement présentée figure 1.11 sera utilisée pour traiter les aspects temporels de l'ICU. Le choix des périodes d'intérêt pour le calcul d'indicateurs de température locale ainsi que le choix des outils statistiques (dont la revue de la littérature est schématisée figure 1.12) seront décidés et justifiés au chapitre 3. La différenciation des jours chômés et des jours travaillés, utilisée pour faire apparaître l'influence des émissions de chaleur d'origine anthropique sur l'ICU, ne sera pas réalisée afin de ne pas ajouter trop de complexité à l'analyse. Des modèles saisonniers plutôt que mensuels seront adoptés, résultat d'un compro-

---

<sup>4</sup> Institut de recherche en sciences et techniques de la ville : institut dans lequel je réalise mon doctorat

mis entre qualité ( finesse de résolution temporelle) et quantité (quantité suffisante de données pour que l'utilisation de techniques statistiques soit pertinente). L'influence des conditions météorologiques sur les variations temporelles de l'ICU sera traitée à partir de lois continues. La méthode proposée aura l'originalité d'être valide pour n'importe quel point de l'agglomération, contrairement à ce qui est pour l'instant proposé dans la littérature.

Un des principaux objectifs de cette thèse est d'expliquer les écarts de température observés à l'échelle d'une zone urbaine à partir d'informations géographiques. Des relations entre  $IG_g$  et  $ITL$  seront donc établies à cette fin. Bien que celles-ci soient empiriques, on veillera à ce qu'elles aient un sens physique. Pour cela, un indicateur géographique sera conservé seulement s'il peut être associé à un flux du bilan énergétique d'un volume urbain. Les relations entre  $ITL$  et  $IG_g$  seront établies à partir d'une seule agglomération et devront être valables pour n'importe quel point de son territoire. Pour cela, les méthodes de traitement devront être homogènes à tout le territoire. Des méthodes automatisées de traitement d'images satellitaires seront donc préférées à des méthodes manuelles souvent chronophages et sources d'hétérogénéités. Des méthodes de traitement simples et génériques seront privilégiées dans la mesure du possible, afin de rendre les tâches facilement reproductibles pour n'importe quelle agglomération. Dans le même souci de pouvoir utiliser les relations géoclimatiques pour d'autres agglomérations, des bases de données homogènes à l'échelle de la France métropolitaine doivent être choisies. Des bases de données produites à l'échelle nationale seront donc préférées à celles produites à l'échelle locale. Les indicateurs thématiques sont souvent insuffisants pour expliquer les variations spatiales de l'ICU. Des indicateurs spatiaux tels que la distance au centre urbain sont parfois utilisés comme variable explicative de l'intensité de l'ICU (Suomi et al. 2012). L'utilisation de tels indicateurs permettra de pondérer l'ICU par rapport à la taille de la ville, ce qui peut s'avérer particulièrement utile pour calculer l'ICU pour des agglomérations de tailles diverses.

Dans le cas de notre étude, les jeux de données utilisés pour calibrer les modèles de régression contiennent seulement 10 échantillons. Dans ces conditions, les méthodes les plus appropriées sont les modèles les plus simples, à savoir des régressions linéaires simples et multiples. La méthodologie mise en place pour traiter le phénomène spatial et temporel d'ICU sera développée dans le chapitre 3.

Les modèles établis à partir du jeu de données de calibration seront vérifiés à partir d'autres jeux de données. On distinguera plusieurs types :

- ceux dont la période d'acquisition est identique à celle des données de calibration, pour

lesquels les stations sont situées au sein de la même agglomération mais à des emplacements différents ;

- ceux dont la période d'acquisition est différente de celle des données de calibration mais utilisant les mêmes sites de mesure ;
- ceux dont la période d'acquisition et les sites de mesure seront différents des données de calibration mais qui ont cependant été produites au sein de la même agglomération ;
- ceux utilisés pour vérifier l'extrapolation des modèles à d'autres agglomérations. Les données seront donc issues de mesures réalisées dans d'autres agglomérations. La vérification des modèles fera l'objet du chapitre 4.

## Les données météorologiques mesurées

Ce chapitre est consacré à la présentation des réseaux de capteurs utilisés et aux jeux de données qui en découlent. Trois zones d'études sont distinguées : les agglomérations urbaines de Nantes, Angers et de La Roche-sur-Yon. Des campagnes de mesure de température d'air et du type de temps ont été réalisées dans chacune de ces trois agglomérations, chacune à partir :

- d'un réseau destiné à mesurer les microclimats compris dans la zone d'étude (agglomération), constitué exclusivement de capteurs de température (par la suite appelé réseau de capteurs de température) ;
- d'un ensemble de capteurs destiné à mesurer le type de temps (vitesse et direction du vent, rayonnement solaire, température de l'air, etc.) à l'échelle d'une zone d'étude (par la suite appelé ensemble de capteurs météorologiques).

Le terme *jeu de données* est employé pour décrire une base de données qui combine des mesures issues d'un couple {réseau de capteurs de température / ensemble de capteur météorologiques} pendant une période donnée. Ainsi, en utilisant deux périodes de mesure dissemblables, une même combinaison de réseaux/ensembles de capteurs permettra de construire plusieurs jeux de données. L'intérêt du partitionnement de nos données en plusieurs jeux est d'en utiliser un pour établir des modèles (calibration), les autres pour vérifier le bon fonctionnement de ces modèles (vérification).



Dans une première partie, les zones d'études seront décrites brièvement. Ensuite, chacun des réseaux/ensembles de capteurs utilisés dans cette thèse sera décrit dans une deuxième partie. La troisième partie sera consacrée aux jeux de données de mesure construits à partir de ces réseaux. Les systèmes de mesure utilisés entre deux jeux de données peuvent différer. Certains sont moins précis ou engendrent des erreurs de mesure, qui peuvent se répercuter sur les modèles. Une quatrième partie consistera donc à comparer les systèmes de mesure entre eux et corriger les données potentiellement erronées.

## 2.1 Présentation des zones d'étude

Les trois zones d'étude de cette thèse sont trois des plus importantes unités urbaines de la région des Pays de la Loire (une région du grand ouest français) en termes de population : selon les chiffres du recensement réalisé en 2012 par l'Institut National de la Statistique et des Études Économiques (INSEE), les agglomérations de Nantes, Angers et La Roche-sur-Yon comptaient respectivement 606 640, 218 657 et 52 808 habitants. Leur situation géographique à l'échelle du territoire français est présentée figure 2.1.

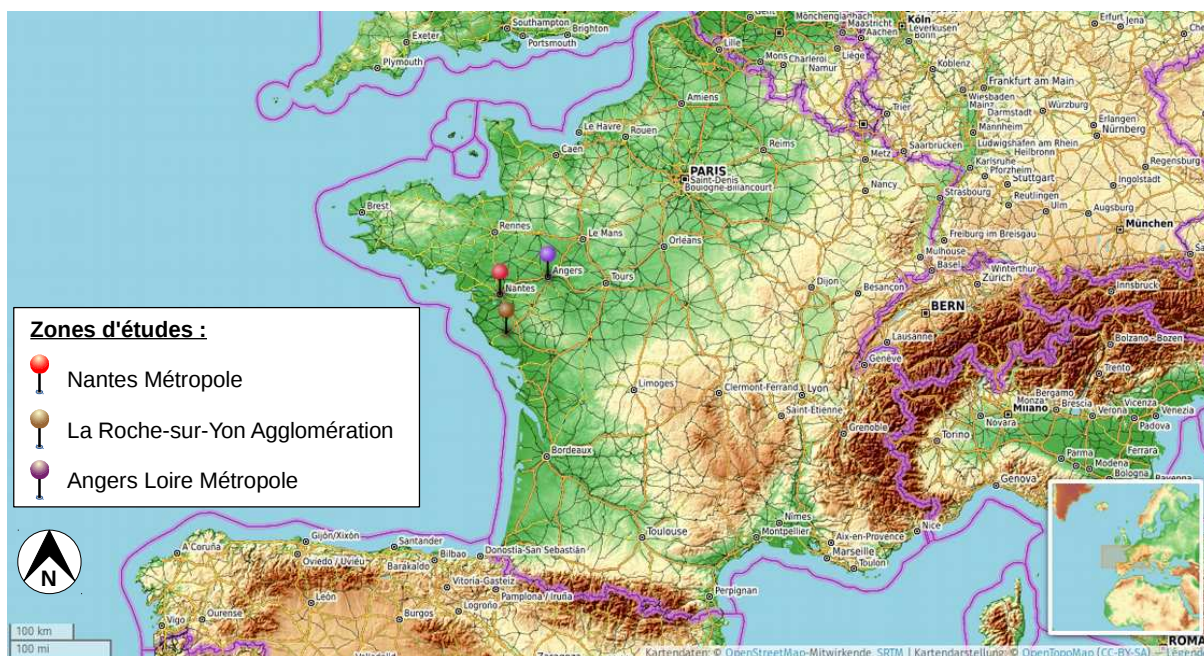


FIGURE 2.1 – Situation géographique des trois zones d'études

La superficie de leur territoire est du même ordre de grandeur, environ 500 km<sup>2</sup>. Chacune de ces agglomérations est constituée d'une partie urbaine compacte et d'une large couronne beaucoup plus rurale. Elles sont toutes les trois situées dans la zone climatique «Cfb»

selon la classification de Peel et al. (2007), où chaque lettre correspond à un critère climatique propre à ce type de région :

- C correspond à une zone à climat tempéré : la température moyenne du mois le plus chaud est supérieure à 10°C et celle du mois le plus froid comprise entre 0 et 18°C ;
- f correspond à une zone climatique sans saison sèche, c'est-à-dire qui n'a ni hiver sec (précipitations du mois hivernal le plus sec inférieures au dixième des précipitations du mois estival le plus humide), ni été sec (précipitations du mois estival le plus sec inférieures à 40 mm de pluie et également inférieures au tiers des précipitations du mois hivernal le plus humide) ;
- b correspond à une zone climatique à été chaud : la température moyenne d'au moins 10 mois est supérieure à 4°C mais la température moyenne du mois le plus chaud reste inférieure à 22°C .

Aucune des agglomérations n'est à proximité de massifs montagneux ou de zones topographiques pouvant impacter de manière importante le climat local. Enfin, les zones d'études de Nantes et de La Roche-sur-Yon sont toutes les deux situées à moins de 50 kilomètres de l'océan tandis que la ville d'Angers est située à une centaine de kilomètres. D'après Suomi et al. (2012), l'océan peut avoir un effet sur le climat local jusqu'à une vingtaine de kilomètres. Aucune des zones d'études ne devrait donc être plus perturbée que les autres par l'influence océanique. Au final, on peut conclure que les conditions météorologiques régionales de chacune de nos agglomérations sont comparables. Les différences de climat local qui pourront être observables seront donc imputables à des causes locales :

- la taille de la tâche urbaine : Balázs et al. (2009) ont montré empiriquement qu'il existait une relation entre l'ICU observé et le logarithme de la distance à la périphérie de la ville ;
- la morphologie qui peut différer d'une ville à l'autre (taille des bâtiments, largeur de rues, etc.) ;
- les types de sols urbains présents (superficie des surfaces en herbe, des surfaces arborées, des surfaces imperméables, etc.).

L'influence de ces trois paramètres sur le climat urbain est le coeur de cette thèse et fera l'objet du chapitre 3 traitant de la modélisation de l'ICU.

## 2.2 Présentation des réseaux ou ensembles de capteurs

Des capteurs sont regroupés au sein d'un réseau/ensemble de capteurs dès lors que :

- leurs données sont fournies ou produites par une même entité (exemple : Météo-France, l'Institut de Recherche en Sciences et Techniques de la Ville - IRSTV) ;
- leur installation répond à un même objectif ou a été réalisée dans le cadre d'une même campagne de mesure.

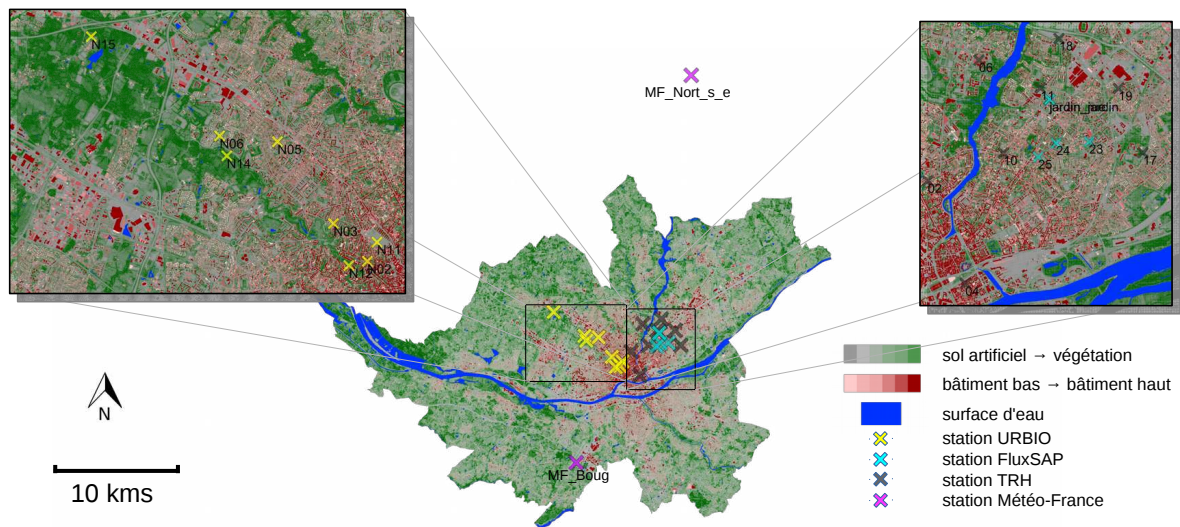
La disposition de chacun des réseaux/ensembles de capteurs utilisés dans cette thèse est représentée sur la figure 2.2. Une description détaillée de chacun d'eux est donnée dans les sections suivantes.

### 2.2.1 Réseaux de capteurs de température

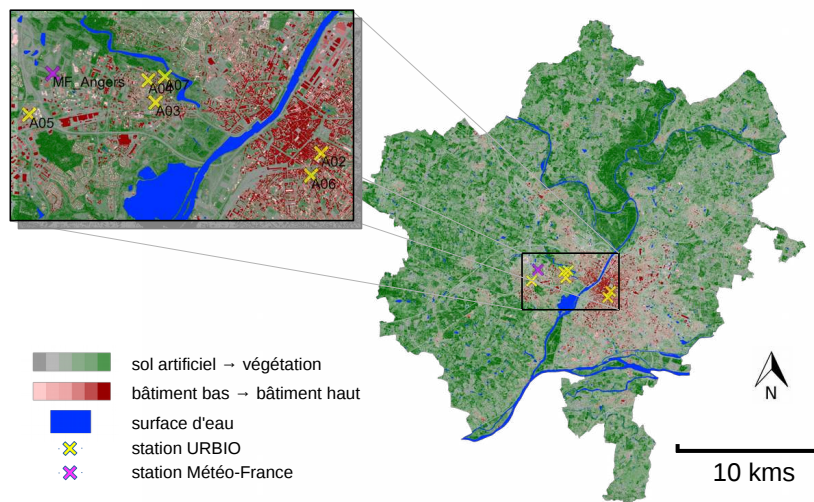
Les réseaux TRH, FluxSAP et URBIO sont tous les trois portés par l'Institut de Recherche des Sciences et Techniques de la Ville (IRSTV). Ils ont vocation à mieux comprendre les disparités spatiales et temporelles de la température et de l'humidité de l'air au sein d'une agglomération. Chacun a ses particularités. Le réseau TRH a la plus grande période de mesure (plusieurs années). Le réseau FluxSAP est concentré dans un quartier de Nantes (quartier du Pin Sec) et restreint à deux campagnes estivales (en 2010 puis en 2012). Les réseaux URBIO ont été déployés la même année sur plusieurs agglomérations (Nantes, Angers, La-Roche-sur-Yon). Les capteurs sont répartis le long d'un transect ville-campagne qui démarre du centre-ville et termine à sa périphérie.

Les derniers réseaux sont ceux de Météo-France. Les stations qui les composent permettent de comparer les différences de climat entre régions françaises. Elles permettent également de suivre l'évolution des conditions météorologiques de l'agglomération qu'elles représentent.

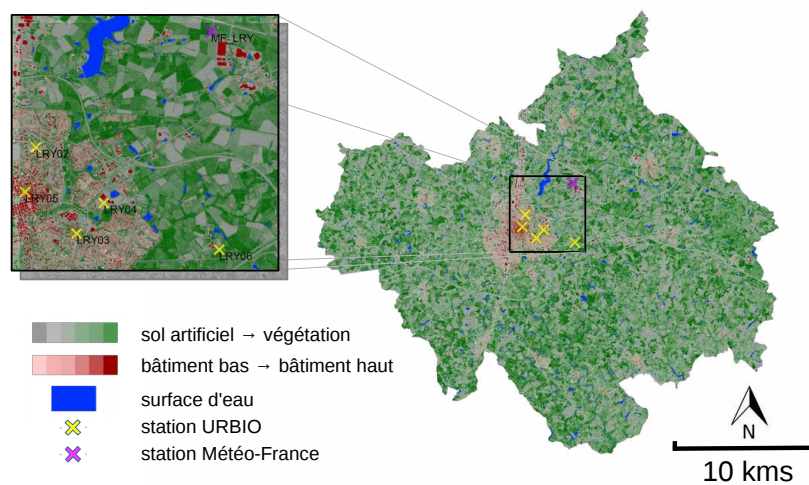
Le support sur lequel est fixé le matériel de mesure est propre à chaque réseau voire chaque site de mesure. De même, le matériel de mesure et d'acquisition n'est pas commun à tous des réseaux, ni même les méthodes d'acquisition de la donnée. Toutes ces différences sont détaillées dans les sections suivantes.



(a)



(b)



(c)

FIGURE 2.2 – Disposition des sites de mesure pour chacune des zones d'études : (a) agglomération de Nantes, (b) agglomération d'Angers, (c) agglomération de la Roche-sur-Yon

### 2.2.1.1 Positionnement des capteurs

Nous avons vu en section 1.2.1.2 que pour mesurer des températures représentatives du climat local dans la canopée urbaine, il était recommandé de fixer les instruments de mesure entre deux et cinq mètres du sol et à plus d'un mètre de toute surface verticale (Nakamura et Oke 1988). Dans des zones rurales plus ouvertes (respectant par exemple les conditions standards d'installation de Météo-France illustrées figure 1.9), cette hauteur peut être comprise entre 1,25 m et 2 m. Si les capteurs des réseaux URBIO et Météo-France respectent toutes ces recommandations, ce n'est pas le cas de tous les capteurs des réseaux TRH ou FluxSAP, qui sont en général placés à une trentaine de centimètres du mur de bâtiments (grâce à un bras de déport fixé au mur). Ce positionnement n'est donc pas conforme avec la condition d'espacement minimum d'un mètre de toute surface urbaine. Le tableau 2.1 rassemble toutes les informations qui concernent le positionnement des capteurs pour chacun des réseaux de capteurs de température.

Réseaux (agglomération)	TRH (Nantes)	FluxSAP (Nantes)	URBIO (Nantes, Angers, LRY)	MF Nantes	MF Angers	MF LRY
Nombre	8	5	19 (8, 6, 5)	2	1	1
Support	Variable (poteau, mur, poutres, etc.)	Variable (poteau, mur, poutres, etc.)	Poteaux (lampadaires ou poteaux en bois)	4 pieds - type trépied		
Hauteur des capteurs	Variable (2 à 5 m - toujours dans la canopée)	2,5 à 3 m	3 m	1,5 m		
Distance aux surfaces verticales	Variable (> 0,3 m)	Variable (> 0,3 m)	Variable (> 1 m)	espace ouvert (> 1 m)		

TABLE 2.1 – Informations sur le positionnement des capteurs de température pour chaque réseau

### 2.2.1.2 Description du matériel de mesure

Afin de limiter les biais de mesure liés au matériel, deux préconisations sont faites :

- utiliser des abris peu (dans l'idéal pas) sensibles au rayonnement solaire et bien ventilés ;
- utiliser un équipement homogène pour toutes les stations dont on utilise les données.

Dans notre cas, tous les capteurs des réseaux TRH et FluxSAP utilisent le même matériel, mais différant de celui utilisé par les réseaux URBIO et Météo-France (tableau 2.2).

Réseaux (agglomération)		TRH (Nantes)	FluxSAP (Nantes)	URBIO (Nantes, Angers, LRY)	MF Nantes	MF Angers	MF LRY
Capteur	Nom	HygroClip S (Rotronic)	HygroClip S (Rotronic)	HygroClip 2 (Rotronic)	N.C.		
	Technologie	Thermo- résistances PT100	Thermo- résistances PT100	Thermo- résistances PT100	Thermo-résistances PT		
	Résolution	0,01 K	0,01 K	0,01 K	0,1 K		
	Précision	0,3 K	0,3 K	0,1 K	Variable (0,1 K ou N.C.)	0,15 K	0,15 K
Abri	Nom	N.C.	N.C.	Davis 7714*	Socrima "modèle réduit"***	N.C.	Socrima "modèle réduit"
	Ventilation	Naturelle					
	Forme	Cylindrique à 9 coupelles	Cylindrique à 9 coupelles	Pavé (7 coupelles)	Pavé (parois à per- siennes)	N.C.	Pavé (parois à per- siennes)
	Taille	Φ 85 x 90 <i>mm</i> <sup>3</sup>	Φ 85 x 90 <i>mm</i> <sup>3</sup>	215 x 191 x 152 <i>mm</i> <sup>3</sup>	N.C.		

\* <http://www.davis-meteo.com/7714-abri-meteo.php> [Accédé le 14/01/2016]

\*\* <http://education.meteofrance.fr/ressources-pour-les-enseignants/observer-et-mesurer/l-abri-meteo> [Accédé le 14/01/2016]

TABLE 2.2 – Caractéristiques du matériel de mesure de température d'air pour chaque réseau de capteurs

L'analyse des mesures réalisées avec les réseaux TRH et FluxSAP a permis de diagnostiquer le peu de fiabilité des abris utilisés face au rayonnement solaire. Le réseau URBIO, installé quelques années plus tard, utilise donc un abri Davis qui a montré de bonnes performances lors d'une expérience réalisée dans le désert par l'Organisation Météorologique Mondiale (Lacombe et al. 2011 ; Lacombe 2010). Le même type d'abri que ceux utilisés par Météo-France faisaient également partie de la comparaison et ont eux-aussi montré de bonnes performances face au rayonnement. Les performances des trois types d'abri utilisés seront analysées plus en détail en section 2.4.

### 2.2.1.3 Méthode et matériel d'acquisition des données

L'acquisition des mesures est réalisée différemment selon l'entité qui produit les données mais également selon le réseau concerné (tableau 2.3).

Réseaux (agglomération)	TRH (Nantes)	FluxSAP (Nantes)	URBIO (Nantes, Angers, LRY)	MF Nantes	MF Angers	MF LRY
Méthode d'acquisition	Relevé manuel des enregistreurs	Relevé manuel des enregistreurs	Envoi automatique via le réseau de téléphonie mobile	Donnée produite par Météo-France		
Fréquence des relevés	Bimestriel	Relevé manuel des enregistreurs	Biquotidien	-		
Période d'échantillonnage	15 minutes	5 minutes	15 minutes	1 h		
Synchronisation de l'échantillonnage	Non (heure de démarrage aléatoire)	Non (heure de démarrage aléatoire)	Oui (démarrage à un quart d'heure entière)	Oui (démarrage aux heures entières)		

TABLE 2.3 – Caractéristiques du matériel et des méthodes utilisés pour l'acquisition des données de température d'air de chaque réseau de capteurs

Les données de Météo-France sont récupérées directement auprès de Météo-France qui fournit des données au pas horaire. Les données issues des réseaux TRH et FluxSAP sont directement produites par l'IRSTV. Elles sont échantillonnées respectivement toutes les 15 et les 5 minutes et stockées temporairement par un enregistreur de données. Il est ensuite nécessaire d'aller relever les données toutes les 10 semaines et de relancer l'acquisition. De plus, cette acquisition ne sera pas synchronisée, c'est-à-dire que chaque station démarre à l'heure à laquelle elle a été reprogrammée. Une synchronisation des données doit donc être mise en place afin de pouvoir comparer les stations intra-réseau et inter-réseaux. Celle-ci est illustrée figure 2.3.

La synchronisation consiste à associer les données de mesure au quart-d'heure le plus proche. Les données TRH et FluxSAP ont été relevées environ tous les deux mois pendant toute la période de fonctionnement du réseau. Au final, il existe une incertitude liée à cette méthode de synchronisation. Celle-ci est due au fait que la température peut diminuer ou augmenter entre l'heure pleine et l'échantillon le plus proche de l'heure pleine. Les valeurs qui nous intéresseront

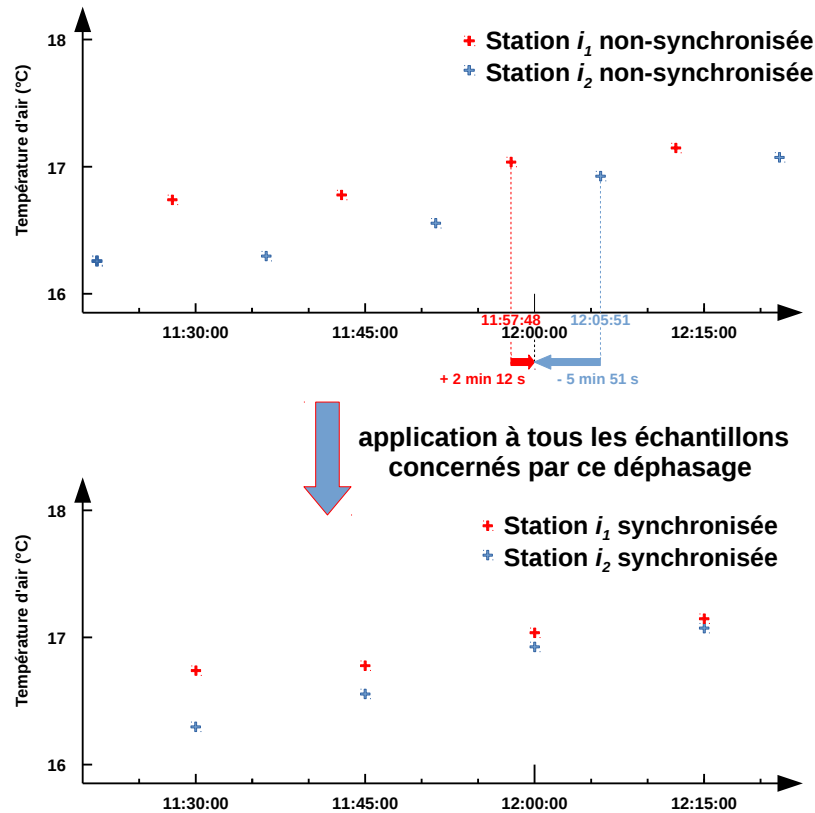


FIGURE 2.3 – Méthode employée pour synchroniser les données issues de deux capteurs de température des réseaux TRH ou FluxSAP

lors de la modélisation seront néanmoins des valeurs moyennées sur des périodes de quelques heures. Par effet de propagation d'incertitudes d'une moyenne, l'incertitude de synchronisation sera considérée comme négligeable par la suite. Les données du réseau URBIO sont également stockées sur des enregistreurs mais qui sont quant à eux relevés de façon automatique et bi-quotidienne par l'envoi d'un message à un serveur central via le réseau de téléphonie mobile. La température est échantillonnée toutes les 15 minutes avec synchronisation à chaque quart de l'heure entière.

Toutes nos données sont ensuite ré-échantillonnées au même pas de temps, c'est-à-dire que seules les données relevées à un multiple de ce pas de temps sont conservées. Une interpolation linéaire est réalisée pour compléter d'éventuelles données manquantes (données pour lesquelles la période d'échantillonnage d'origine serait supérieure à la période d'échantillonnage finale). Celle-ci est définie par l'équation 2.1.

$$T(t_{interp}) = T(t_1) + \frac{T(t_2) - T(t_1)}{3600} \tau \quad (2.1)$$

avec  $T(t_{interp})$  le résultat de l'interpolation linéaire des températures  $T(t_1)$  et  $T(t_2)$  à l'instant



$t_{interp} = t_1 + \tau$  (tel que  $t_1 < t_{interp} < t_2$ ). Les temps sont ici en secondes.

Un pas de temps commun à tous les réseaux doit être utilisé. Ceci implique de trouver un compromis entre les pas de temps de 5, 15 et 60 minutes. Utiliser le pas de cinq minutes nécessite d'interpoler fréquemment et donc d'utiliser d'importants volumes de données sans entraîner de réel gain d'information. Le pas de temps d'une heure peut s'avérer long pour étudier l'ICU (temps pendant lequel l'ICU peut évoluer rapidement entre une situation à l'ombre et une situation au soleil). Le pas de temps de 15 minutes est donc choisi comme pas de temps commun à tous les réseaux. Les données Météo-France, qui sont acquises au pas horaire, sont donc interpolées linéairement pour combler les valeurs manquantes.

Toutes ces données sont finalement stockées dans des fichiers au format *Comma-separated values* (csv). Ce format, lisible avec de nombreux logiciels quel que soit le système d'exploitation, facilite leur traitement a posteriori.

#### 2.2.1.4 Maintenance et calibration du matériel

La fréquence de maintenance et de calibration du matériel est une source supplémentaire d'hétérogénéité entre les réseaux de mesure. Les abris de Météo-France sont en effet nettoyés une fois par an (contre deux à six fois pour les autres réseaux) et repeints tous les deux ans (jamais pour les autres). Les capteurs de température ne sont pas ré-étalonnés après leur mise en place, et ce quel que soit le réseau. Aucun moyen de post-calibrer ces capteurs n'existe non plus. Sans vérification régulière, tous les capteurs de température ont potentiellement pu dériver dans le temps, ce qui pourrait conduire à une mauvaise interprétation des écarts de température. Cependant, les sondes de platine utilisées sont réputées très stables dans le temps. On fait donc l'hypothèse que les sondes PT100 utilisées par les capteurs de température ont une erreur négligeable devant l'ordre de grandeur de l'ICU qui sera mesuré. Ceci est d'autant plus vrai pour les capteurs du réseau URBIO qui ont fonctionné moins d'une année depuis leur calibration.

#### 2.2.1.5 Conclusion sur les réseaux de capteurs de température

Les réseaux TRH et FluxSAP se distinguent nettement des réseaux URBIO et Météo-France. Si une station des réseaux TRH ou FluxSAP était située sur le même site qu'une station des réseaux Météo-France ou URBIO, il serait fort probable qu'elles ne mesurent pas les mêmes températures. Ceci peut s'expliquer par deux raisons :

- l'incertitude sur les appareils de mesure. On notera que celle-ci est d'ailleurs plus élevée

pour les capteurs des réseaux TRH et FluxSAP (en raison des instruments de mesure moins précis et d'une méthode de synchronisation qui engendre des incertitudes) ;

- les erreurs de mesure qui sont probablement plus importantes pour les capteurs des réseaux TRH et FluxSAP et qui sont dues aux capteurs et aux abris utilisés, ainsi qu'à leur positionnement par rapport aux surfaces urbaines.
  - Les instruments de mesure utilisés n'ont pas été calibrés depuis leur mise en service (ceci est particulièrement vrai pour les réseaux TRH et FluxSAP qui ont fonctionné pendant plusieurs années). Ces erreurs ne peuvent malheureusement pas être corrigées.
  - Les abris ventilés des réseaux TRH et FluxSAP sont plus sensibles au rayonnement solaire que ceux utilisés par les réseaux Météo-France et URBIO. Ces erreurs peuvent néanmoins être estimées. Ce sera l'objet de la section 2.4.
  - Les capteurs des réseaux TRH et FluxSAP sont positionnés trop près des murs au regard des recommandations de Oke (2004). Ces erreurs peuvent néanmoins être estimées et seront également l'objet de la section 2.4.

### 2.2.2 Ensembles de capteurs météorologiques

Les ensembles de capteurs dits *météorologiques* mesurent tout un panel de grandeurs physiques descriptives des conditions atmosphériques. Afin de pouvoir comparer les analyses menées sur chacune de nos zones d'études, il est préférable que les mêmes grandeurs physiques soient relevées dans les mêmes conditions pour toutes les zones d'études. Ceci implique l'utilisation des stations du réseau de Météo-France, permettant par la suite de généraliser nos observations à d'autres agglomérations françaises. Les principales grandeurs physiques utilisées pour des applications de microclimatologie urbaine sont : la hauteur de pluie, la direction et la vitesse du vent, la température de l'air, le rayonnement global, l'humidité relative, la pression atmosphérique, la nébulosité totale. Elles sont chacune décrites plus en détail dans le tableau 2.4. Les définitions et les méthodes de mesure sont issues du standard Météo-France<sup>1</sup>.

Ces grandeurs physiques sont toutes mesurées pour chaque agglomération. Elles seront supposées représentatives des conditions météorologiques observables à l'échelle de la zone d'étude (mésos-échelle).

---

<sup>1</sup> disponible sur la page [http : //publitheque.meteo.fr/okapi/composantsHTML/listeParametres/infoParam.jsp?id=DH](http://publitheque.meteo.fr/okapi/composantsHTML/listeParametres/infoParam.jsp?id=DH) [Accédé avec mot de passe le 31/10/2014]

Nom	Description	Unité	Méthode d'échantillonnage	Hauteur de mesure (en m)
Hauteur de pluie	Hauteur de pluie (ou de fusion de la neige)	millimètres	Total recueillie dans l'heure précédente	N.C.
Direction du vent	Direction du vent, de 0 (vent calme) à 360 ° (vent du nord) dans le sens horaire	Degrés	Direction moyennée sur les 10 minutes précédant l'heure ronde	10 - 11
Vitesse du vent	Vitesse du vent	mètres par seconde	Vitesse moyennée sur les 10 minutes précédant l'heure ronde	10 - 11
Température de l'air	Température de l'air mesurée sous abri	degrés Celsius	Relevée à l'heure ronde	1,5
Rayonnement global (ou rayonnement solaire descendant)	Il s'agit de l'ensemble des rayonnements solaires (direct et diffus) parvenant de toutes les directions sur une surface horizontale	Joules par centimètre carré	Total recueilli l'heure précédente (UTC)	N.C.
Humidité relative	Humidité relative de l'air sous abri	Pourcentage	Relevée à l'heure ronde	1,5
Pression	Pression atmosphérique à l'altitude de la station	hectopascal	Relevée à l'heure ronde	27 - 91
Nébulosité totale	Observation humaine de la couverture de la voûte céleste par des nuages	Octa ([0-8] : huitième de voûte céleste couverte, 9 : ciel invisible notamment par brouillard)	Mesurée dans les 10 minutes qui précèdent l'heure ronde	N.C.

TABLE 2.4 – Définition des principales grandeurs physiques météorologiques mesurées par Météo-France

Les stations Météo-France de Beaucouzé et de La-Roche-sur-Yon sont utilisées respectivement pour les zones d'étude d'Angers et de La-Roche-sur-Yon. Deux stations Météo-France sont présentes dans un rayon de 30 kilomètres de la ville de Nantes : celle de Bouguenais et celle de Nort-sur-Erdre. Selon la grandeur physique considérée, l'une ou l'autre des deux stations sera utilisée.

La pression atmosphérique ainsi que la nébulosité ne sont mesurées qu'à Bouguenais. Bouguenais est donc retenue pour la mesure de ces grandeurs physiques.

Comme pour les mesures de température (cf. figure 1.9), une qualité d'emplacement et de capteur sont associées à la mesure de chaque variable météorologique. Quelle que soit la grandeur physique considérée, la qualité du capteur du site de Bouguenais est toujours meilleure ou égale à celle du site de Nort-sur-Erdre. La qualité de l'emplacement est quant à elle dépendante de la grandeur physique considérée : elle est meilleure à Bouguenais pour la température de l'air, l'humidité relative et la pluviométrie tandis qu'elle est moins bonne pour le rayonnement solaire, la vitesse et la direction du vent. Dans le cadre de ce travail de thèse, ces mesures sont utilisées pour estimer le type de temps à l'échelle de la zone d'étude concernée. La précision de la mesure n'est donc pas un critère déterminant quant au choix de la station utilisée. Excepté pour la température et l'humidité relative, les grandeurs physiques utilisées seront toutes issues de la station de Bouguenais, qui est la plus proche de la ville. Concernant la température de l'air, celle mesurée à la station de Bouguenais pour des vents provenant des directions nord-nord-est est en moyenne supérieure à celle de plusieurs stations du réseau TRH (situées en centre-ville de Nantes), ce qui n'est pas le cas pour les autres directions de vent. Ces mêmes stations ont par contre une température moyenne toujours largement supérieure à la station de Nort-sur-Erdre quelle que soit la direction du vent. La station Météo-France de Bouguenais est située sous le vent de la tâche urbaine et des pistes de l'aéroport Nantes Atlantique lorsque celui-ci vient du nord-est et sous le vent d'éléments plus ruraux lorsque le vent vient des autres directions. Selon la direction du vent, il se pourrait donc que cette station mesure tantôt un microclimat urbain, tantôt un microclimat rural. Pour cette raison, la station de Nort-sur-Erdre lui est préférée pour les mesures de température et d'humidité relative de l'air.

Les données de chacune des grandeurs mesurées sont disponibles au pas horaire. Afin de coller aux observations des réseaux de capteurs de température, les données correspondant à chacune des grandeurs physiques sont interpolées linéairement toutes les 15 minutes, sauf celles correspondant à la hauteur de pluie et à la direction du vent.

La hauteur de pluie est représentative de la quantité d'eau accumulée pendant une certaine période. C'est donc une variable extensive, par opposition à toutes les autres variables qui sont intensives. Ceci implique que l'interpolation linéaire telle que définie par l'équation 2.1 ne peut être utilisée pour cette variable (elle surestimerait les précipitations). L'alternative est de considérer que l'intensité des précipitations a été constante durant l'heure qui a précédé le décompte de hauteur de pluie. Cela implique que si l'on considère une observation horaire de hauteur de pluie  $HP(t_{ini})$  (pour l'heure  $t_{ini}=11:00$ ) que l'on veut transformer au pas quart-

horaire  $t_{fin}$ , il faut utiliser la relation suivante :

$$\begin{aligned}
 HP(t_{fin} = 10 : 15) &= HP(t_{fin} = 10 : 30) \\
 &= HP(t_{fin} = 10 : 45) \\
 &= HP(t_{fin} = 11 : 00) \\
 &= HP(t_{ini})/4
 \end{aligned}
 \tag{2.2}$$

L'interpolation linéaire ne peut pas non-plus être utilisée pour la direction du vent, car non-seulement les changements de direction de vent peuvent s'effectuer de manière assez brutale, mais surtout l'interpolation linéaire n'est pas adaptée à un raisonnement trigonométrique. La figure 2.4 illustre le problème auquel peut donner lieu l'interpolation linéaire appliquée à la direction du vent.

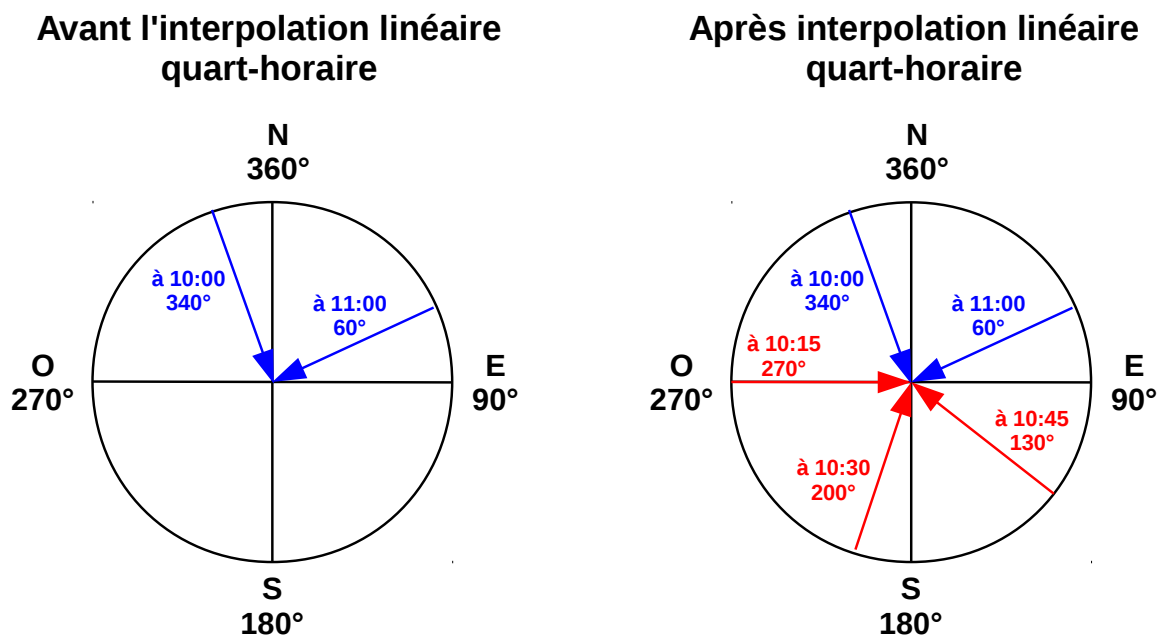


FIGURE 2.4 – Interpolation linéaire quart-horaire appliquée à un signal de direction de vent

La direction initialement nord-ouest à 10h tourne vers le nord-est à 11h. Entre ces deux instants (qui sont en fait des périodes puisque le vent est mesuré et moyenné pendant les 10 dernières minutes avant l'heure pile), la direction du vent n'est pas connue. On pourrait s'attendre intuitivement à ce que la direction du vent dérive lentement vers le nord-est dans le sens horaire (en passant donc par le cadran nord), ce qui n'est pas le cas si une interpolation linéaire est utilisée. Pour parer à ce problème, la direction du vent est considérée comme stable entre deux heures. Ainsi, dans notre exemple, la direction de vent à 10:15, 10:30 et 10:45 serait

constante et égale à celle de 11:00 (c'est-à-dire à 340°).

## 2.3 Présentation des jeux de données

Six jeux de données sont utilisés dans le cadre de cette thèse. L'un d'entre eux, appelé jeu de données de calibration et noté  $JDD_{calib}$ , est utilisé pour calibrer les modèles microclimatiques. Les cinq autres sont utilisés pour vérifier les estimations générées à partir du premier. On distingue les jeux qui sont produits pour l'agglomération de Nantes (agglomération utilisée pour la calibration) de ceux produits pour une autre agglomération (qui seront utilisés pour vérifier l'extrapolation des modèles). Les réseaux ou ensembles de capteurs ainsi que les périodes qui constituent ces jeux de données sont schématisés figure 2.5.

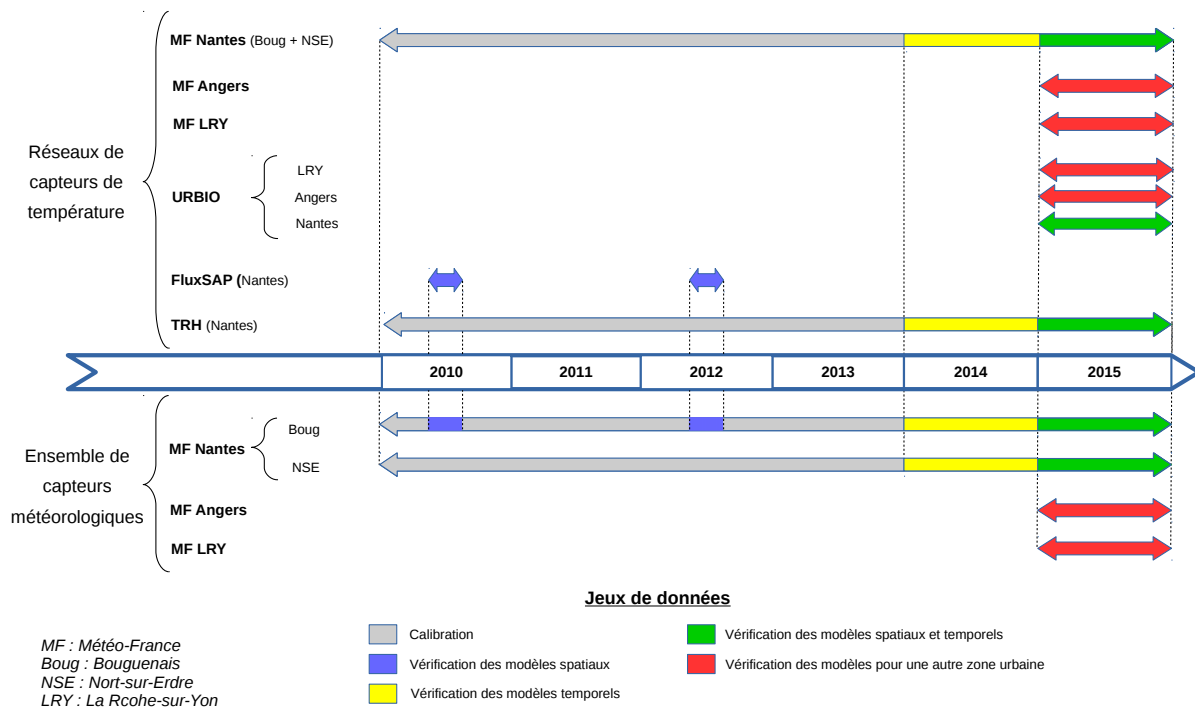


FIGURE 2.5 – Réseaux/ensembles de capteurs et jeux de données utilisés pour calibrer et vérifier la modélisation de l'ICU à l'échelle d'une zone urbaine

Que ce soit pour les réseaux de capteurs de température ou les ensembles de capteurs météorologiques, des valeurs manquantes peuvent apparaître dans les données (dues à la maintenance, à des problèmes techniques, etc.). Afin de comparer la température mesurée en chacun de nos sites, seules les périodes où tous les capteurs de température et tous les capteurs météorologiques fonctionnent sont utilisées. De plus, l'unité de référence pour l'analyse temporelle étant le jour (cycle naturel du soleil le plus court observé de la Terre), seuls ceux dont au-

cune donnée ne manque sur la journée entière sont utilisés. La liste des stations qui constituent chaque réseau de capteurs de température a été présentée section 2.2.1. Cette liste n'est en fait pas exhaustive. Seules les stations utilisées dans cette thèse ont été présentées. Si toutes les stations avaient été utilisées, il ne resterait que très peu de jours comportant des données complètes pour l'ensemble des capteurs). Le gain spatial (en nombre de points) engendré par l'utilisation de toutes les stations serait ainsi fait au détriment du gain temporel (le nombre de jours étudiés). Afin de trouver pour chaque jeu de données un bon compromis nombre de points de mesure / nombre de jours étudiés, toutes les combinaisons de stations sont constituées. L'évolution du nombre de jours utilisables en fonction du nombre de stations utilisées est ensuite tracée afin de choisir la combinaison de stations qui maximise à la fois le nombre de jours et le nombre de stations. La figure 2.6 représente cette évolution pour le jeu de données de calibration  $JDD_{calib}$ . Chaque croix bleue représente une combinaison de capteurs de température.

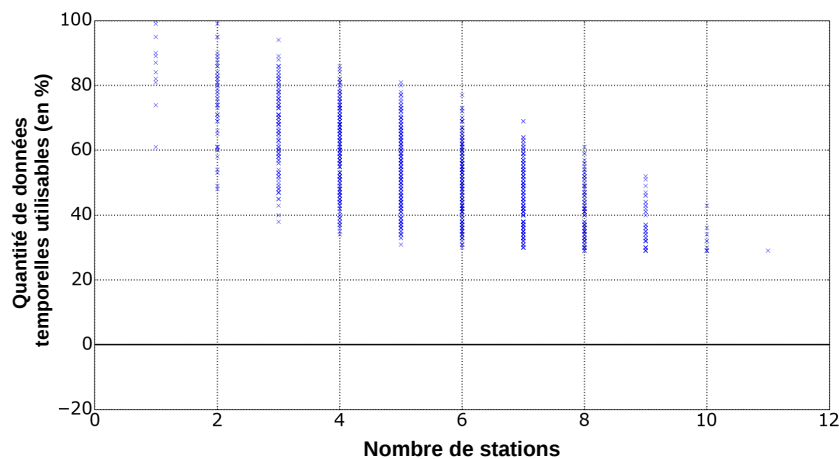


FIGURE 2.6 – Évolution de la quantité de données utilisables selon le nombre de stations de mesure utilisées ( $JDD_{calib}$ ). Une journée de données est comptée comme *utilisable* si aucune station n'a de valeur manquante.

Afin de conserver le maximum de journées pour un nombre de stations donné, il est donc préférable de privilégier les combinaisons de stations qui donnent les valeurs les plus hautes selon l'axe des ordonnées. À partir de deux stations, la perte de données utilisables observée pour chaque nouvel ajout d'une station est stable et est en moyenne de 7%. Entre les combinaisons comportant 10 et 11 stations, la perte de données est de l'ordre de 14%. Le gain d'une station est jugé trop faible pour la perte de données importante qu'il engendre. Ce raisonnement a été effectué pour chaque jeu de données, aboutissant à la constitution des réseaux de capteurs de température présentés en section 2.2.1.

En réalité, si cette méthode était appliquée à la lettre, peu de jours subsisteraient. En

effet, des périodes de 2h pendant lesquelles aucune valeur n'est enregistrée sont régulièrement rencontrées dans les données de nébulosité de la station de MF Bouguenais. La nébulosité est également rarement mesurée la nuit pour les sites d'Angers et de La Roche-sur-Yon. Pour éviter de supprimer toutes les journées qui ont des valeurs de nébulosité manquantes (plus de 90% des jours dans certains jeux de données), il faut combler ces trous dans les données. Pour cela, la méthode d'interpolation linéaire des données (telle que définie par l'équation 2.1) est utilisée.

Au final, un calendrier représentant les jours où les données peuvent être utilisées est établi pour chaque jeu de données. Une représentation des types de temps observés en fonction de la direction du vent est également produite. Pour cela, la procédure suivie est la suivante : un premier tri est réalisé afin de séparer les données météorologiques par saisons. La rose des vents est ensuite divisée en sixièmes. La direction du vent à chaque pas de temps d'une journée est placée dans un de ces sixièmes afin d'identifier la direction de vent prédominante (fréquence maximale) pour chaque jour du jeu de données. Chaque journée est ainsi associée à un sixième de rose des vents correspondant à la direction d'où vient le vent. Les autres variables météorologiques (toutes sauf la direction du vent) sont moyennées à l'échelle journalière. À chaque combinaison {saison  $S_s$  - part de rose des vents  $dir_{pr}$  - variable météorologique  $VM_m$ } est ainsi associée une série de  $nb_j$  valeurs correspondant à la valeur moyenne journalière de la variable  $VM_m$  pour les  $nb_j$  jours de la saison  $S_s$  où la direction de vent prédominante appartient à la part  $dir_{pr}$ . Un diagramme en étoile à six branches (chaque branche correspondant à un sixième de la rose des vents) est utilisé pour représenter graphiquement les résultats de chaque combinaison {saison  $S_s$  - variable météorologique  $VM_m$ }. Les valeurs qui apparaissent sont les centiles 25% et 75% de chaque série, sauf dans le cas de l'analyse de la direction de vent. Dans ce cas, le diagramme en étoile est associé à une rose des vents, la valeur indiquée correspondant à la fréquence (comprise entre 0 et 1) pour laquelle une direction de vent est observée.

Ces représentations permettent de mettre en évidence :

- la variabilité d'une variable météorologique par rapport à une direction de vent donnée (si les centiles de la vitesse de vent sont par exemple très espacés, cela démontre que la vitesse du vent est changeante lorsque le vent vient de cette direction - tantôt forte, tantôt faible) ;
- l'ordre de grandeur de la variable météorologique pour une direction de vent par rapport aux cinq autres (si les centiles de la vitesse du vent sont par exemple situés vers l'extérieur de la toile, cette direction de vent peut être associée à de fortes vitesses de vent) ;



- la variabilité des types de temps entre les saisons (les vents peuvent provenir plus souvent d'une direction de vent que d'une autre selon la saison, ou une même direction de vent peut ne plus être associée au même type de temps, etc.).

### 2.3.1 Jeu de données de calibration

Cette phase permet de calibrer les modèles temporels ainsi que les modèles spatiaux. Les réseaux TRH, MF Bouguenais et MF Nord-sur-Erdre sont utilisés comme réseau de capteurs de température. Les stations Météo-France choisies pour l'ensemble de capteurs météorologiques sont MF Nord-sur-Erdre pour la température et l'humidité et MF Bouguenais (pour les autres variables météorologiques). Les journées qui pourront être utilisées pour la calibration sont en vert sur la figure 2.7. Un tri saisonnier est réalisé pour rendre compte du nombre de jours exploitables par saison.

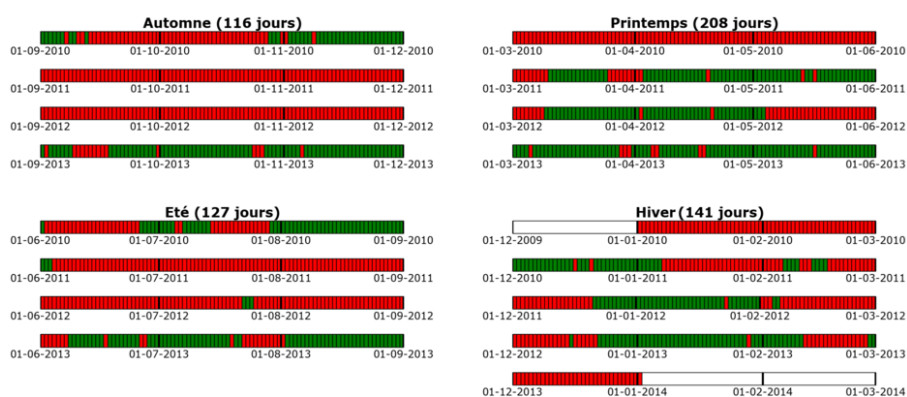


FIGURE 2.7 – Répartition saisonnière des journées du jeu de données de calibration utilisées

L'automne et l'été ont nettement moins de jours que le printemps et dans une moindre mesure que l'hiver. Ceci s'explique par la quasi absence de données à ces saisons en 2011 et en 2012.

L'analyse des types de temps (figure 2.8) permet d'observer les directions principales de vent et les conditions météorologiques qui leur sont associées.

- Les directions principales de vent diffèrent selon les saisons : ils viennent principalement de la moitié nord en été et au printemps, de la moitié ouest en automne tandis qu'aucune tendance ne se dégage pour l'hiver.
- Les précipitations viennent le plus souvent avec des vents du quart sud-ouest (donc de l'océan) quelle que soit la saison. Cette direction est donc sans surprise associée à une

forte nébulosité et un rayonnement solaire souvent faible. Les vents les plus forts proviennent également de cette direction.

- Les jours les plus clairs (nébulosité faible et rayonnement solaire fort) sont souvent associés à des directions de vent nord et nord-est quelle que soit la saison.
- Les directions de vent associées aux températures les plus élevées ne sont par contre pas communes à toutes les saisons : la chaleur provient du tiers est / nord-est (des terres) en été, du tiers ouest / sud-ouest (de l'océan) en hiver et aucune réelle tendance n'est observable pour le printemps et l'automne.

D'une manière générale, le matériel, le positionnement des stations et la durée des campagnes de mesure utilisés pour la calibration sont similaires à un bon nombre des études microclimatiques décrites dans le chapitre 1 (cf. tableau 2.5). Ceci ne permet cependant pas de conclure sur la fiabilité de nos données ou de nos modèles, puisque le nombre de stations est faible, les capteurs sont trop proches des murs et n'ont jamais été ré-étalonnés depuis leur implantation sur site. De plus, les abris utilisés génèrent d'importants biais de mesure causés par le rayonnement solaire. Pour tenter de pallier ces limites, plusieurs mesures seront adoptées.

- Le nombre de stations sera artificiellement augmenté au chapitre 3 en considérant qu'une station peut être dissociée en six stations selon la direction du vent.
- On considérera dans la suite de ce travail que l'erreur due à la dérive des sondes PT100 est négligeable devant les différences de température à mesurer.
- L'influence de la proximité des capteurs avec le mur et les performances des abris utilisés seront étudiés plus en détail en section 2.4. À noter que des 17 études présentées dans ce travail, seuls Fenner et al. (2014) comparent les températures mesurées sous chacun des abris à celle mesurée sous un abri à ventilation forcée pris comme référence. Ils montrent que l'erreur moyenne est maximale en période diurne mais qu'elle reste faible (inférieure à 0,4°C pour tous leurs abris). L'écart-type de l'erreur est également du même ordre de grandeur.

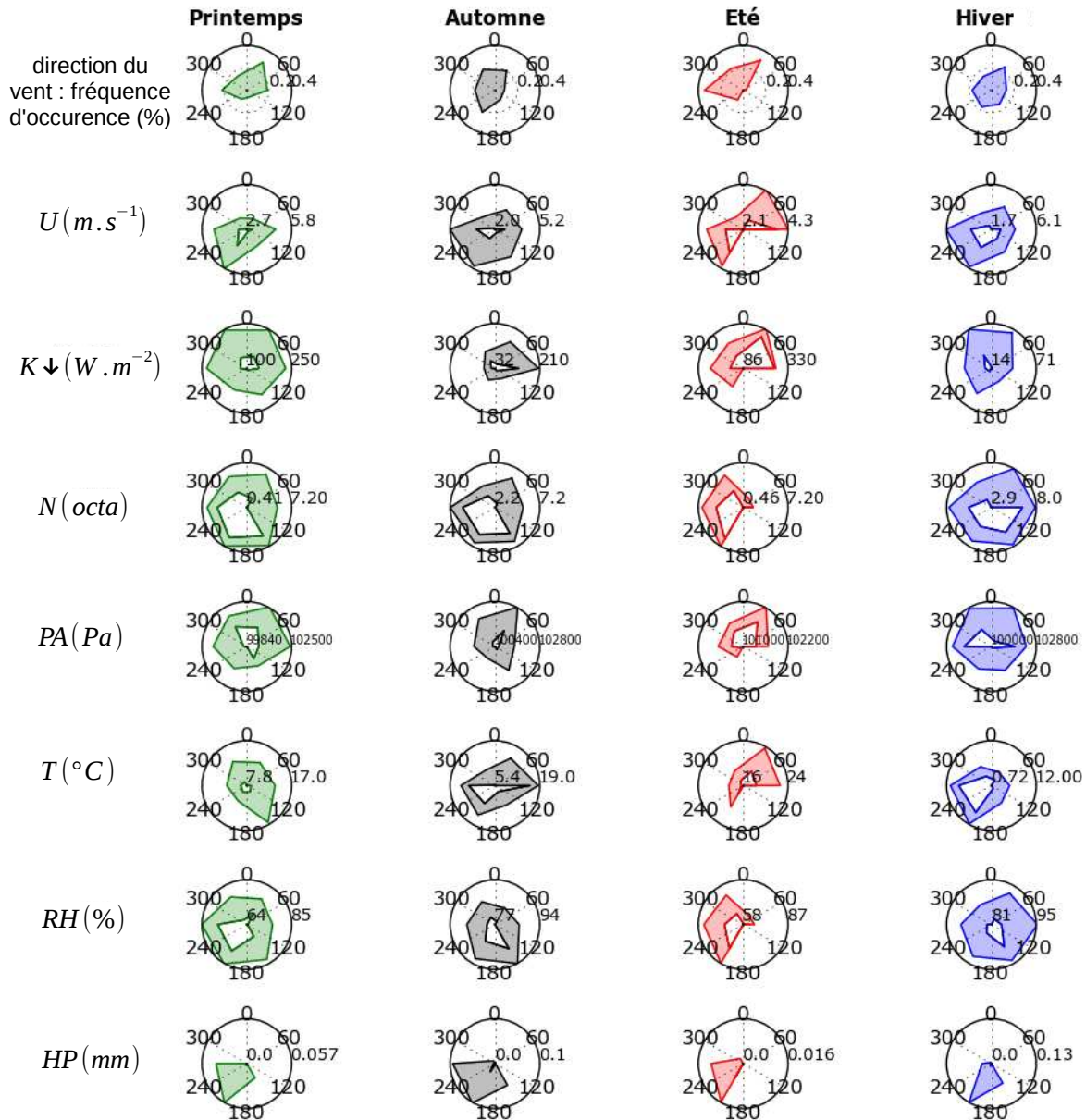


FIGURE 2.8 – Répartition du type de temps selon la saison et la direction du vent pour le jeu de données de calibration. Ligne continue extérieure : pourcentile 0,75 de la variable  $VM_m$ , ligne continue intérieure : pourcentile 0,25 de la variable  $VM_m$

Étude	Méthode de mesure	Matériel				Précision sur le positionnement du capteur par rapport aux surfaces verticales	Précision sur le positionnement du capteur par rapport au sol	Durée des campagnes
		Nombre de stations	Précision du capteur	Description de l'abri	Intercomparaison			
État de l'art	Fixe (8 études*)	entre 2 et 50 (moyenne : 16, médiane : 10)	5 études décrivent l'abri utilisé, seule une compare les performances de l'abri par rapport à un abri à ventilation forcée	seules 2 études précisent que les capteurs viennent d'être certifiés ou ont été étalonnés avant mise en oeuvre	4 études à une distance > 0,5 m d'un mur  1 à moins de 0,5 m  N.C. pour 3 études	5 études ont toutes leurs stations entre 2 et 5 m du sol  2 ont une partie de leurs stations au-dessus de 5 m  N.C. pour 1 étude	0,2 K pour 5 études  N.C. pour les autres	3 études > 4 ans  4 études entre 1 et 4 ans  1 étude < 1 an
	Mobile (8 études**)	Station mobile	0,1 K à 0,2 K  N.C. pour 3 études	3 n'en utilisent pas  3 le décrivent  N.C. pour 2 études	4 utilisent un seul capteur  1 précise que les capteurs sortent d'usine et sont étalonnés  N.C. pour 3 études	Station mobile	7 études ont toutes leurs stations entre 1,5 et 2 m du sol  N.C. pour 1 étude	5 études > 30 jours de campagnes répartis sur plus d'une année  3 études < 30 jours de campagnes
Cette thèse	Mobile ( <i>JDD<sub>calib</sub></i> )	10	0,3 K	Oui (tableau 2.2)	Non	Oui (souvent < 50 cms de surfaces urbaines)	Oui (entre 2 et 5 m)	4 années

\* (Alonso et al. 2007 ; Erell et Williamson 2007 ; Fenner et al. 2014 ; Hjort et al. 2011 ; Krüger et Givoni 2007 ; Petralli et al. 2014 ; Steeneveld et al. 2014 ; Suomi et al. 2012)

\*\* (Balázs et al. 2009 ; Chen et al. 2012 ; Gal et al. 2009 ; Hart et Sailor 2009 ; László et Szegedi 2012 ; Oke 1976 ; Yan et al. 2014 ; Yokobori et Ohta 2009)

TABLE 2.5 – Caractéristiques du matériel utilisé et des méthodes employées relativement à 16 études issues de la littérature pour produire des données de température d'air

### 2.3.2 Jeux de données de vérification provenant de la même zone d'étude

Ces phases permettent de vérifier les modèles temporels et spatiaux établis auparavant à partir de jeux de données obtenus sur la même zone d'étude que les données de calibration (l'agglomération de Nantes). Les données de température utilisées sont donc acquises soit en d'autres points de l'agglomération, soit à d'autres périodes.

D'une manière générale, comme décrit section 1.2.5.2, peu d'auteurs ont publié des articles traitant de la vérification de leurs modèles sur le territoire où a eu lieu la calibration. Sur les 17 études décrites au chapitre 1, seulement trois en traitent : Chen et al. (2012), Hjort et al. (2011) et Svensson et al. (2002). Ils sont également les seuls à préciser que les capteurs de température ont été étalonnés avant la campagne de vérification. Toutes les vérifications utilisent des jeux de données qui diffèrent tant spatialement (d'autres sites de mesure que ceux utilisés pour la calibration sont utilisés) que temporellement (d'autres périodes de mesure que celles utilisées pour la calibration sont utilisées), sauf celle réalisée par Hjort et al. (2011), qui diffère seulement spatialement.

#### 2.3.2.1 Acquisition des données à une autre période

Cette phase permet de vérifier comment les modèles se comportent lorsqu'ils sont appliqués aux mêmes points du territoire mais pour des périodes différentes que celles utilisées pour la calibration. Presque toutes les stations choisies pour la calibration sont donc réutilisées (réseau TRH et réseaux MF Nort-sur-Erdre et MF Bouguenais) mais pour une période de mesure différente (2014 – 2016). Seule la station 10 du réseau TRH est exclue, faute d'une quantité suffisante de données. Les mêmes capteurs météorologiques sont également réutilisés (la station MF Nort-sur-Erdre pour la mesure de température et d'humidité de l'air et la station MF Bouguenais pour la mesure des autres variables météorologiques - cf. figure 2.5). Le jeu de données utilisé pour cette phase est noté  $JDD_{temp}$ .

Les jours où les données peuvent être utilisées sont beaucoup moins bien répartis entre les saisons pour le jeu de données  $JDD_{temp}$  que pour le jeu  $JDD_{calib}$ . En effet, comme l'illustre la figure 2.9, aucun jour complet de données n'est enregistré en hiver et seulement huit jours complets sont enregistrés au printemps. L'été et l'automne ont quant à eux des jeux de données plus complets (plus de 30 enregistrements). Les résultats de la vérification seront donc sans doute plus fiables pour ces deux saisons.

L'analyse des types de temps selon la direction de vent est moins évidente pour ce jeu

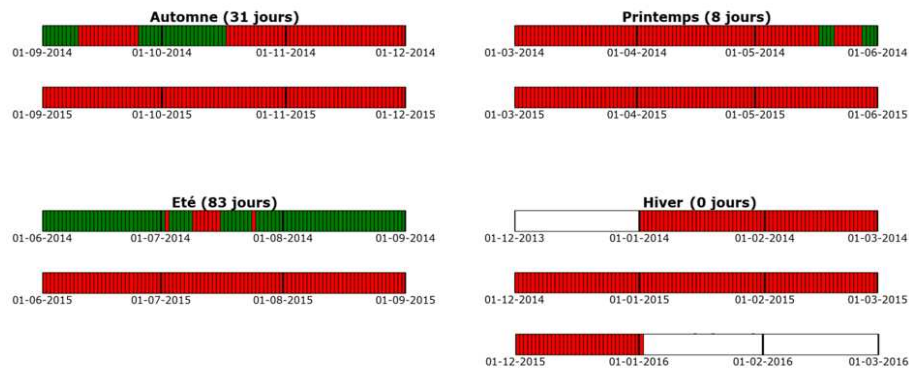


FIGURE 2.9 – Répartition saisonnière des journées du jeu de données de vérification  $JDD_{temp}$  utilisées

de données du fait d'une quantité de données moindre. Les précipitations sont peu fréquentes (< 25 % du temps) quelles que soient la saison et la direction du vent puisque les centiles 0,75 et 0,25 sont nuls pour toutes les directions de vent (cf. figure 2.10). Seules les roses des vents associées à la saison estivale peuvent donner lieu à des interprétations fiables, du fait d'une quantité de données suffisamment importante pour chaque direction de vent. Pour cette saison, toutes les conclusions qui avaient été tirées du jeu de données  $JDD_{calib}$  restent valables, à savoir : les précipitations viennent du sud-ouest, les vents dominants sont principalement des vents de nord, les températures les plus élevées ainsi que les temps clairs sont associés à des vents de nord-est.

D'une manière générale, les mêmes stations que pour la phase de calibration étant utilisées, toutes les remarques émises en section 2.3.1 à l'encontre du matériel utilisé et de son positionnement restent valables. Des solutions pour tenter de pallier ces problèmes seront proposées section 2.4.

### 2.3.2.2 Acquisition des données pour d'autres stations

Cette phase permet de vérifier comment les modèles se comportent lorsqu'ils sont appliqués à des points du territoire qui diffèrent de ceux utilisés pour la calibration mais pour des périodes de mesure similaires. Les stations utilisées pour la vérification doivent donc être différentes de celles utilisées pour la calibration mais la période de mesure doit être similaire. En réalité, ce n'est pas la même période qui est utilisée. Celle de la campagne de vérification est beaucoup plus courte (les étés 2010 et 2012 contre les années 2010-2014) mais est incluse dans la période de mesure de la campagne de calibration. Les stations utilisées sont par contre bien différentes

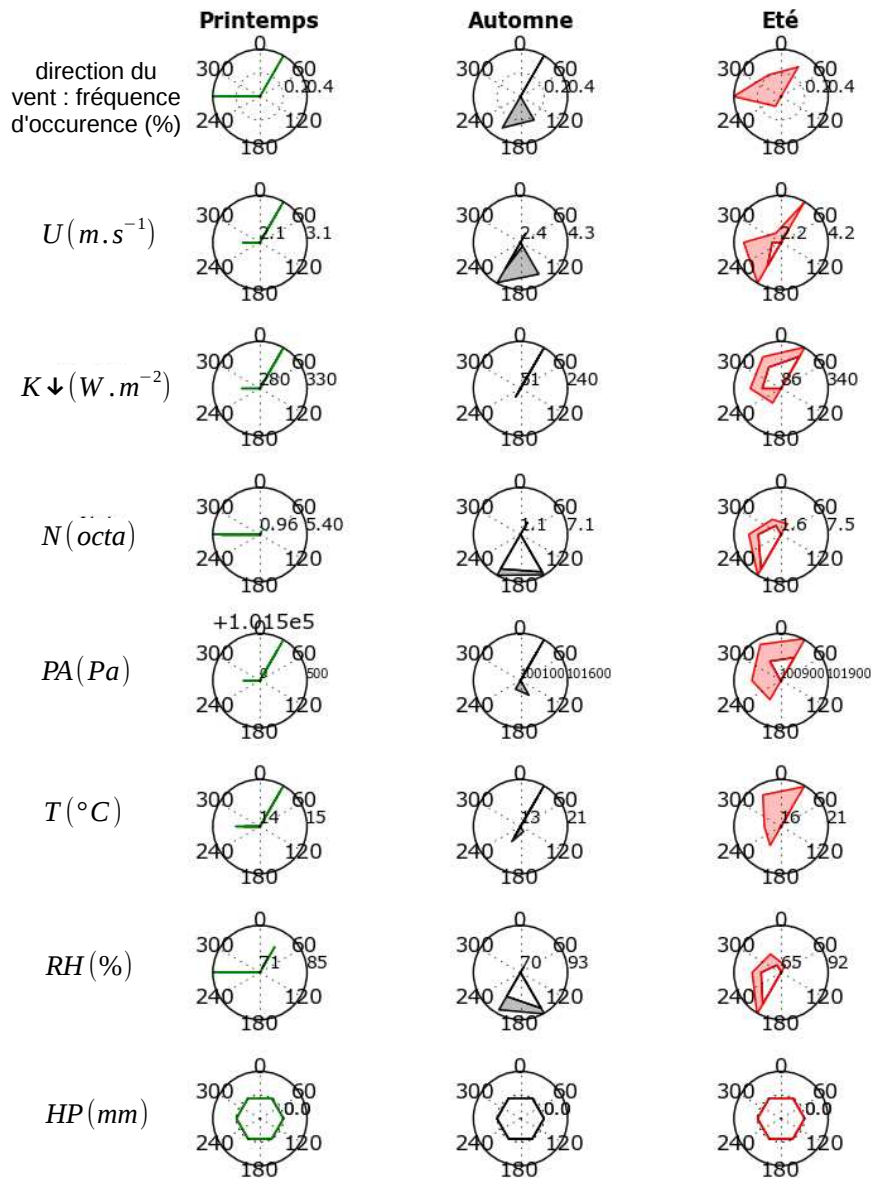


FIGURE 2.10 – Répartition du type de temps selon la saison et la direction du vent pour le jeu de données de vérification  $JDD_{temp}$ . Ligne continue extérieure : pourcentile 0,75 de la variable  $VM_m$ , ligne continue intérieure : pourcentile 0,25 de la variable  $VM_m$

(réseau FluxSAP). Les mêmes capteurs météorologiques sont également réutilisés (la station MF Nort-sur-Erdre pour la mesure de température et d'humidité de l'air et la station MF Bouguenais pour la mesure des autres variables météorologiques - cf. figure 2.5). Le jeu de données utilisé pour cette phase est noté  $JDD_{spat}$ .

Les périodes de fonctionnement des stations de ce jeu de données sont courtes. De ce fait, il n'est pas possible de faire correspondre son calendrier des jours utilisables avec celui du jeu de données  $JDD_{calib}$ . Ainsi,  $JDD_{spat}$  sera donc utilisé pour vérifier les modèles temporels

en plus des modèles spatiaux. Seuls 14 jours de début d'été et six jours de fin de printemps sont utilisables, représentés sur la figure 2.11.



FIGURE 2.11 – Répartition saisonnière des journées du jeu de données de vérification  $JDD_{spat}$  utilisées

Les directions du vent pour la période estivale sont principalement ouest et sud-ouest, associées à des précipitations fréquentes et donc à des valeurs faibles de rayonnement solaire (figure 2.12). En effet, les centiles 0,75 pour le rayonnement solaire moyen journalier étaient pour les deux précédents jeux de données de l'ordre de  $350 \text{ W/m}^2$ , contre  $270 \text{ W/m}^2$  pour ce jeu de données. Le nombre de jours utilisés pour le printemps est relativement faible. Il est donc difficile de tirer des conclusions sur une quelconque influence de la direction du vent sur les types de temps observés, mais des observations peuvent être effectuées pour les six jours de mesure : le rayonnement est plutôt fort (avec  $330 \text{ W/m}^2$  pour le centile 75% contre  $250 \text{ W/m}^2$  pour celui du jeu de données  $JDD_{calib}$ ) et les températures plutôt élevées (le centile 0,25 est de  $17^\circ\text{C}$  contre  $7,8^\circ\text{C}$  pour le jeu de calibration).

D'une manière générale, le même matériel et les mêmes méthodes de positionnement que pour la phase de calibration sont utilisés. Toutes les remarques émises à leur rencontre sont donc valables également pour ce jeu de données.

### 2.3.2.3 Acquisition des données à une autre période et pour d'autres stations

Cette phase permet de vérifier comment les modèles se comportent lorsqu'ils sont appliqués à d'autres points du territoire et pour des périodes différentes de celles utilisées pour la calibration. Le réseau URBIO (période 2015 - 2016) est donc utilisé. Les mêmes capteurs météorologiques sont également réutilisés sur la même période (la station MF Nort-sur-Erdre pour la mesure de température et d'humidité de l'air et la station MF Bouguenais pour la mesure des autres variables météorologiques - cf. figure 2.5). Le jeu de données utilisé pour cette phase est noté  $JDD_{spat\&temp}$ .

Comme pour le jeu de données précédent, seuls le printemps et l'été recensent des journées de données complètes, avec respectivement 81 et 39 jours (figure 2.13). Sous les latitudes de nos zones d'études, le climat urbain est souvent regardé à travers le prisme des canicules, qui sont des événements météorologiques estivaux. Des mesures de température pendant



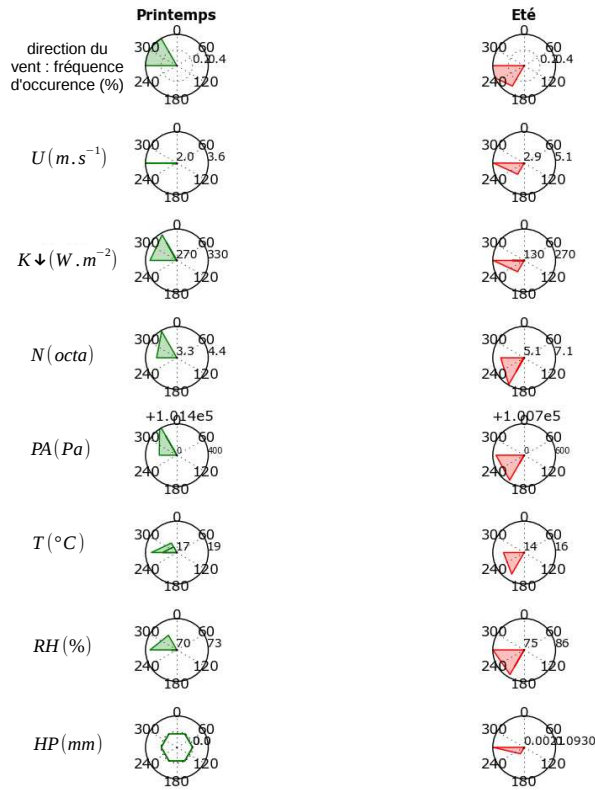


FIGURE 2.12 – Répartition du type de temps selon la saison et la direction du vent pour le jeu de données de vérification  $JDD_{spat}$ . Ligne continue extérieure : pourcentile 0,75 de la variable  $VM_m$ , ligne continue intérieure : pourcentile 0,25 de la variable  $VM_m$

les périodes qui sont a priori les plus intéressantes sont donc bien enregistrées.

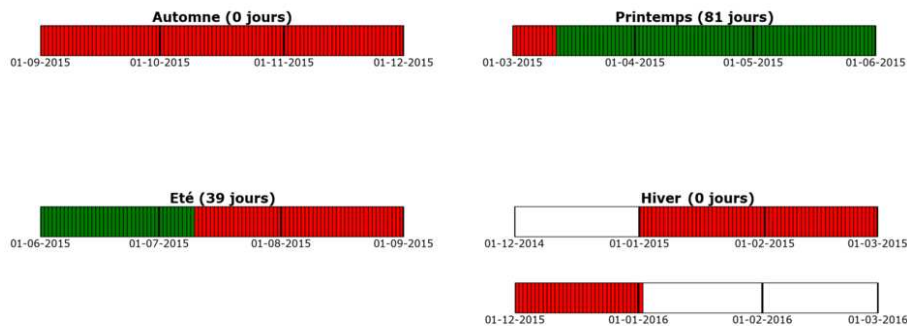


FIGURE 2.13 – Répartition saisonnière des journées du jeu de données de vérification  $JDD_{spat\&temp}$  utilisées

Les liens entre types de temps et direction de vent, établis à partir du jeu de données  $JDD_{calib}$ , se vérifient presque tous pour ce nouveau jeu de données (figure 2.14). La seule différence peut être imputée au manque de données en période estivale : durant les 39 jours

qu'elle comporte, aucun n'a une direction prédominante provenant du tiers sud de la rose des vents.

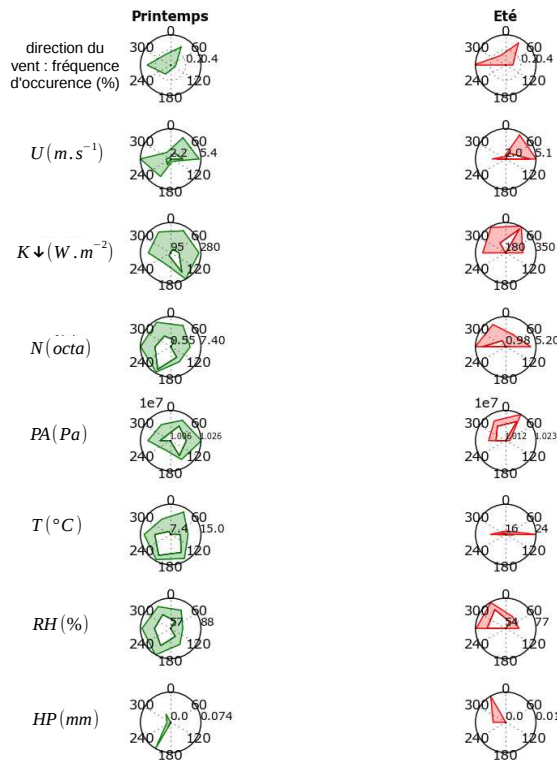


FIGURE 2.14 – Répartition du type de temps selon la saison et la direction du vent pour le jeu de données de vérification  $JDD_{spat\&temp}$ . Ligne continue extérieure : pourcentile 0,75 de la variable  $VM_m$ , ligne continue intérieure : pourcentile 0,25 de la variable  $VM_m$

Le réseau de capteurs de température utilisé pour ce jeu de données est beaucoup moins critiquable que celui utilisé pour les jeux de données précédents :

- la précision des capteurs (0,1 K) est du même ordre de grandeur que les meilleurs capteurs utilisés dans les études microclimatiques présentées tableau 2.5 ;
- les abris ventilés URBIO sont a priori aussi voire plus performants que certains abris de référence : Lacombe (2010) a montré que la température mesurée sous ces abris en période diurne était inférieure à celle mesurée sous des abris à ventilation forcée ou sous des abris dits de Stevenson pourtant très peu sensibles au rayonnement solaire.

Les emplacements utilisés pour les capteurs respectent les recommandations de Oke (2004) concernant les distances au sol et aux surfaces verticales. Notons également que les surfaces au-dessus desquelles les capteurs ont été implantés sont représentatives des types de surfaces qui se trouvent dans un rayon alentour de 100 m (les capteurs ont ainsi été placés au-dessus

d'une surface enherbée si la station était placée dans un parc majoritairement enherbé). Enfin, même si les capteurs n'ont pas été calibrés depuis leur mise en place, ils sortaient d'usine et avaient donc été étalonnés juste avant leur implantation sur site. La durée de la campagne étant relativement courte ( $< 1$  an), l'erreur due à la dérive des capteurs est supposée négligeable devant les ordres de grandeur des différences à mesurer.

D'une manière générale, les mesures de ce jeu de données sont beaucoup plus fiables que celles des jeux précédents. Il est cependant considéré trop pauvre en données (nombre de jours enregistrés et nombre de stations utilisées) pour permettre d'établir des modèles fiables, ce qui explique (entre autres raisons) son utilisation pour la vérification plutôt que pour la calibration.

### **2.3.3 Jeux de données de vérification provenant d'une autre zone d'étude**

Les jeux de données de vérification de l'extrapolation permettent de vérifier comment les modèles temporels et spatiaux se comportent lorsqu'ils sont appliqués à une autre zone d'étude.

Seuls Balázs et al. (2009) et Chen et al. (2012) ont publié des articles traitant de la vérification de l'extrapolation de leurs modèles. Les modèles ont été, pour ces deux études, établis et vérifiés à partir de campagnes de mesure mobile. En réalité, Chen et al. (2012) n'ont pas réalisé une campagne de vérification d'extrapolation telle que définie au paragraphe précédent : ils ont réalisé leurs mesures dans un autre quartier de la même ville (Hong-Kong) en se replaçant dans un contexte géographique et morphologique (formes urbaines) identique. Cependant, le nouveau quartier est situé à plus de 10 kilomètres de la zone de calibration, ce qui de notre point de vue justifie la classification de leur campagne de vérification dans la partie extrapolation. Balázs et al. (2009) ont quant à eux effectué leurs campagnes de vérification dans quatre villes situées dans la même zone climatique et dans les mêmes conditions topographiques que les deux villes qui ont accueilli les campagnes de calibration. La taille des quatre villes utilisées pour la vérification de l'extrapolation varie entre 23 000 et 191 000 habitants contre 200 000 pour les villes utilisées pour la calibration. Les agglomérations utilisées pour la vérification de nos modèles sont toutes les deux moins peuplées que celle utilisée pour la calibration mais sont situées dans une même zone climatique et dans un même contexte géographique régional.

### 2.3.3.1 Agglomération d'Angers

Le réseau URBIO (période 2015 - 2016) est utilisé. L'ensemble de capteurs météorologiques de la station MF Angers est également utilisé sur la même période (cf. figure 2.5). Le jeu de données correspondant à cette phase de vérification est noté  $JDD_{extrapol}^{(Angers)}$ .

Ce jeu de données est particulièrement complet et bien équilibré entre l'été, l'automne et le printemps (cf figure 2.15). Seul l'hiver est sous-représenté avec seulement cinq jours de données utilisables.

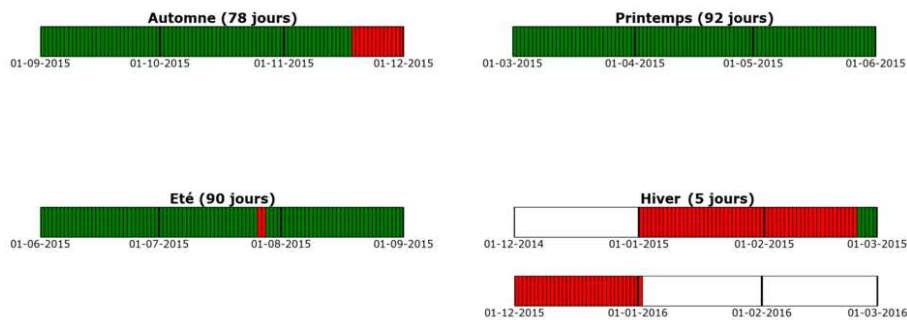


FIGURE 2.15 – Répartition saisonnière des journées du jeu de données de vérification  $JDD_{extrapol}^{(Angers)}$  utilisées

L'analyse des types de temps selon la direction du vent peut être réalisée à partir de la figure 2.16.

De même que pour Nantes :

- les directions de vent estivales et printanières proviennent rarement du sud, direction pourtant quasi-prédominante en automne ;
- les nuages (nébulosité forte) et les précipitations sont souvent apportés par le vent du sud ;
- les temps clairs et ensoleillés sont souvent associés à un vent du nord.

Une différence notable apparaît avec l'agglomération de Nantes : les températures estivales chaudes viennent de l'ouest alors qu'elles provenaient du nord-est pour Nantes. Les jeux de données sont néanmoins trop réduits pour pouvoir conclure à une réelle différence.

Le matériel utilisé pour mesurer les températures est le même que pour le jeu de données de vérification  $JDD_{spat\&temp}$  (section 2.3.2.3). Les mesures enregistrées pour ce jeu de données sont donc beaucoup plus fiables que celles enregistrées avec le jeu de données  $JDD_{calib}$ .

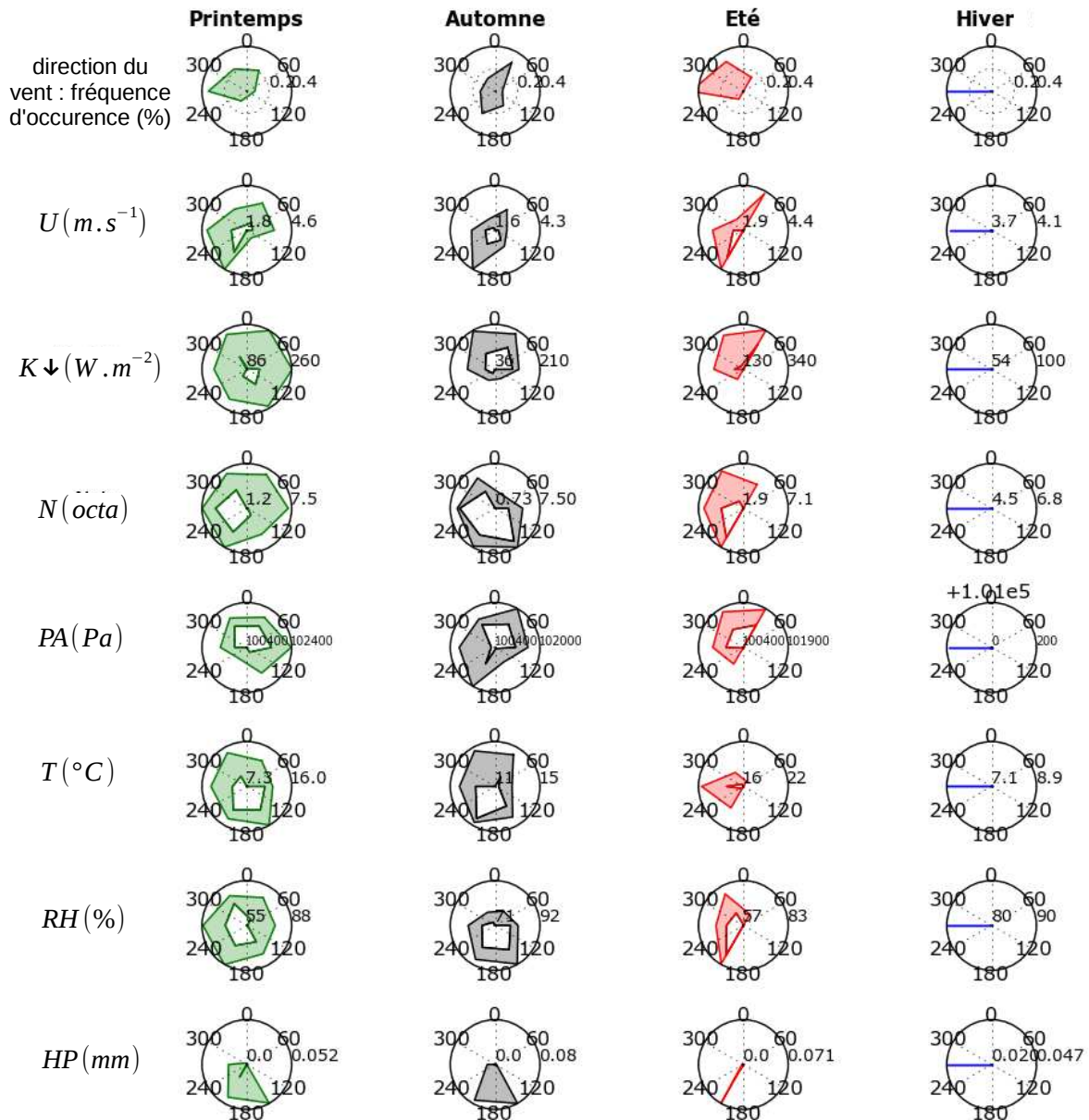


FIGURE 2.16 – Répartition du type de temps selon la saison et la direction du vent pour le jeu de données de vérification  $JDD_{extrapol}^{(Angers)}$ . Ligne continue extérieure : pourcentile 0,75 de la variable  $VM_m$ , ligne continue intérieure : pourcentile 0,25 de la variable  $VM_m$

### 2.3.3.2 Agglomération de la Roche-sur-Yon

Le réseau URPIO (période 2015 - 2016) est utilisé. L'ensemble de capteurs météorologiques de la station MF LRY est également utilisé sur la même période (cf. figure 2.5). Le jeu de données correspondant à cette phase de vérification est noté  $JDD_{extrapol}^{(LRY)}$ .

Ce jeu de données est, comme le précédent, complet et bien équilibré entre l'été, l'automne et le printemps. Seul l'hiver est sous-représenté avec seulement trois jours de données

utilisables (figure 2.17).

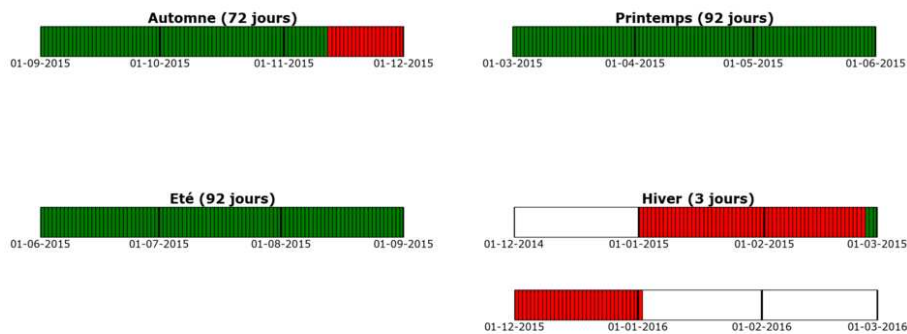


FIGURE 2.17 – Répartition saisonnière des journées du jeu de données de vérification  $JDD_{extrapol}^{(LRY)}$  utilisées

L'analyse des types de temps selon la direction du vent peut être réalisée à partir de la figure 2.18.

Les mêmes conclusions que celles énoncées pour l'agglomération nantaise section 2.3.1 peuvent être tirées. Les températures estivales chaudes proviennent donc également de l'est, contrairement à Angers où elles étaient associées à un vent d'ouest.

Le matériel utilisé pour mesurer les températures est le même que pour le jeu de données de vérification  $JDD_{spat\&temp}$  (section 2.3.2.3). Les mesures enregistrées par ce jeu de données sont donc beaucoup plus fiables que celles enregistrées avec le jeu de données  $JDD_{calib}$ .

## 2.4 Méthode de correction automatique des températures erronées

### 2.4.1 Mise en évidence de l'erreur

Une première analyse des mesures effectuées à partir de quelques capteurs du réseau TRH montrent des températures qui augmentent puis diminuent brutalement à certains instants de la journée. C'est par exemple le cas de la température mesurée par le capteur TRH11 (figure 2.19b). Celui-ci est situé à une trentaine de centimètres d'un mur exposé nord-est comme l'illustre la figure 2.19a.

Les périodes de surchauffe coïncident avec celles où l'abri ventilé voit directement le

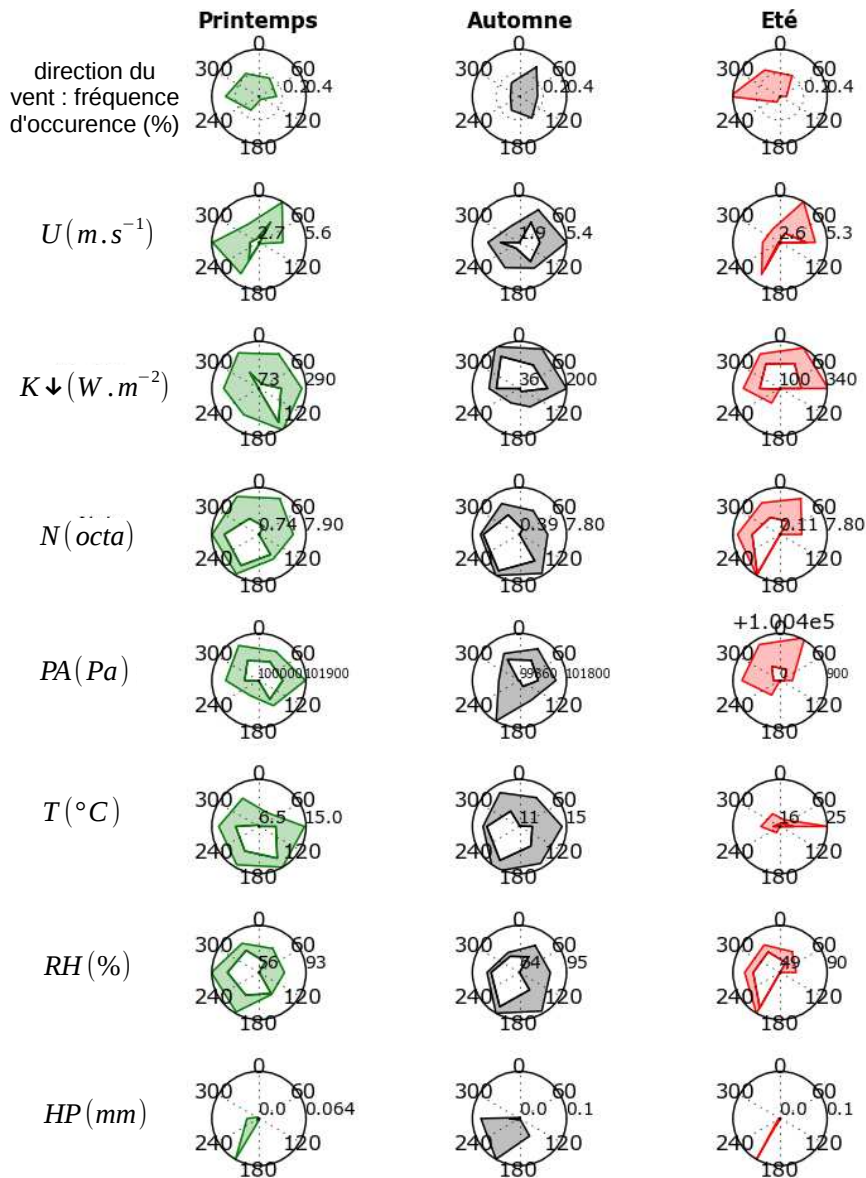


FIGURE 2.18 – Répartition du type de temps selon la saison et la direction du vent pour le jeu de données de vérification  $JDD_{extrapol}^{(LRY)}$ . Ligne continue extérieure : pourcentile 0,75 de la variable  $VM_m$ , ligne continue intérieure : pourcentile 0,25 de la variable  $VM_m$

soleil. On considère que la température mesurée pendant cette période est erronée par rapport à l'objectif de l'étude : elle n'est pas représentative de la température ambiante du quartier et elle doit donc être corrigée. L'ordre de grandeur de cette erreur est évalué à partir de la figure 2.19b à 5°C.

Plusieurs hypothèses sont avancées pour expliquer cette erreur : elle peut être due au mauvais positionnement du capteur (hypothèse 1), au matériel (hypothèse 2) ou à une combinaison des deux hypothèse 3).

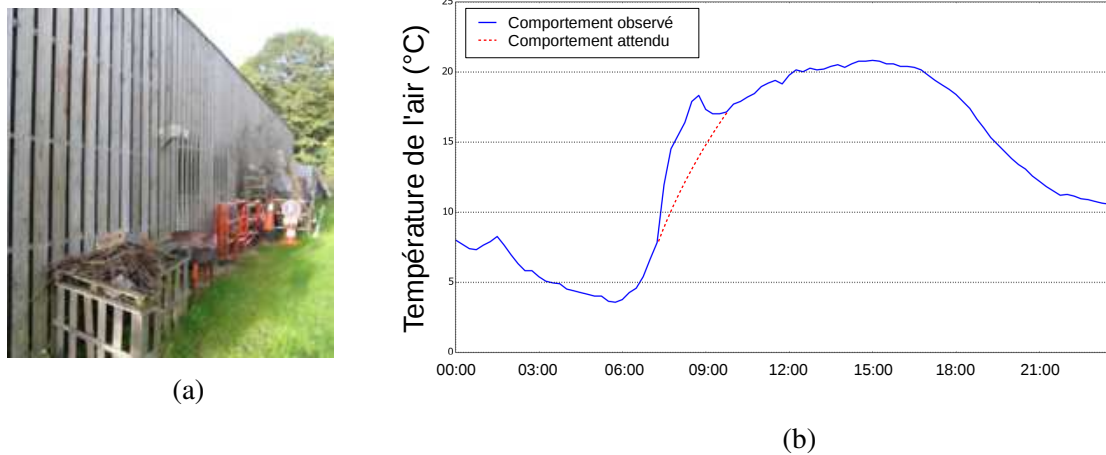


FIGURE 2.19 – Station de mesure où une erreur de mesure de température est observée : (a) photographie de la station sur site, (b) illustration de l'évolution temporelle de la température observée un jour de ciel clair

- Hypothèse 1 : l'erreur est due au positionnement du capteur. La température près d'une paroi augmente très rapidement lorsque le rayonnement l'atteint (pour rappel plusieurs capteurs du réseau TRH sont positionnés à moins de 50 centimètres d'un mur). La mesure ne serait donc pas biaisée mais le capteur ne serait pas positionné en un point représentatif du climat du quartier alentour.
- Hypothèse 2 : l'erreur est d'origine matérielle. L'abri ventilé remplit mal son rôle de protection du capteur de température contre le soleil car il est de mauvaise qualité. Le système {abri + capteur} est donc plus chaud que l'air environnant : c'est donc une erreur de mesure que l'on qualifiera par la suite de biais de mesure.
- Hypothèse 3 : l'erreur est due à une mauvaise combinaison {matériel / positionnement du capteur}. L'abri ventilé remplit mal son rôle de protection du fait de son positionnement. La proximité du mur empêcherait ici la bonne ventilation du système abri - capteur, qui se réchauffe donc sous l'effet du rayonnement solaire. Ceci implique finalement une mesure biaisée également.

Une procédure de correction est proposée dans ce chapitre. Elle est schématisée sur la figure 2.20. Une phase de diagnostic (section 2.4.2) permettra de vérifier la cohérence des hypothèses énoncées ci-dessus : une analyse de la littérature ainsi que des campagnes de mesure seront réalisées. Une deuxième partie (section 2.4.3) permettra de modéliser cette erreur afin de corriger les données. La vérification de cette modélisation sera opérée en troisième partie (section 2.4.4) et une dernière partie sera consacrée à l'application du modèle de corrections



aux jeux de données entachés d'erreur (section [2.4.5](#)).

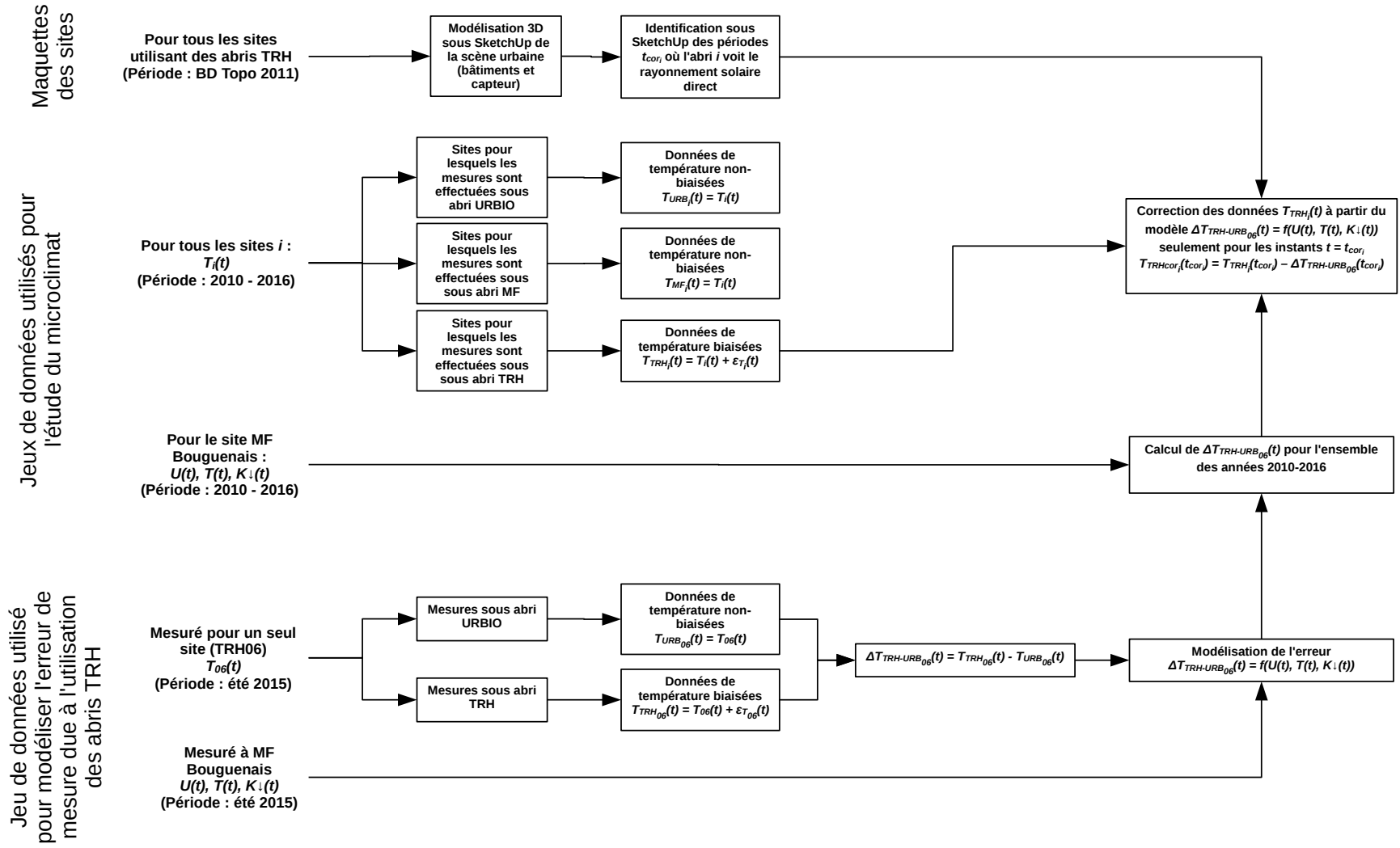


FIGURE 2.20 – Méthodologie adoptée pour corriger un jeu de données dont certaines stations engendrent des erreurs de mesure de la température d’air

## 2.4.2 Hypothèses explicatives de l'erreur

### 2.4.2.1 Erreur de positionnement

Oke (2004) recommande de positionner les capteurs de température à plus d'un mètre de toute surface verticale. En-dessous de cette distance, un gradient important de température peut s'établir sous certaines conditions d'ensoleillement. Niachou et al. (2008) ont mesuré pour une journée d'été les conditions de vent et les températures de l'air au sein d'une rue canyon (de rapport d'aspect  $H/W = 1,7$ ) de la ville d'Athènes (Grèce). Des thermomètres étaient positionnés à une dizaine de mètres du sol et à 0,5 m des parois de part et d'autres de la rue. Lorsque le soleil frappait la façade sud, les différences de température entre l'air proche de la façade sud et celui proche de la façade nord était de l'ordre de 4 à 5°C, ordre de grandeur cohérent avec les observations de Nakamura et Oke (1988) et Santamouris et al. (1999) ont également réalisé des mesures de température au sein d'une rue canyon à Kyoto (Japon). La position des instruments de mesure est précisée figure 2.21. Les isothermes au sein de la rue canyon sont également représentées sur la même figure à 11h et à 14h, périodes de la journée où le gradient horizontal est le plus fort (11 et 14h). À noter qu'entre 11h et 11h10, le vent est transverse au canyon tandis qu'il est parallèle entre 14h et 14h10.

À une hauteur de deux mètres, les différences de température entre l'air situé à 0.5 m de la façade ensoleillée et celui situé au milieu de la rue sont de l'ordre de 1°C. Ces observations sont réalisées lorsque la vitesse du vent dans le canyon est comprise entre 0,5 et 1,6 m/s et la hauteur du soleil dans le ciel entre 72 et 81°. Le gradient reste également légèrement positif la nuit, avec des températures proches des parois de l'ordre de 0,5°C supérieures aux températures au milieu de la rue.

Toutes ces études montrent que des différences de température apparaissent spatialement au sein d'une rue canyon, notamment lorsque celle-ci est en grande partie ensoleillée. Le gradient de température est souvent marqué lorsque l'on s'approche des façades au soleil, et varie selon la vitesse du vent, sa direction et la stabilité de l'air dans le canyon.

Dans notre cas, peu de capteurs sont situés dans des espaces assimilables à des rues canyons. Les périodes de mesure qui apparaissent clairement comme erronées correspondent à des hauteurs solaires faibles ( $< 60^\circ$ ). Ces caractéristiques diffèrent de celles décrites auparavant, où les rues sont des canyons, conditionnant des hauteurs solaires importantes lorsque le soleil frappe les façades. De plus, les mesures qui ont été réalisées ne permettent pas d'observer aisément le gradient de température dans les 50 premiers centimètres d'air proches de la paroi

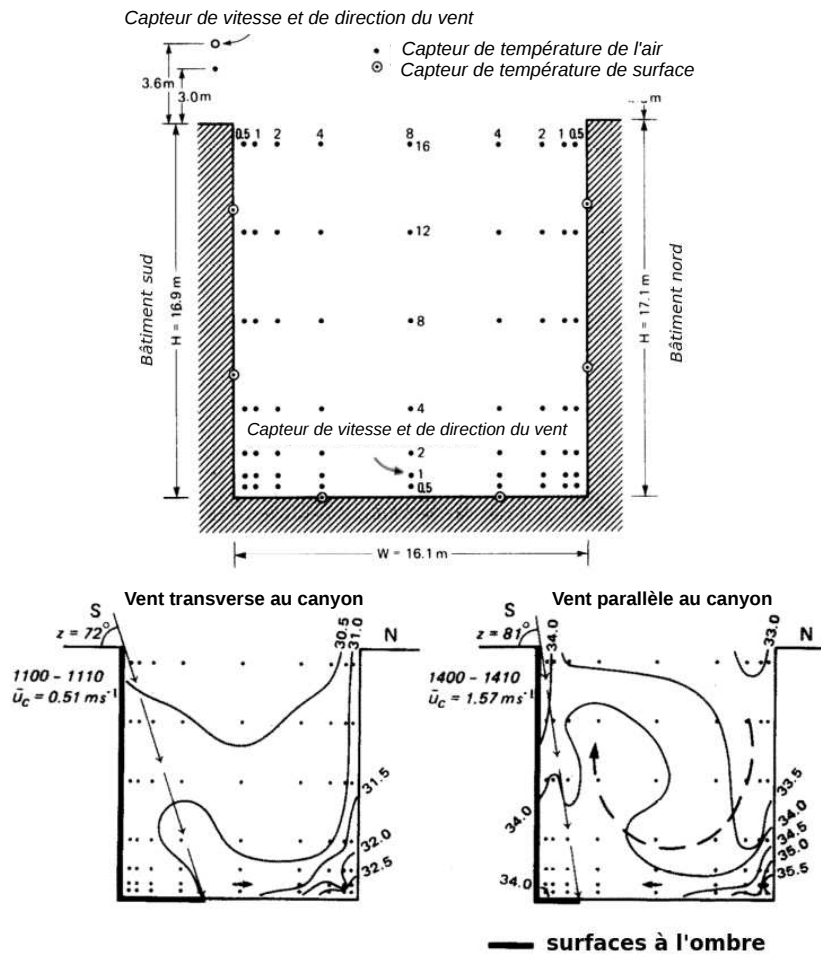


FIGURE 2.21 – Isothermes mesurés au sein d'une rue canyon orientée est-ouest. Source : adapté de Nakamura et Oke (1988)

(couche d'air dans laquelle se trouve le capteur pour plusieurs de nos sites). Les données d'une campagne de mesure (numérotée 1 par la suite) réalisée par le LHEEA<sup>2</sup> sont donc utilisées pour quantifier ce gradient dans des conditions urbaines plus ouvertes. Un peigne de thermocouples a été positionné à l'extérieur d'un bâtiment, orthogonalement à sa paroi orientée au sud. Les points de mesure de température sont situés à 0,005, 0,01, 0,02, 0,05 et 1 m du mur et à 1,5 m au-dessus du sol (figure 2.22).

La température d'air à un mètre du mur est prise comme température de référence et est supposée représentative de la température de l'air de l'espace ouvert considéré. Seule la plus ensoleillée des 10 journées estivales de la campagne 1 est conservée pour l'analyse. L'évolution de la température de l'air est tracée en fonction de la distance au mur en soustrayant à chaque point de mesure la température de référence (cf. figure 2.23). Elle est réalisée toutes les deux heures afin d'illustrer les variations temporelles du gradient. Les courbes de tendance

<sup>2</sup> Laboratoire de recherche en Hydrodynamique, Énergétique et Environnement Atmosphérique (école centrale de Nantes)

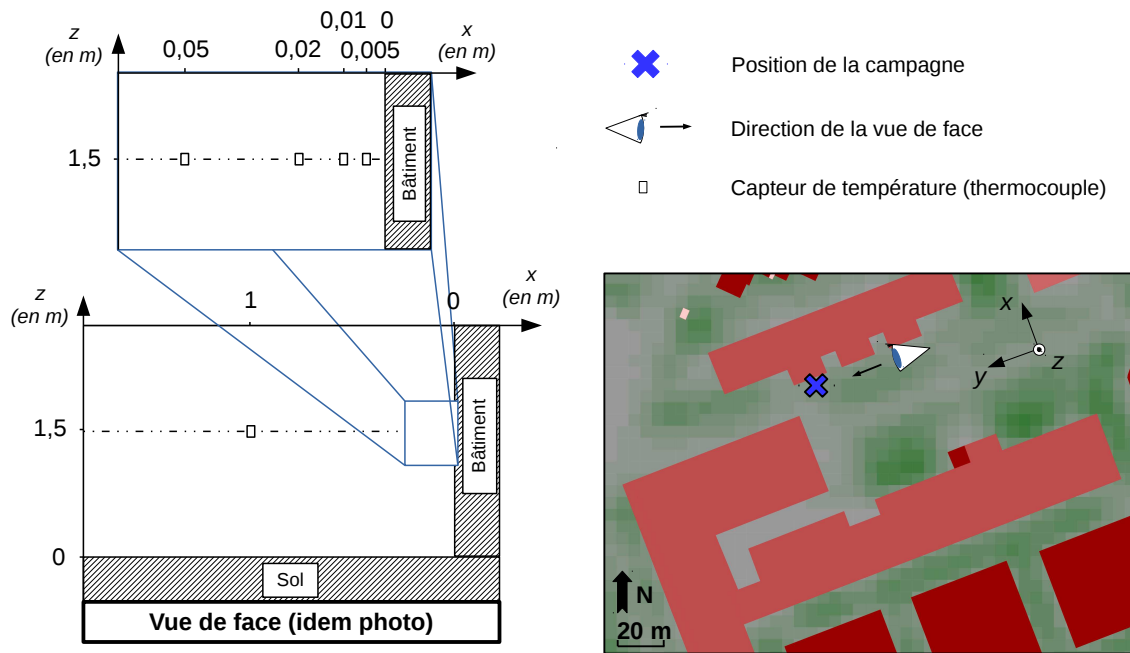


FIGURE 2.22 – Description du site de mesure et du positionnement du matériel utilisé pendant la campagne de mesure 1. Cette campagne est utilisée pour quantifier l'évolution de la température de l'air au voisinage d'un mur exposé au rayonnement solaire

sont tracées de façon à passer par le point de coordonnées  $(x = 100 \text{ cm}, y = 0^\circ \text{C})$ . Quelle que soit l'heure de la journée, il apparaît que la différence de température au-delà de 20 centimètres du mur est toujours inférieure à  $0,5^\circ\text{C}$ . Cette valeur ne reste cependant qu'un ordre de grandeur propre aux conditions spatiales et temporelles de la campagne de mesure.

- Le site a des formes urbaines spécifiques, qui modifient de manière importante les conditions de vent. Les matériaux qui constituent la paroi ont une influence sur les échanges thermiques à sa surface et en son sein : ils peuvent modifier indirectement le gradient de température extérieure.
- Le mur voit le soleil la quasi totalité de la journée. L'utilisation d'une seule journée estivale est donc représentative de tous les angles zénithaux que peut prendre le soleil par rapport au mur au cours d'une année. La vitesse du vent (de 0 à 2 m/s - mesurée à 3 m du sol et 2 mètres du mur) est quant à elle propre à la journée choisie, de même que le gradient.

D'après l'ensemble de ces informations, l'erreur due au positionnement ne devrait guère excéder les  $1^\circ\text{C}$ , largement en-deçà des  $5^\circ\text{C}$ , erreur estimée pour la station TRH11 (figure 2.19b). Une autre cause est donc recherchée.

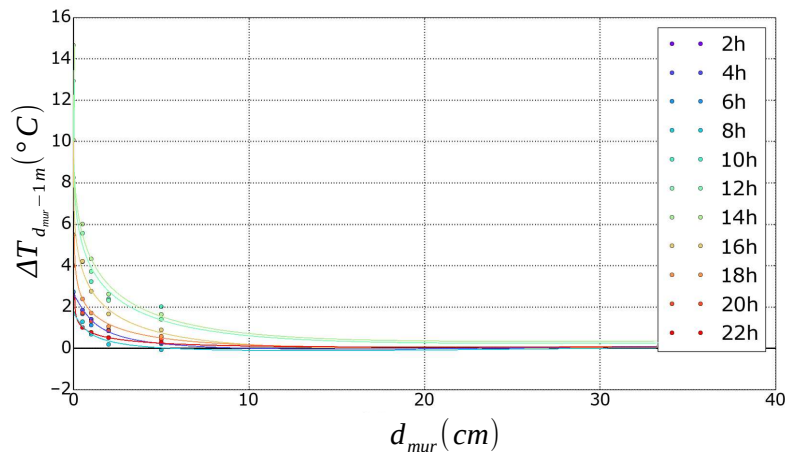


FIGURE 2.23 – Différence de température entre l’air au voisinage d’un mur soumis au rayonnement solaire et l’air mesuré à 1 m du mur

### 2.4.2.2 Erreur matériel

De nombreuses études ont montré que deux abris ventilés soumis au rayonnement solaire fournissaient rarement les mêmes performances (Cheng et al. 2014 ; Erell et al. 2005 ; Lacombe 2010 ; Nakamura et Mahrt 2005). Afin d’identifier si les abris ventilés utilisés dans cette thèse sont sensibles au rayonnement solaire, une campagne d’inter-comparaison (numérotée 2 par la suite) est menée. Trois sondes PT100 neuves (étalonnées récemment) de précision 0,1 K sont placées chacune dans les abris ventilés suivants : l’abri utilisé dans le réseau TRH (appelé par la suite abri TRH), l’abri utilisé dans le réseau URBIO (appelé par la suite abri URBIO) et un abri ventilé mécaniquement (photographie figure 1.8c). Ce dernier est constitué d’un cylindre principal orienté verticalement et ouvert aux deux extrémités. Ce cylindre abrite le capteur de température et le ventilateur, qui est situé en partie haute et qui aspire l’air de bas en haut. Ce système est protégé du rayonnement direct descendant par une casquette en forme de disque et par un second cylindre pour le rayonnement solaire latéral. Cet abri, destiné initialement à la mesure de l’hygrométrie, est un prototype. Ses performances ne sont donc pas connues et il sera dans un premier temps considéré comme parfait. Un dernier abri est utilisé dans les jeux de données de cette thèse. Il s’agit d’un abri de Stevenson (Socrima *modèle réduit*<sup>3</sup>), utilisé par tous les sites Météo-France (tableau 2.3). D’après l’étude réalisée par Lacombe et al. (2011) dont les résultats sont synthétisés par Lacombe (2010), ces abris ont de meilleures performances que de nombreux abris ventilés mécaniquement. Ils ont également des performances très proches de l’abri URBIO utilisé dans cette intercomparaison. Nous considérerons donc que les résultats

<sup>3</sup> [http : //education.meteofrance.fr/ressources – pour – les – enseignants/observer – et – mesurer/l – abri – meteo](http://education.meteofrance.fr/ressources-pour-les-enseignants/observer-et-mesurer/l-abri-meteo) [Accédé le 14/01/2016]

de cette campagne de mesure pour l'abri URBIO peuvent être extrapolés également à l'abri Socrima.

La campagne 2 s'est déroulée du 9 au 30 juillet 2016. Les trois abris sont disposés à une même distance de 1,5 m d'un mur. La vitesse et la direction du vent ainsi que les rayonnements en courtes et grandes longueur d'ondes montants et descendants sont mesurés sur le site respectivement à l'aide d'un anémomètre sonique (WindMaster 3D de la marque Gill<sup>4</sup>) et d'un bilanmètre (modèle CNR4 de Kipp and Zonen<sup>5</sup>). L'emplacement exact de chacun des capteurs est donné figure 2.24. Les campagnes de mesure se déroulent près de l'emplacement du site 06 du réseau TRH (se reporter à la figure 2.2a pour son positionnement au sein de l'agglomération).

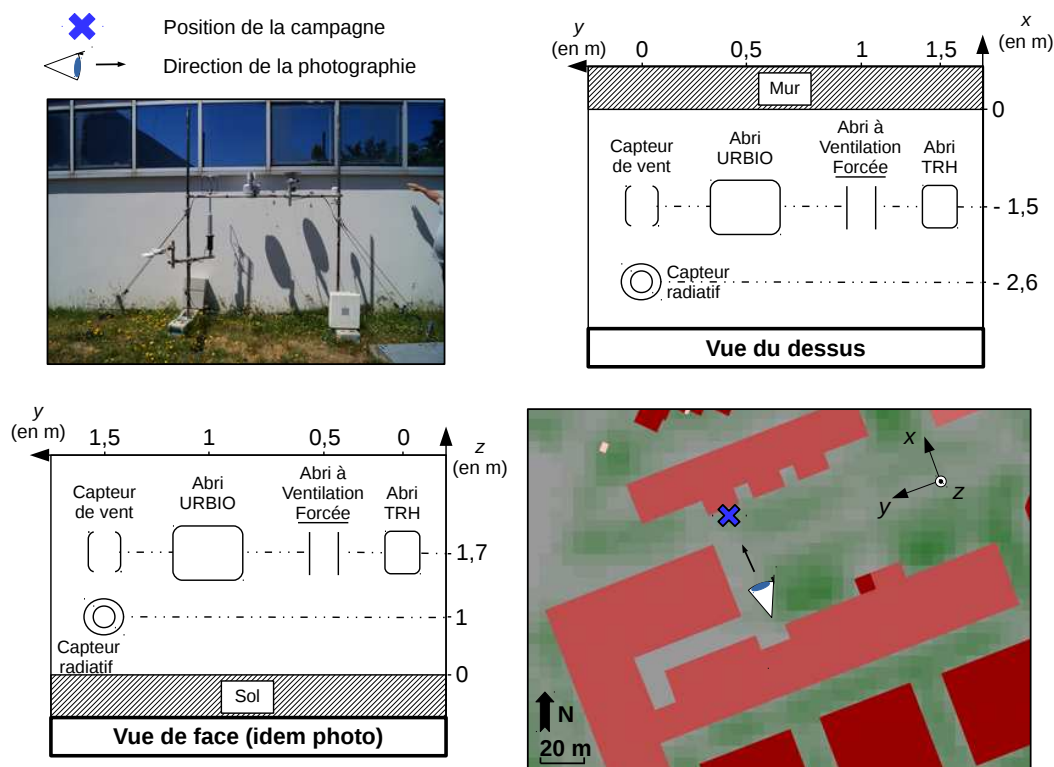


FIGURE 2.24 – Description du site de mesure et du positionnement du matériel utilisé pendant la campagne de mesure 2. Cette campagne est utilisée pour comprendre et modéliser les erreurs de mesure de température causées par l'utilisation de certains abris ventilés naturellement.

L'évolution temporelle de chacune des variables météorologiques au cours de la campagne de mesure est tracée sur la figure 2.25.

<sup>4</sup> disponible sur la page [http : //gillinstruments.com/products/anemometer/windmaster.htm](http://gillinstruments.com/products/anemometer/windmaster.htm) [Accédé le 08/07/2016]

<sup>5</sup> disponible sur la page [http : //www.kippzonen.fr/Product/136/CNR4](http://www.kippzonen.fr/Product/136/CNR4) – Radiometre – Net#.V396F15b80o [Accédé le 08/07/2016]

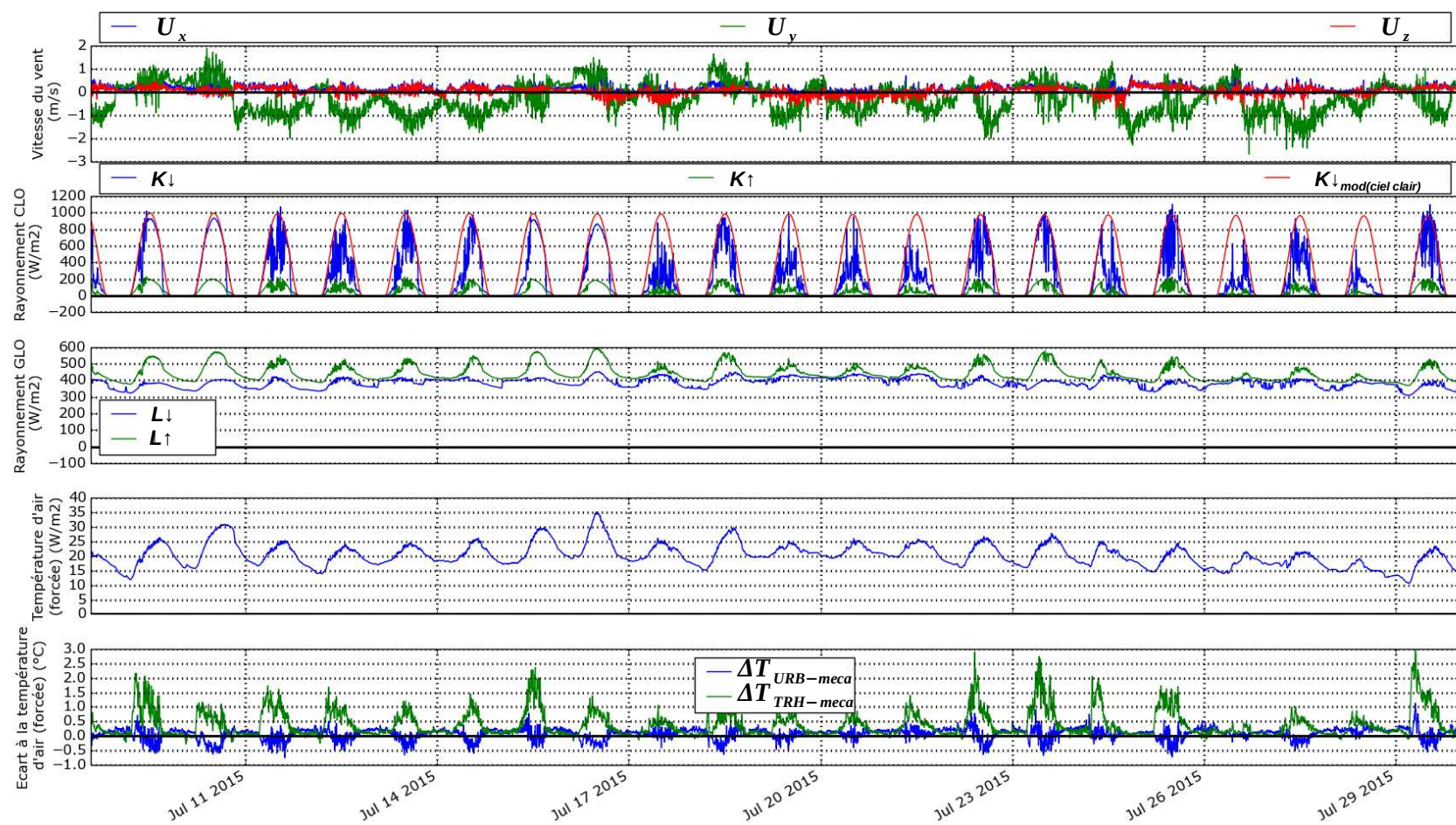


FIGURE 2.25 – Évolution temporelle de chacune des variables météorologiques mesurées localement lors de la campagne de mesure 2.



La température de l'air est celle mesurée par l'abri à ventilation forcée (considéré dans un premiers temps comme l'abri de référence). Les écarts de température sont ceux observés pour chacun des abris à ventilation naturelle par rapport à l'abri à ventilation forcée. Les valeurs du rayonnement solaire *modèle ciel clair* sont calculées à partir de l'outil de modélisation Solene. Par comparaison aux mesures, ces valeurs permettent d'identifier visuellement les journées les plus ensoleillées. Le modèle de ciel utilisé par Solene est le modèle de Perez (Perez et al. 1993), repris par Miguet (2000). La décomposition du vent selon ses trois composantes permet clairement d'identifier la direction principale comme parallèle au mur. Il apparaît également que le vent faiblit la nuit. L'écart de température enregistré entre les abris diffère selon les abris et la période de la journée considérés.

- La nuit, les deux abris à ventilation naturelle mesurent une température légèrement supérieure à celle mesurée avec l'abri à ventilation forcée.
- Le jour, l'écart mesuré à l'aide de l'abri URBIO est souvent négatif. Celui mesuré par l'abri TRH augmente assez brutalement dès que le jour se lève. Il se situe autour de 1°C en moyenne et atteint jusqu'à 3°C certaines journées.

L'augmentation brutale de l'erreur de mesure de l'abri TRH lors de l'apparition du soleil ressemble fortement à celle observable pour la station TRH11 (figure 2.19b). Cependant, son amplitude reste largement inférieure (3°C au maximum contre plus de 5°C pour la station TRH11). Même en retirant à l'erreur de la station TRH11 les 0,5°C attribuables au gradient de température qui s'établit près du mur (l'abri TRH11 est situé à environ 30 cm du mur), l'écart d'amplitude est toujours de 1,5°C. La troisième hypothèse explicative de l'erreur énoncée en section 2.4.1 permet peut-être de combler les quelques degrés celsius restants. Selon cette hypothèse, l'erreur observée peut varier selon la distance de l'abri à des obstacles au vent. En effet, pour un même rayonnement, l'équation 1.8 montre que plus le coefficient de convection (dépendant de la vitesse du vent) régissant les échanges de chaleur entre la surface du capteur et l'air est faible, plus l'erreur est élevée. La distance séparant les abris du mur est fixée à 1,5 m durant toute la campagne 2. Un deuxième jeu de capteurs (de température, de vent et de rayonnement) aurait pu être placé plus près du mur durant la même période de mesure. Cela aurait pu permettre de vérifier que l'erreur près du mur était plus élevée et de relier ce delta d'erreur à la diminution de la vitesse du vent.

Dans un premier temps, nous considérerons que le positionnement de l'abri à une distance plus ou moins proche du mur n'a pas d'influence sur ses performances. L'erreur de

5°C observée figure 2.19b serait ainsi seulement due à des valeurs de vent très faibles et donc à une mauvaise ventilation de chacun des abris ce jour-là. Nous verrons les limites de cette hypothèse en section 2.4.5.5.

### 2.4.3 Formulation d'un modèle de correction de l'erreur matérielle

La section précédente a permis de postuler que seul le matériel était responsable de l'erreur de mesure, ce qui implique un réchauffement anormal du capteur au-dessus de la température de l'air ambiant. Cette section sera consacrée à la modélisation de cette erreur en fonction des conditions météorologiques locales (mesurées sur le site) mais également régionales (mesurées par une station Météo-France). Ces modèles seront établis à partir de la campagne de mesure 2 présentée figure 2.24. Ils seront ensuite vérifiés à partir d'une autre campagne de mesure (campagne 3) qui sera présentée dans la section 2.4.4. Enfin, les modèles d'erreur seront utilisés pour corriger les jeux de données qui utilisent des abris dont les mesures sont biaisées.

#### 2.4.3.1 Hypothèses explicatives de l'erreur

Pour rappel, la différence de température existant entre le capteur et l'air (correspondant à l'erreur de mesure) peut être modélisée en première approximation par l'équation 1.8. L'analyse de cette équation permet de proposer deux causes explicatives d'une erreur de mesure :

- l'abri est un très mauvais protecteur solaire : le rayonnement net que reçoit le capteur est dans ce cas largement supérieur à zéro, ce qui conduit à son réchauffement ;
- l'abri est très mal ventilé : le système {abri + capteur} a donc tendance à se réchauffer même sous un faible rayonnement.

Une troisième source d'erreur peut également être entrevue : l'abri se réchauffe sous l'effet du rayonnement, qui va à son tour réchauffer l'air entrant dans l'abri ventilé. L'air qui entre dans l'abri, dont la température est mesurée par le capteur, est donc plus chaud que l'air ambiant.

#### 2.4.3.2 Mise en équation de l'erreur : modélisation

Le but de cette section est de modéliser par une méthode de régression l'erreur de mesure en fonction des conditions météorologiques mesurées soit localement, soit régionalement. Afin

d'identifier quel type de fonctions usuelles peut être approprié pour la modélisation, une approche plus formelle est utilisée. L'ensemble {capteur + abri} est assimilé au système schématisé sur la figure 2.26 afin de limiter la complexité des échanges de chaleur.

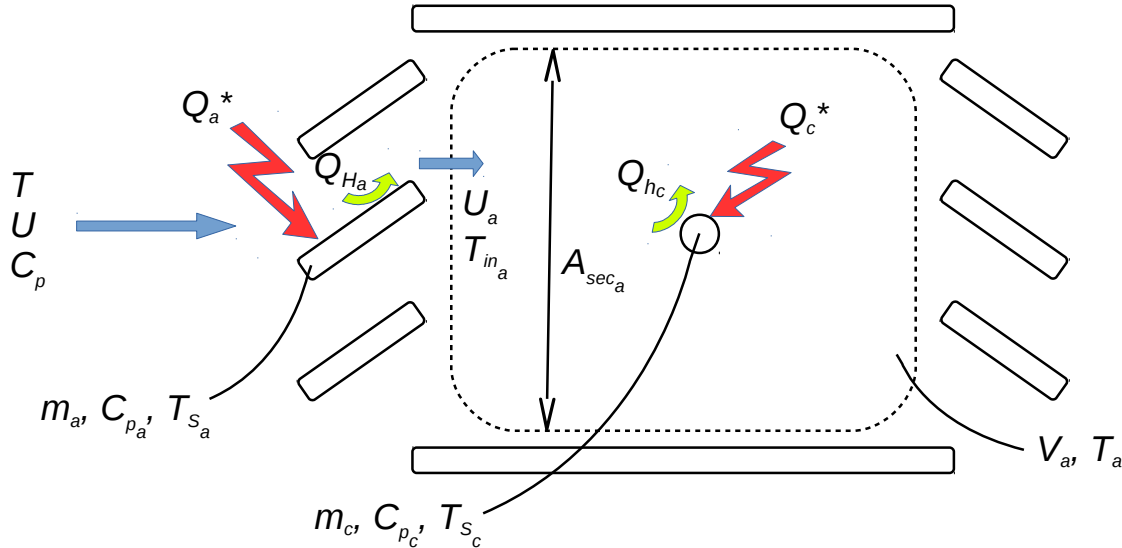


FIGURE 2.26 – Représentation schématique des échanges de chaleur observables entre un abri ventilé renfermant un capteur de température d'air et leur environnement

Pour simplifier le problème, on fait l'hypothèse que le mélange d'air dans le volume de l'abri  $V_a$  est homogène et à la même température que le capteur à tout instant  $t$ . Cela implique que la constante de temps du capteur est négligeable devant le pas de temps qui nous intéresse (ici cinq minutes), soit que les échanges de chaleur  $Q_{Hc}$  entre l'air de l'abri et le capteur se font instantanément. On note  $T_{a+c}$  la température du système air de l'abri - capteur,  $T_{S_c}$  la température du capteur (supposé homogène au sein du solide),  $T_a$  celle de l'air contenu dans l'abri.

$$\text{Ainsi : } T_{a+c} = T_{S_c} = T_a$$

On considère que la capacité thermique volumique de l'air  $C_{p_a}$  est la même pour l'air extérieur comme intérieur à l'abri.

Un bilan d'énergie est réalisé sur le système {air de l'abri + capteur} :

$$(m_c C_{p_c} + V_a C_{p_a}) \dot{T}_{a+c} = A_{sec_a} U_a C_p (T_{IN_a} - T_{a+c}) + A_c Q_c^* \quad (2.3)$$

avec  $m_c$  et  $C_{p_c}$  respectivement la masse et la capacité thermique massique du capteur,  $V_a$  le volume de l'air contenu au sein de l'abri,  $A_{sec_a}$  la section de l'abri qui permet le renouvellement de l'air,  $U_a$  la vitesse de l'air qui entre dans l'abri,  $T_{IN_a}$  la température de l'air qui entre dans

l'abri après avoir échangé de la chaleur par convection avec ses ailettes,  $A_c$  la surface du capteur et  $Q_c^*$  l'énergie radiative surfacique nette absorbée directement par la surface du capteur et par conséquent le système {air de l'abri + capteur}.

Un bilan d'énergie sur l'air entrant dans l'abri est également réalisé :

$$A_{sec_a} U_a C_p T_{IN_a} = A_{sec_a} U_a C_p T + h_a A_a (T_{S_a} - T)$$

avec  $T$  la température de l'air extérieur (que l'on souhaite mesurer avec le capteur),  $h_a$  et  $A_a$  respectivement le coefficient de convection (en  $W.K^{-1}$ ) et la surface des ailettes de l'abri (en  $m^2$ ),  $T_{S_a}$  la température de surface de l'abri (considéré homogène pour toutes les ailettes et à l'intérieur de chaque ailette).

On en déduit :

$$T_{IN_a} = T + \frac{h_a A_a (T_{S_a} - T)}{A_{sec_a} U_a C_p} \quad (2.4)$$

Le bilan de chaleur de la structure de l'abri s'écrit :

$$m_a C_{p_a} \dot{T}_{S_a} = A_a Q_a^* - h_a A_a (T_{S_a} - T) \quad (2.5)$$

avec  $m_a$  et  $C_{p_a}$  respectivement la masse et la capacité thermique massique de la structure de l'abri.

En considérant le régime comme établi (les variables météorologiques sont supposées constantes dans le temps), les équations 2.3 et 2.5 peuvent s'écrire respectivement :

$$T_{a+c} = T_{IN_a} + \frac{A_c Q_c^*}{A_{sec_a} U_a C_p} \quad (2.6)$$

et

$$h_a A_a (T_{S_a} - T) = A_a Q_a^* \quad (2.7)$$

En combinant les équations 2.4 et 2.7 avec l'équation 2.6, la différence de température entre l'air intérieur de l'abri et l'air extérieur (appelée erreur en conditions stationnaires et notée  $E_T$ ) est formulée :

$$E_T = T_{a+c} - T = \frac{A_a Q_a^* + A_c Q_c^*}{A_{sec_a} U_a C_p} \quad (2.8)$$

L'observation des données permet de mettre en évidence que l'erreur nocturne dépend non-seulement de la vitesse du vent mais également du taux de réchauffement de l'air (défini comme la dérivée première de la température mesurée sous l'abri de référence). En effet, la

figure 2.27 illustre une variation linéaire de l'erreur nocturne (pour l'abri UR-BIO) avec le taux de réchauffement de l'air : l'erreur croît d'autant plus que le taux de réchauffement est faible (taux de refroidissement élevé).

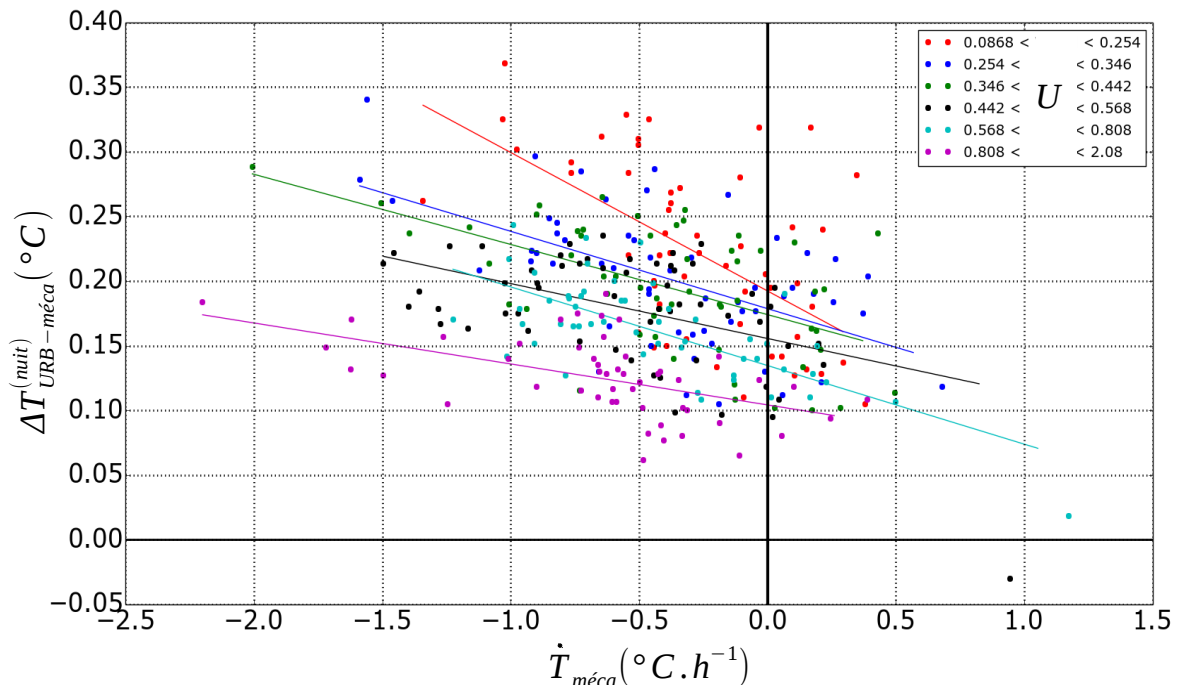


FIGURE 2.27 – Évolution de la différence de température nocturne mesurée entre l'abri UR-BIO et l'abri ventilé mécaniquement en fonction du taux de réchauffement de l'air et après classification par intervalles de vitesse du vent

La pente de la droite est également d'autant plus inclinée que le vent est faible. Le déphasage que subissent certains abris volumineux est cohérent avec ce constat. Si l'on fait l'hypothèse que le renouvellement de l'air dans l'abri est proportionnel à la vitesse du vent, ceci implique que lorsque la vitesse du vent est faible, le renouvellement est limité. Si des variations importantes de la température de l'air apparaissent par vitesse de vent faible, un déphasage de température pourra donc apparaître pour les abris dont l'air est faiblement renouvelés au regard de leur volume. Ce déphasage entraînera une erreur négative si la température de l'air augmente rapidement et positive si elle diminue rapidement. Cette théorie est cohérente avec le diagnostic formulé par Lacombe (2010) qui montre qu'un déphasage temporel existe, dû à l'utilisation de larges abris. Dans son cas, le déphasage est le temps qui sépare l'observation du maximum ou du minimum de température journalier entre l'abri à caractériser et l'abri de référence (ventilé mécaniquement). Selon cette définition, Lacombe (2010) met en évidence que tous les abris ventilés naturellement ont un déphasage médian inférieur à cinq minutes, avec les valeurs les plus fortes pour les abris les plus volumineux. Ce résultat est également

confirmé par Van der Meulen et Brandsma (2008) qui observent (en utilisant la même méthode de calcul) un déphasage négligeable pour la majorité des abris sauf pour les abris les plus larges (majoritairement des abris équivalents à des abris de Stevenson) pour lesquels le déphasage médian est de l'ordre de sept à huit minutes.

Afin de prendre en compte ce phénomène qui n'apparaît pas dans l'équation 2.7, il est nécessaire de considérer la température de l'air comme une variable dans le temps (et non plus comme une valeur constante comme c'était le cas auparavant). Dans l'équation 2.3, le terme  $(m_c C_{pc} + V_a C_{pa}) \dot{T}_{a+c}$  n'est dans ce cas plus considéré comme nul, mais les autres variables météorologiques (vitesse du vent, rayonnement, etc.) sont supposées constantes pour simplifier le problème. L'équation 2.3, combinée avec les équations 2.7 et 2.4 puis mise sous la forme d'une équation différentielle, devient :

$$\dot{T}_{a+c}(t) + \frac{A_{seca} U_a}{V_a} T_{a+c}(t) = \frac{A_a Q_a^* + A_c Q_c^*}{V_a C_p} + \frac{A_{seca} U_a}{V_a} T(t) \quad (2.9)$$

Soit  $\epsilon = \frac{A_{seca} U_a}{V_a}$ ,  $\zeta = \frac{A_a Q_a^* + A_c Q_c^*}{V_a C_p}$ . En posant également  $e_T(t) = T_{a+c}(t) - T(t)$ , l'équation différentielle 2.9 devient :

$$\dot{e}_T(t) + \epsilon e_T(t) = \zeta - \dot{T}(t) \quad (2.10)$$

On considère qu'entre deux pas de temps  $\Delta t$ , la température augmente linéairement avec une pente de valeur  $\dot{T}$  telle que  $T(t + \Delta t) = T(t) + \dot{T} \Delta t$  où  $\dot{T}$  le taux de réchauffement de l'air entre deux pas de temps  $\Delta T$ .

$$\dot{e}_T(t) + \epsilon e_T(t) = \zeta - \dot{T} \quad (2.11)$$

Cette équation est une équation différentielle linéaire du premier ordre à coefficient et second membre constants. Les solutions générale, particulière (pour le second membre constant) et finale sont donc respectivement  $e_{T_G}(t) = \Gamma_0 e^{-\epsilon t}$ ,  $e_{T_P}(t) = \Gamma_1$  et  $e_T(t) = \Gamma_0 e^{-\epsilon t} + \Gamma_1$ .

Afin de déterminer la valeur des constantes  $\Gamma_0$  et  $\Gamma_1$ , la solution finale  $e_T(t)$  est injectée dans l'équation 2.11 et conduit à :

$$\Gamma_1 = \frac{\epsilon - \dot{T}}{\epsilon}$$

À l'instant  $t = 0$ , on a aussi :  $e_T(t = 0) = \Gamma_0 + \Gamma_1$  soit  $\Gamma_0 = e_T(0) - \frac{\zeta - \dot{T}}{\epsilon}$

En remplaçant  $\Gamma_0$  et  $\Gamma_1$  par leur valeur dans l'expression de la solution  $e_T$ , on obtient :

$$e_T(t) = e_T(0)e^{-\epsilon t} + \frac{\dot{T} - \zeta}{\epsilon} e^{-\epsilon t} - \frac{\dot{T} - \zeta}{\epsilon} \quad (2.12)$$

Les constantes  $\epsilon$  et  $\zeta$  peuvent être reformulées de façon à distinguer les paramètres géométriques de l'abri des variables météorologiques. Pour cela, on pose  $\epsilon_1 = \frac{A_{sec a}}{V_a}$ ,  $\zeta_1 = \frac{A_a}{V_a C_p}$  et  $\zeta_2 = \frac{A_c}{V_a C_p}$ . Ainsi, après simplification,  $e_T(t)$  devient :

$$e_T(t) = e_T(0)e^{-\epsilon_1 U_a t} + \frac{1}{\epsilon_1} \frac{\dot{T}}{U_a} e^{-\epsilon_1 U_a t} - \frac{\zeta_1 Q_a^*}{\epsilon_1 U_a} e^{-\epsilon_1 U_a t} - \frac{\zeta_2 Q_c^*}{\epsilon_1 U_a} e^{-\epsilon_1 U_a t} - \frac{1}{\epsilon_1} \frac{\dot{T}}{U_a} + \frac{\zeta_1 Q_a^*}{\epsilon_1 U_a} + \frac{\zeta_2 Q_c^*}{\epsilon_1 U_a}$$

Soit après factorisation :

$$e_T(t) = (e^{-\epsilon_1 U_a t} - 1) \left( \frac{1}{\epsilon_1} \frac{\dot{T}}{U_a} - \frac{\zeta_1 Q_a^* + \zeta_2 Q_c^*}{\epsilon_1 U_a} \right) + e_T(0) e^{-\epsilon_1 U_a t} \quad (2.13)$$

Dans le cas de notre campagne de mesure, le rayonnement net reçu respectivement par l'abri et par le capteur ne sont pas mesurés. Pour pallier ce manque, on considère que le terme  $\zeta_1 Q_a^* + \zeta_2 Q_c^*$  peut être approché par  $\zeta_{eq} Q_{\downarrow+\uparrow}$  où  $\zeta_{eq}$  est une constante représentative des paramètres géométriques de l'abri et  $Q_{\downarrow+\uparrow}$  la somme des rayonnements globaux montant et descendant mesurés localement sur le site. La constante de temps  $\tau = \frac{1}{\epsilon_1 U_a}$  est également fixée. Elle correspond physiquement à la rapidité à laquelle une erreur s'établit ou s'annule. Plus la vitesse du vent est élevée, plus une erreur pourra se développer ou disparaître rapidement. L'équation 2.13 devient ainsi :

$$e_T(t) = (e^{-\frac{t}{\tau}} - 1) \left( \frac{1}{\epsilon_1} \frac{\dot{T}}{U_a} - \zeta_{eq} \frac{Q_{\downarrow+\uparrow}}{U_a} \right) + e_T(0) e^{-\frac{t}{\tau}} \quad (2.14)$$

Les mesures étant échantillonnées à chaque pas de temps  $\Delta t = 5 \text{ min} = 300 \text{ s}$ , les variables doivent être discrétisées. On écrit donc l'erreur  $e_T(t_n)$  comme la discrétisation de l'erreur  $e_T(t)$  aux instants  $t_n = n\Delta t$  où  $n$  est le n-ième échantillon de mesure depuis le début de la campagne  $t_0$  :

$$e_T(t_n) = (e^{-\frac{\Delta t}{\tau_n}} - 1) \left( \frac{1}{\epsilon_1} \frac{\dot{T}_n}{U_{a_n}} - \zeta_{eq} \frac{Q_{\downarrow+\uparrow n}^*}{U_{a_n}} \right) + e_T(t_{n-1}) e^{-\frac{\Delta t}{\tau_n}} \quad (2.15)$$

La constante de temps  $\tau_n$  est dépendante de la vitesse du vent  $U_{a_n}$  et est donc propre à chaque échantillon  $t_n$ . Ceci rend la compréhension des phénomènes et la modélisation plus

complexe (termes non-linéaires). Or, lorsque  $\tau_n \ll \Delta t$ , les termes  $e^{-\frac{\Delta t}{\tau_n}}$  tendent vers zéro. Nous ferons dans un premier temps l'hypothèse que le pas de temps  $\Delta t = 300$  s est suffisamment élevé pour considérer chacun des termes  $e^{-\frac{\Delta t}{\tau_n}}$  comme négligeable. Cette hypothèse sera vérifiée par la suite.

L'équation 2.15 est donc simplifiée en l'équation suivante :

$$e_T(t_n) = \zeta_{eq} \frac{Q_{\downarrow+\uparrow n}^*}{U_{a_n}} - \frac{1}{\epsilon_1} \frac{\dot{T}_n}{U_{a_n}} \quad (2.16)$$

Cette équation implique que l'erreur de mesure engendrée par l'utilisation de certains abris ventilés est dépendante du rayonnement global net reçu par l'abri, à la vitesse de l'air qui entre à l'intérieur et à la vitesse de réchauffement de l'air extérieur. Par la suite, nous allons vérifier que les termes de cette équation sont pertinents. L'écriture des indices correspondant à chaque échantillon  $n$  sera abandonnée en vue d'alléger les notations.

### 2.4.3.3 L'abri ventilé mécaniquement : une référence ?

La température de l'air a jusqu'à maintenant toujours été assimilée à la température mesurée par le capteur placé sous l'abri ventilé mécaniquement. Nous allons voir dans cette section les limites de cette hypothèse. Nous distinguerons pour cela les situations diurnes et nocturnes.

#### L'abri TRH, référence pour la mesure des températures nocturnes

La nuit, la différence de température entre l'abri TRH et l'abri ventilé mécaniquement est faible et caractérisée par une moyenne de  $0,092^\circ\text{C}$  et un écart-type de  $0,061^\circ\text{C}$ . Aucune tendance ne se dégage lorsque la différence de température est tracée en fonction des autres variables météorologiques de l'équation 2.16. Cet écart tend seulement vers sa valeur moyenne lorsque la vitesse de vent augmente (figure 2.28).

L'écart-type de l'écart ( $0,061^\circ\text{C}$ ) reste faible au regard de son incertitude ( $0,1414^\circ\text{C}$  - définie comme la somme quadratique de la précision ( $0,1$  K) des deux capteurs utilisés). Il est donc assimilé à l'incertitude de mesure. L'écart moyen ( $0,092^\circ\text{C}$ ) peut quant à lui être interprété de deux manières :

- l'un des capteurs ou les deux capteurs sont mal calibrés : une erreur permanente est donc associée à la mesure de l'écart ;
- l'abri à ventilation forcée aspire de l'air frais en provenance des couches inférieures ce qui fausse la mesure.



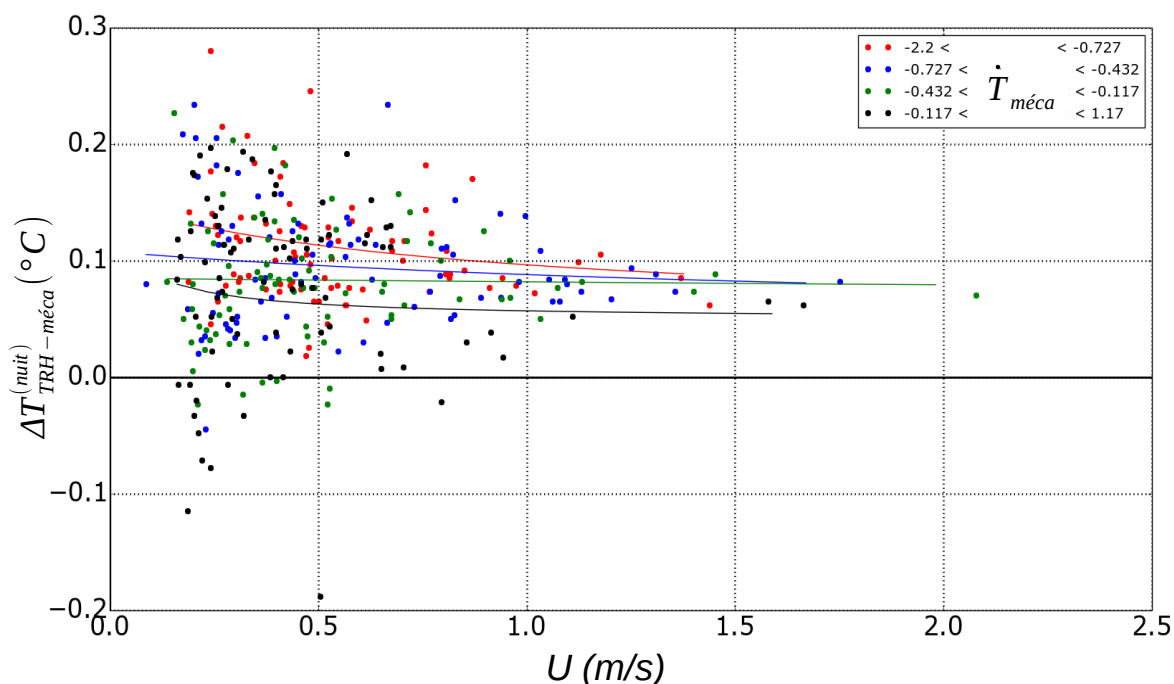


FIGURE 2.28 – Évolution de la différence de température nocturne mesurée entre l’abri TRH et l’abri ventilé mécaniquement en fonction de la vitesse du vent et après classification par intervalles de taux de réchauffement de l’air

La différence de température entre l’abri URBIO et l’abri mécaniquement ventilé est ensuite tracée en fonction du vent et du taux de réchauffement (figure 2.29).

Une tendance nette est ici observée : la différence est presque toujours positive et elle augmente lorsque le vent ou le taux de réchauffement diminuent. Cette observation est cohérente avec l’équation 2.16 : la température mesurée sous l’abri URBIO est biaisée en raison du faible renouvellement d’air qui a lieu en son sein en regard de son volume. Plus le taux de réchauffement tend vers zéro ou plus la vitesse du vent est élevée, plus l’erreur tend vers une ordonnée à l’origine positif du même ordre de grandeur que celui relevé entre l’abri TRH et l’abri ventilé mécaniquement ( $0,092^{\circ}\text{C}$ ). On considérera donc par la suite que la valeur de cette ordonnée à l’origine est entièrement imputable au système {abri à ventilation forcée + capteur}, quelle que soit son origine.

Au regard de ses performances, l’abri TRH sera donc par la suite considéré comme un abri de référence la nuit.

### L’abri URBIO, référence pour la mesure des températures diurnes

L’analyse de la figure 2.25 a permis de mettre en évidence que l’abri TRH mesurait presque toujours les températures diurnes les plus élevées, suivi de l’abri ventilé mécaniquement

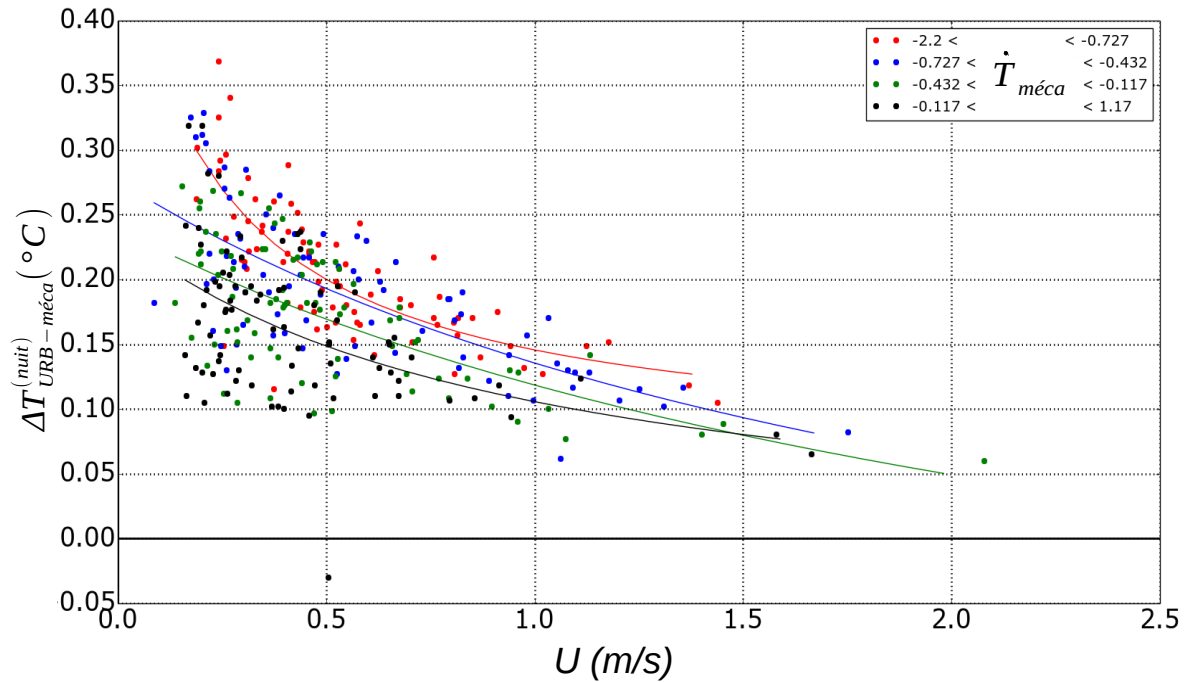


FIGURE 2.29 – Évolution de la différence de température nocturne mesurée entre l’abri URBIO et l’abri ventilé mécaniquement en fonction de la vitesse du vent et après classification par intervalles de taux de réchauffement de l’air

puis enfin de l’abri URBIO. Le standard ISO17714 :2007 stipule que les abris dont les mesures de température sont les plus faibles pendant le jour ont de fortes chances d’être les plus proches de la température d’air réelle (ISO 2007). Si l’abri TRH ne peut clairement pas prétendre au titre d’abri de référence diurne, le choix entre l’abri URBIO et l’abri ventilé mécaniquement n’est pas aussi trivial. L’évolution de la différence de température entre ces deux abris est tracée en fonction du rayonnement et de la vitesse du vent (figure 2.30) afin d’analyser les données au regard de l’équation 2.16.

La différence de température (assimilée à l’erreur  $e_T(t_n)$ ) décroît en fonction de l’inverse de la vitesse du vent et tend pour les fortes valeurs de vitesse vers une erreur résiduelle négative qui est variable selon la valeur du rayonnement. Or, au vu de l’équation 2.16 (et donc en considérant l’abri à ventilation forcée comme parfait), les courbes de tendance devraient avoir une ordonnée à l’origine nulle lorsque la vitesse du vent tend vers l’infini (ou égale à celle observée durant la nuit). Elles devraient ainsi être de la forme :

$$y = \eta \frac{1}{U_a} + \mu \quad (2.17)$$

avec  $\eta$  le coefficient directeur qui varie selon la valeur du rayonnement ( $\eta = \eta_0 Q_{\downarrow+\uparrow}$ ),  $\mu$  l’or-

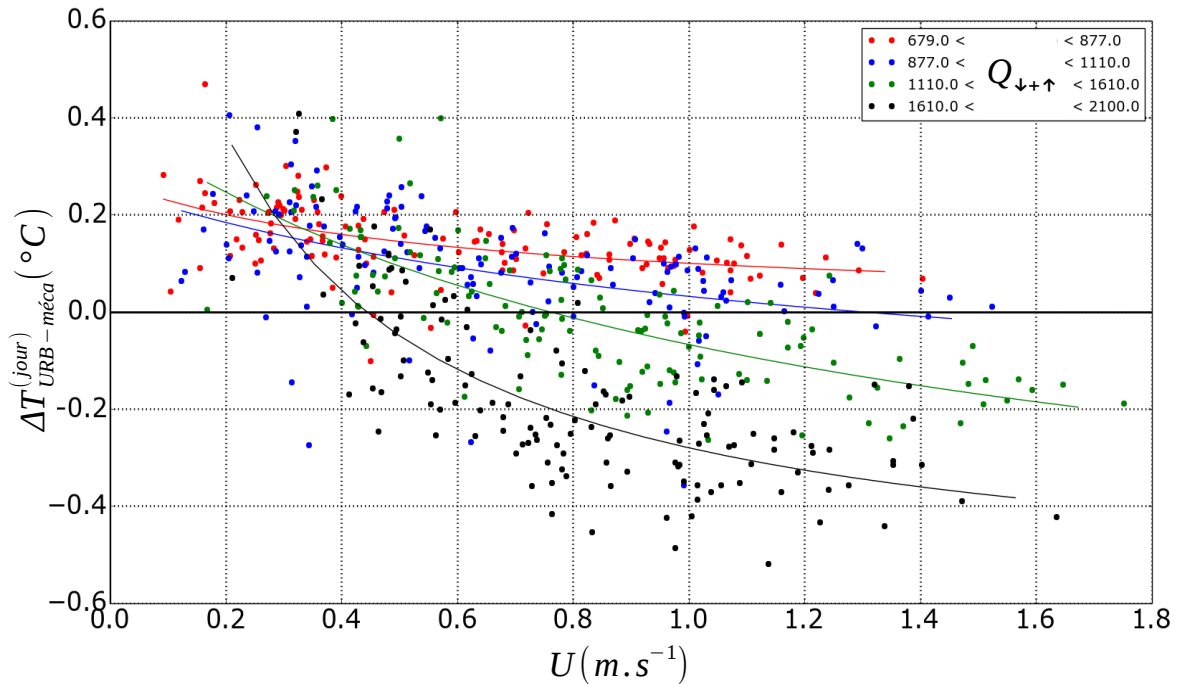


FIGURE 2.30 – Évolution de la différence de température diurne mesurée entre l’abri URBIO et l’abri ventilé mécaniquement en fonction de la vitesse du vent et après classification par intervalles de rayonnement solaire montant et descendant

donnée à l’origine lorsque la vitesse du vent tend vers l’infini.

L’ordonnée à l’origine  $\mu$  devrait être nulle ou égale à la valeur observée durant la nuit. Or, la figure 2.30 montre que  $\mu$  est variable et dépendant de la valeur du rayonnement. De plus, pour de fortes valeurs de rayonnement, la température mesurée sous l’abri à ventilation forcée est plus élevée que celle mesurée sous l’abri URBIO. Ceci implique que l’abri à ventilation forcée surchauffe sous un rayonnement important quelle que soit la vitesse du vent. Il mesure donc une température erronée avec une erreur  $e_{T_{meca}}$  assimilable à :

$$e_{T_{meca}} = \nu_{meca} Q_{\downarrow+\uparrow} \quad (2.18)$$

avec  $\nu_{meca}$  un coefficient fonction des paramètres géométriques de l’abri mais également de la vitesse d’aspiration mécanique de l’air à travers le tube vertical de l’abri.

Cela implique que la différence de température  $\Delta T_{nat-meca}$  entre un abri ventilé naturellement et l’abri ventilé mécaniquement est de la forme :

$$\Delta T_{nat-meca} = \zeta_{eq} \frac{Q_{\downarrow+\uparrow}}{U_a} - \frac{1}{\epsilon_1} \frac{\dot{T}}{U_a} - \nu_{meca} Q_{\downarrow+\uparrow} + \kappa \quad (2.19)$$

avec  $\kappa$  une constante qui est ajoutée pour mieux correspondre aux observations.

Au final, le jour, la température mesurée sous l'abri URBIO dans des conditions de vent faible ( $< 0,5$  m/s) est légèrement supérieure (environ  $0,15^\circ\text{C}$ ) à celle mesurée sous l'abri mécanique. Au contraire, pour des vitesses de vent supérieures à  $0,5$  m/s et sous fort rayonnement ( $Q_{\downarrow+\uparrow} > 1600\text{W}/\text{m}^2$ ), l'air dans l'abri URBIO se réchauffe moins que dans l'abri mécanique. Les différences pour des vents supérieurs à  $1$  m/s sont de l'ordre de  $0,3$  à  $0,4^\circ\text{C}$ . L'abri URBIO sera donc considéré par la suite comme l'abri de référence pour les mesures de jour.

#### 2.4.3.4 Confrontation du modèle de l'erreur aux observations

Afin de déterminer la pertinence des modèles proposés, l'équation 2.19 est confrontée directement aux données en déterminant la valeur des coefficients  $\zeta_{eq}$ ,  $\epsilon_1$ ,  $\nu_{meca}$  et  $\kappa$  par régression linéaire. Les régressions sont réalisées pour chaque période de la journée (jour et nuit) et pour chacun des deux abris ventilés naturellement. Les résultats sont présentés dans le tableau 2.6.

Coefficient	TRH / mécanique		URBIO / mécanique	
	jour	nuit	jour	nuit
$\zeta_{eq}$	0,0000609	-0,000004*	0,0000724	0,0000182
$\epsilon_1$	-0,0457	0,018	0,0169	0,019
$\nu_{meca}$	-0,000698	-0,00067	0,00035	0,00034
$\kappa$	-0,456	-0,44	0,29	0,39

\* coefficient pour lequel la statistique  $p_{val} > 0,05$  (significativité faible)

TABLE 2.6 – Valeurs des coefficients de l'équation 2.19 déterminées par régression linéaire

Tous les coefficients pour l'abri URBIO sont positifs, en cohérence avec la formulation de l'équation 2.19. De plus, tous les termes sont significatifs (pour  $p_{val} < 0,05$ ). Seul le coefficient  $\zeta_{eq}$ , qui correspond au terme de rayonnement de l'abri à ventilation naturelle, voit sa valeur changer d'ordre de grandeur entre le jour et la nuit. Ce terme dérivant d'une grossière approximation du terme initial  $\zeta_1 Q_a^* + \zeta_2 Q_c^*$ , une telle différence dans la valeur du coefficient n'est pas étonnante.

Tous les coefficients pour l'abri TRH ne sont pas positifs, que ce soit le jour ou la nuit. Le jour, seul le coefficient  $\zeta_{eq}$  est positif, en cohérence avec l'équation 2.19. Les termes en rayonnement étant prédominants durant le jour pour cet abri, il n'est pas impossible que le coefficient  $\epsilon_1$ , en théorie positif, soit mal évalué. Le vrai problème vient en fait du coefficient

$\mu_{mecca}$ , associé à un terme radiatif, dont la valeur aurait dû être positive. Le fait que cette valeur soit négative interroge sur l'exhaustivité de l'équation 2.19. Il semblerait qu'un terme de la forme  $\nu_{nat}Q_{\downarrow+\uparrow}$  (où l'influence de la vitesse du vent est donc considérée comme négligeable) doive être ajouté à l'équation 2.19 afin de mieux représenter la sensibilité au rayonnement des abris ventilés naturellement. Cette hypothèse est cohérente avec l'analyse que l'on peut faire de la figure 2.31 : aussi élevée soit la vitesse du vent, il n'est pas possible de réduire l'écart de température enregistré entre l'abri TRH et l'abri mécanique lorsque le rayonnement est élevé ( $>$  à  $1800 \text{ W/m}^2$ ). Un terme de la forme  $\nu_{nat}Q_{\downarrow+\uparrow}$  sera finalement ajouté aux termes d'erreur des abris ventilés naturellement (terme qui sera a priori négligeable le jour pour l'abri URBIO mais prépondérant pour l'abri TRH).

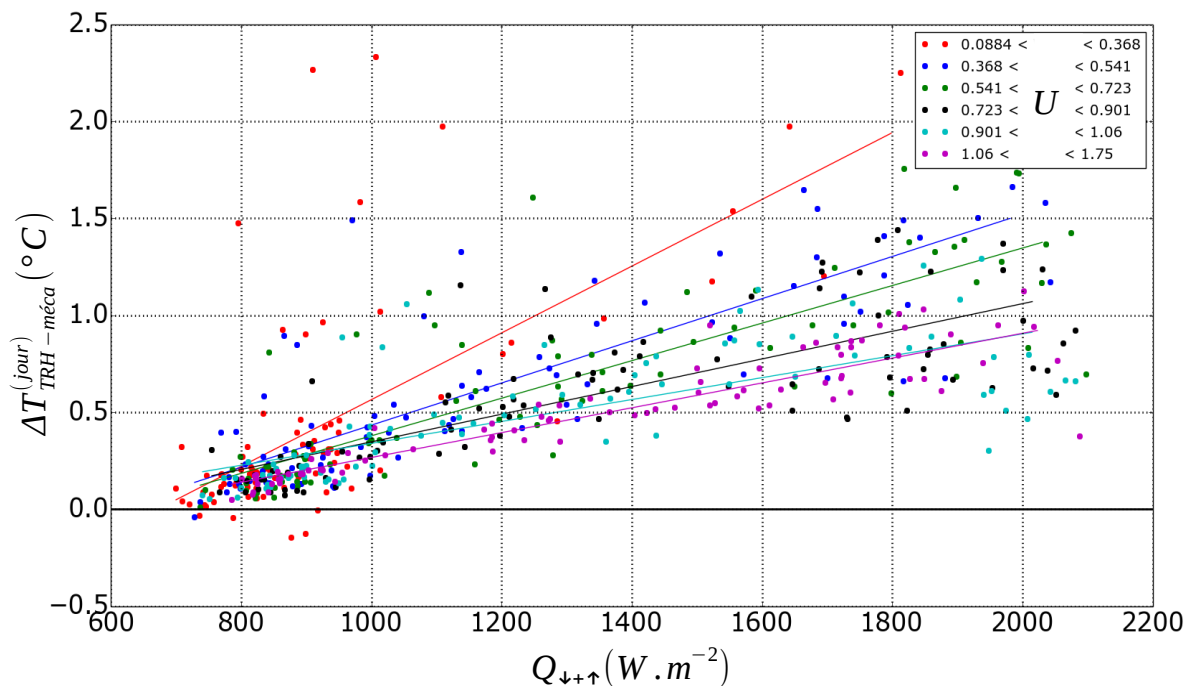


FIGURE 2.31 – Évolution de la différence de température diurne mesurée entre l'abri TRH et l'abri ventilé mécaniquement en fonction du rayonnement global descendant et montant et après classification par intervalles de vitesse du vent.

L'erreur de mesure due à l'utilisation d'un abri ventilé naturellement (noté par l'indice  $abri$ ) sera donc finalement de la forme :

$$e_{T_{abri}}(t) = \zeta_{eq} \frac{Q_{\downarrow+\uparrow}}{U_a} - \frac{1}{\epsilon_1} \frac{\dot{T}}{U_a} + \mu_{nat} Q_{\downarrow+\uparrow} \quad (2.20)$$

### 2.4.3.5 Calibration d'un modèle de correction

Pour rappel, l'objectif de l'intercomparaison de nos abris est simple : elle permet d'évaluer l'erreur de mesure engendrée par l'utilisation d'abris ventilés (naturellement ou mécaniquement). Afin de corriger les jeux de données présentés section 2.3, lorsque c'est possible, cette erreur est modélisée pour chacun des abris utilisés. Cependant, la température réelle de l'air, nécessaire pour modéliser l'erreur, ne peut pas être connue avec certitude. En effet, dans tous les abris utilisés pour l'intercomparaison, les mesures sont entachées d'une erreur plus ou moins grande selon les conditions météorologiques :

- dans l'abri TRH, on mesure des températures largement trop élevées la journée par rapport à celle mesurée dans les deux autres abris (surtout sous fort rayonnement solaire et faible vitesse de vent) ;
- dans l'abri URBIO, on mesure des températures légèrement trop élevées la nuit mais également le jour lorsque la vitesse du vent est faible ;
- dans l'abri ventilé mécaniquement, on surestime (par rapport à l'abri URBIO) les températures sous fort rayonnement et fortes vitesses de vent.

Dans ce cas, la vérification des modèles d'erreur définis pour les abris ventilés naturellement par l'équation 2.20 semble impossible à mettre en oeuvre. Afin de tout de même prendre en compte les différences de performances liées au type d'abri utilisé, des modèles d'erreur sont établis, en considérant un abri de référence (abri dont le capteur mesure a priori la température au plus proche de la température réelle de l'air) pour chaque période de la journée. Les abris TRH et URBIO sont ainsi définis respectivement comme abris de référence nocturne et diurne. À l'abri URBIO sera donc associé un modèle d'erreur nocturne tandis que l'abri TRH aura un modèle d'erreur diurne. Les modèles d'erreur, une fois vérifiés, seront ainsi utilisés pour corriger les mesures de tous les jeux de données afin de les rendre intercomparables.

La vitesse du vent  $U$  mesurée localement est supposée proportionnelle à la vitesse du vent  $U_a$  entrant dans l'abri. Finalement, les modèles d'erreur proposés sont les suivants :

$$e_{T_{TRH},\text{jour}} = \omega_{1_{TRH}} \frac{Q_{\downarrow+\uparrow}}{U} - \omega_{2_{TRH}} \frac{\dot{T}}{U} + \omega_{3_{TRH}} Q_{\downarrow+\uparrow} + \omega_{4_{TRH}} \quad (2.21)$$

$$e_{T_{URB},\text{nuit}} = \omega_{1_{URB}} \frac{Q_{\downarrow+\uparrow}}{U} - \omega_{2_{URB}} \frac{\dot{T}}{U} + \omega_{3_{URB}} Q_{\downarrow+\uparrow} + \omega_{4_{URB}} \quad (2.22)$$

avec  $e_{T_{TRH,jour}}$  l'erreur de température due à l'abri TRH durant la période diurne,  $e_{T_{URB,nuit}}$  l'erreur de température due à l'abri URBIO durant la période nocturne,  $\omega_{1_{abri}} = \zeta_{eq_{abri}} - \zeta_{eq_{ref}}$ ,  $\omega_{2_{abri}} = \frac{1}{\epsilon_{1_{abri}}} - \frac{1}{\epsilon_{1_{ref}}}$ ,  $\omega_{3_{abri}} = \nu_{nat_{abri}} - \nu_{nat_{ref}}$  pour l'abri *abri* par rapport à l'abri de référence durant cette période de la journée,  $\omega_{4_{abri}}$  une constante qui est ajoutée à chaque modèle pour la régression.

Ces équations diffèrent des formulations proposées par Nakamura et Mahrt (2005) ou Cheng et al. (2014). Nakamura et Mahrt (2005) proposent de modéliser l'erreur par un polynôme d'ordre deux :

$$e_T = \phi_0 + \phi_1 X_{naka} + \phi_2 X_{naka}^2 \quad (2.23)$$

$$X_{naka} = \frac{Ray}{\rho C_p T_{ref} U}$$

avec  $\phi_0$ ,  $\phi_1$  et  $\phi_2$  les coefficients de régressions du modèle,  $Ray$  le rayonnement solaire net le jour et le rayonnement GLO net la nuit mesuré à 1 m du sol ( $W/m^2$ ),  $\rho$  la masse volumique de l'air ( $kg/m^3$ ),  $T_{ref}$  la température mesurée sous l'abri de référence ( $^{\circ}C$ ) et  $U$  la vitesse du vent mesurée à 1 m du sol ( $m/s$ ).

Ils montrent que le terme du second ordre est négligeable et donc que l'erreur peut être modélisée seulement par  $y = \phi_0 + \phi_1 X_{naka}$ .

Pour la période diurne, Cheng et al. (2014) proposent de reprendre cette dernière formulation en pondérant la variable  $X_{naka}$  par l'angle solaire zénithal  $Z$  (en  $^{\circ}$ ) :

$$X_{chen} = \frac{Ray}{\rho C_p T_{ref} U} \frac{Z}{180} \quad (2.24)$$

Les différences qui existent entre ces modèles et les modèles proposés dans cette thèse sont principalement le nombre de variables explicatives utilisées et la période de validité des modèles.

- Les modèles Cheng et Nakamura n'utilisent qu'une seule variable explicative, issue de la combinaison de trois variables météorologiques. Notre modèle combine les trois mêmes variables météorologiques (le terme radiatif n'est pas tout-à-fait le même) mais pour créer trois variables explicatives.
- Le modèle Nakamura ainsi que le modèle que nous proposons sont valables de jour comme de nuit tandis que le modèle Cheng est valable exclusivement le jour.

Afin de comparer les performances de ces trois modèles d'erreur (que l'on appellera

par la suite modèle Nakamura, modèle Cheng et modèle proposé), ils seront calibrés à partir des données de la campagne de diagnostic et de modélisation de l'erreur (utilisées jusqu'alors pour établir les modèles). Dans un deuxième temps, les modèles seront recalibrés à partir de données météorologiques mesurées à une station Météo-France. Nous verrons par la suite les motivations de cette décision.

L'intégralité de la campagne de diagnostic n'est pas utilisée pour la calibration. Les données de rayonnement sont utilisées pour expliquer les erreurs de mesure causées par les abris. Or, le capteur de rayonnement n'est pas positionné à la même hauteur du sol et à la même distance du mur que les autres capteurs (figure 2.24). Du fait que le mur puisse constituer un masque solaire et une source de réflexion, deux types de périodes de mesure sont retirées des données :

- les périodes où le capteur de rayonnement est au soleil pendant que les abris sont à l'ombre (ou bien le contraire) ;
- les périodes où le capteur de rayonnement subit la réflexion spéculaire de la vitre alors que les abris ne la subissent pas (ou bien le contraire).

Ces périodes sont déterminées géométriquement en utilisant les côtes des bâtiments, vitres, distance des appareils au mur et au sol, taille des appareils, etc. Les résultats de ces calculs théoriques sont confrontés à l'observation des signaux de température et de rayonnement. Des périodes légèrement plus larges sont finalement retirées de l'analyse : elles correspondent grossièrement (car elles varient selon chaque journée) aux plages horaires de 07h30 à 09h30 et de 16h05 à 17h30.

Les modèles sont calibrés deux fois : une première fois à partir de variables météorologiques mesurées localement (sur le site de mesure), le second à partir de variables météorologiques mesurées à la station Météo-France la plus proche et dont les valeurs sont représentatives du type de temps à l'échelle de l'agglomération. Pour rappel, notre objectif principal est de corriger des données de température acquises pendant plusieurs années en plusieurs points du territoire (plusieurs stations). Calibrer les modèles proposés aux équations 2.21 et 2.22 à partir de données météorologiques locales permet de juger la pertinence de chacun des termes mais ne permet pas de corriger les données a posteriori. En effet, ces modèles reposeraient sur des données de vent, de rayonnement et de température acquises :

- seulement pour un site de mesure (TRH06) ;



- pendant une durée limitée (quelques semaines).

Pour corriger l'ensemble des sites à partir de mesures locales, il aurait fallu mesurer chacune des grandeurs physiques utilisée dans les équations 2.21 et 2.22 sur chaque site pendant plusieurs années, ce qui n'est pas le cas. Les grandeurs physiques mesurées sur un site Météo-France sont par contre représentatives des conditions météorologiques de l'agglomération. Elles sont également disponibles pendant les périodes de mesure qui nous intéressent. Leur mesure présente enfin l'avantage d'être standardisée pour tous les sites Météo-France. Les modèles calibrés à partir de ces données pourront ainsi être appliqués à n'importe quelle autre agglomération française si besoin.

### Utilisation de données locales

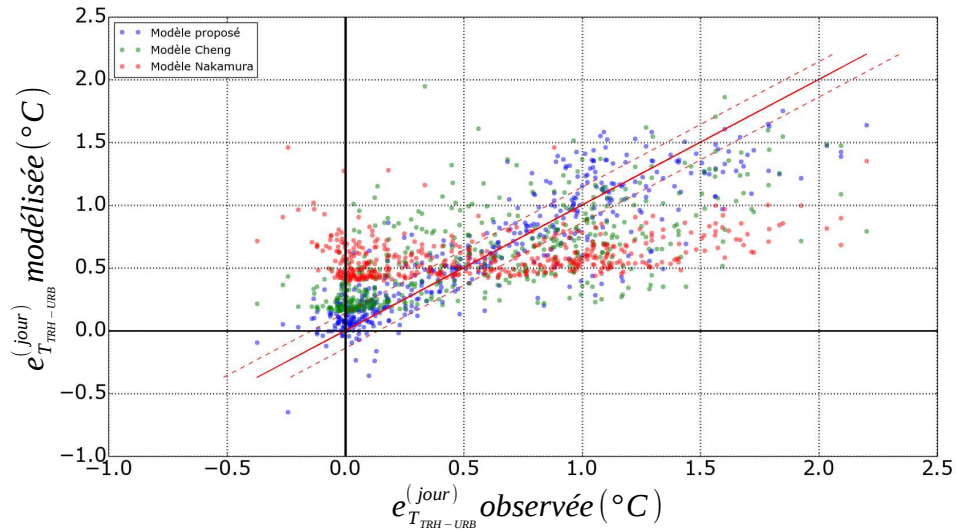
Dans un premier temps, les données météorologiques mesurées dans le cadre de la campagne de diagnostic sont utilisées pour calibrer les modèles. Le rayonnement global montant et descendant  $Q_{\downarrow+\uparrow}$  est choisi comme variable commune représentative du rayonnement reçu par l'abri tant pour la période diurne que nocturne. Le résultat de la régression des trois modèles est tracé en fonction des données observées figure 2.32a (pour les données diurnes) et 2.32b (pour les données nocturnes). Comme défini auparavant, l'abri TRH est pris comme référence la nuit tandis que l'abri URBIO est pris comme référence le jour.

La figure 2.32b (illustrant les résultats de la régression nocturne) met en évidence que la variabilité de l'erreur est du même ordre de grandeur que l'incertitude de mesure. Cela contraint les performances des modèles nocturnes, qui ont donc peu de chance d'expliquer une très grande partie de la variance.

Le coefficient de détermination ( $R^2$ ) prend selon la période de la journée et le modèle utilisé les valeurs suivantes :

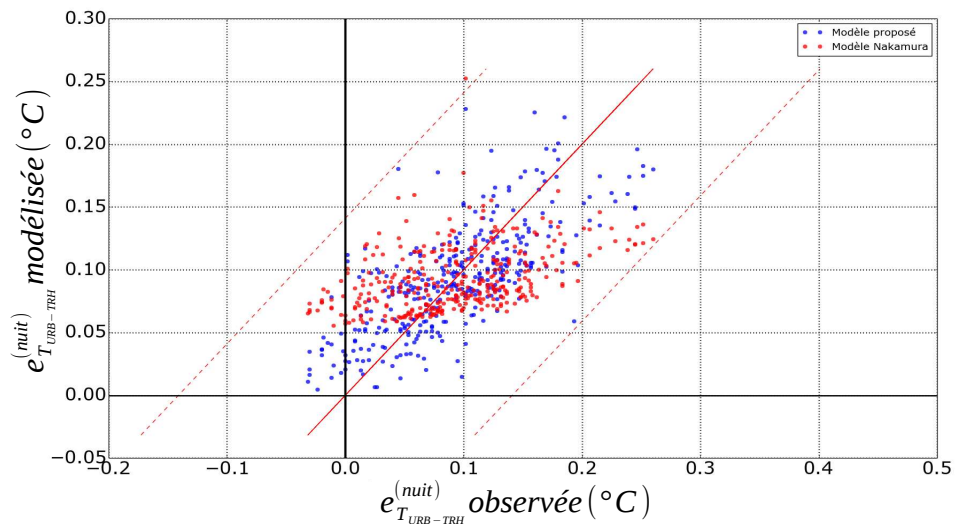
- le jour : 0,80 pour le modèle proposé, 0,46 pour le modèle Cheng et 0,09 pour le modèle Nakamura ;
- la nuit : 0,54 pour le modèle proposé et 0,18 pour le modèle Nakamura.

De jour comme de nuit, le modèle proposé explique mieux la variance des données que les deux autres, même si les valeurs d'erreur les plus élevées sont légèrement sous-évaluées. En période diurne, du fait de l'utilisation de l'angle zénithal dans la variable explicative, le modèle Cheng est plus approprié que le modèle Nakamura.



pointillés rouges : incertitude sur la différence de température associée à la précision des capteurs

(a)



pointillés rouges : incertitude sur la différence de température associée à la précision des capteurs

(b)

FIGURE 2.32 – Erreur (a) diurne et (b) nocturne estimée à partir des trois modèles étudiés en fonction de l'erreur observée. Les données météorologiques utilisées en entrée des modèles sont les données mesurées *localement* sur le site

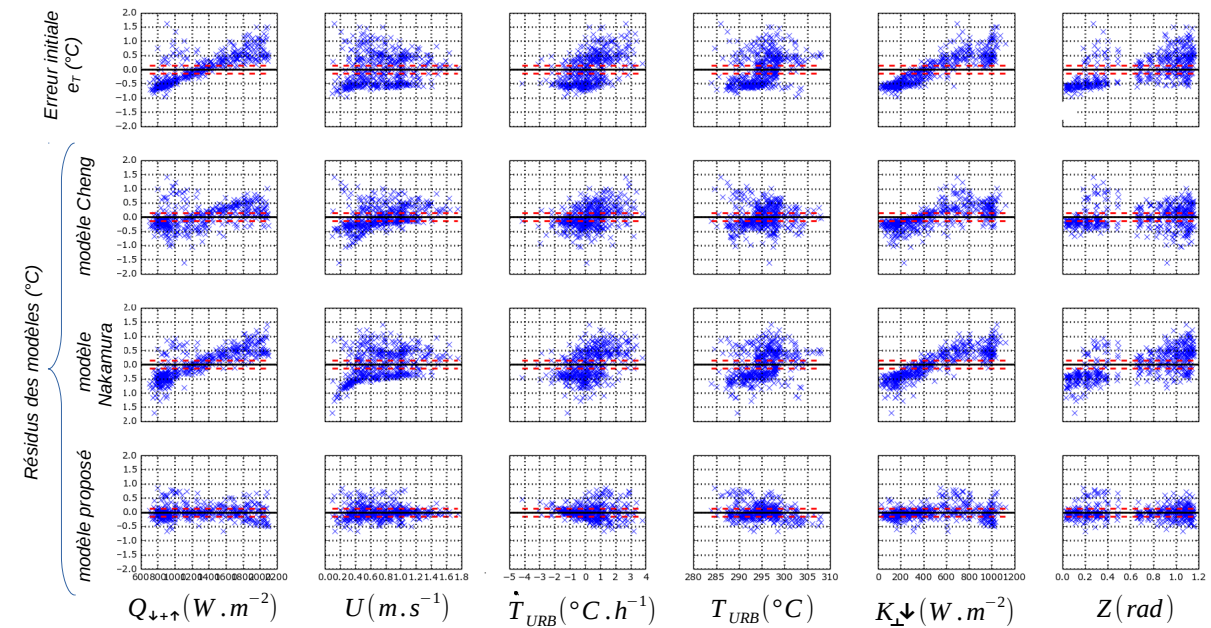
Les résidus de chaque modèle correspondent à la différence entre les valeurs observées et les valeurs modélisées. Ils sont représentés (figure 2.33a pour les données diurnes et figure 2.33b pour les données nocturnes) au regard des variables météorologiques utilisées pour le calcul des variables explicatives (rayonnement global montant et descendant, vitesse de vent, taux de réchauffement, température de l'air). Deux autres variables sont maintenant ajoutées à cette analyse : l'angle solaire (utilisé par le modèle de Cheng) et le rayonnement solaire reçu sur une surface perpendiculaire aux rayons du soleil (noté  $K_{\perp} \downarrow$ ). Cette dernière variable est utilisée

pour pallier au fait que le rayonnement est mesuré sur une surface horizontale alors que l'abri est constitué de surfaces horizontales mais également de surfaces verticales. Le rayonnement total reçu par l'abri est donc sûrement mieux estimé par le rayonnement reçu sur une surface perpendiculaire aux rayons du soleil plutôt que sur une surface horizontale, cette affirmation étant plus ou moins vraie selon les rapports hauteur / largeur de chaque abri. L'augmentation brutale de l'erreur observée dès le lever du soleil (figure 2.25) est également observée par Nakamura et Mahrt (2005) ainsi que par Cheng et al. (2014). En première analyse, elle peut être associée au rayonnement reçu par une surface perpendiculaire au rayonnement (ce qui expliquerait que dès qu'il y a du soleil, l'erreur est très forte). Cependant, l'utilisation de ce rayonnement plutôt que du rayonnement global montant et descendant ne permet pas d'améliorer les performances des modèles. Il est malgré tout utilisé dans l'analyse des résidus afin de veiller à ce qu'après régression, aucune erreur explicable par ce rayonnement ne subsiste.

De jour, une relation linéaire est identifiable entre l'erreur avant régression et la variable  $Q_{\downarrow+\uparrow}$ . Des relations existent également entre l'erreur et chacune des autres variables mais la linéarité de ces relations est moins évidente car l'erreur suit plusieurs tendances selon la plage de valeurs de la variable météorologique considérée. L'évolution de l'erreur en fonction de  $K_{\perp} \downarrow$  (figure 2.33a) pourrait ainsi être découpée en trois modèles linéaires : un premier de pente nulle correspondant à la plage  $[0 - 200]W/m^2$ , un second de pente positive pour la plage  $[200 - 800]W/m^2$  et un dernier de pente nulle pour la plage  $[800 - 1200]W/m^2$ .

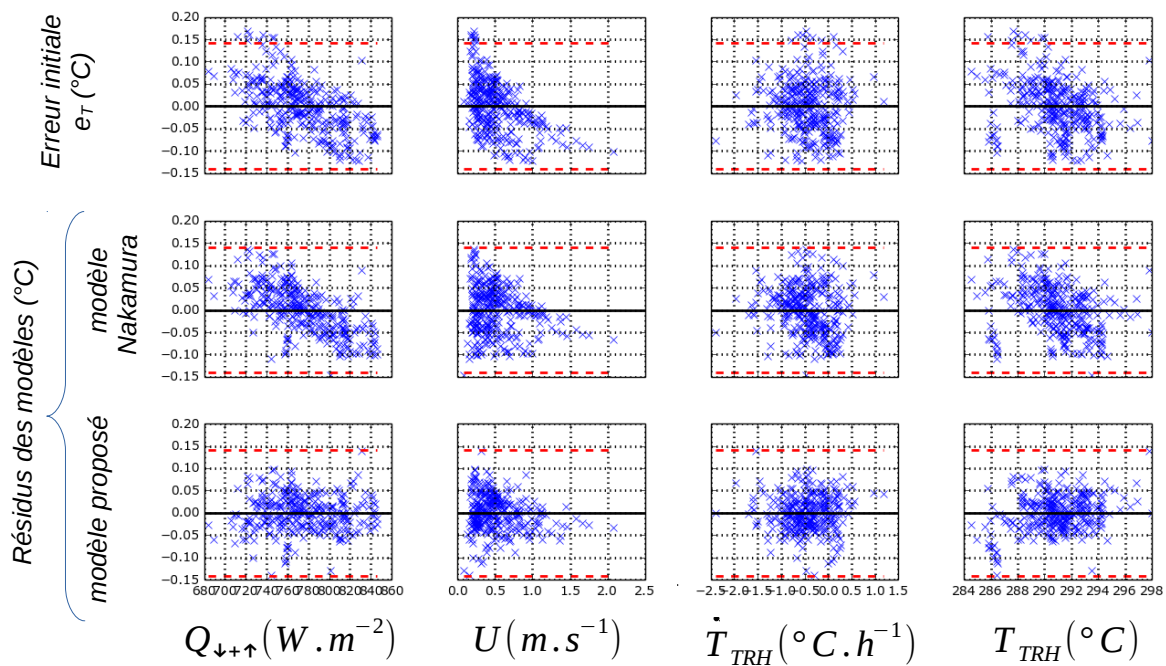
Le modèle Nakamura ne contribue que très peu à distribuer normalement les résidus le long des variables étudiées. Par le simple ajout du terme d'angle solaire, le modèle Cheng améliore de façon notable cette situation : les résidus tendent vers une distribution normale quelle que soit la variable étudiée (sauf éventuellement l'angle solaire). Le modèle proposé est donc le seul qui permet de distribuer normalement les résidus pour chacune des variables météorologiques, même si certaines valeurs s'éloignent notablement (environ  $0,75^{\circ}C$ ) de la tendance. Ces points correspondent toujours à des angles solaires zénithaux compris entre 0,2 et 0,5 radians ( $11,5$  et  $28,9^{\circ}$ ), soit généralement les périodes de 6h à 7h pour ce mois de l'année. Ils ont la particularité d'être enregistrés pour des journées ensoleillées (cf. figure 2.25) : les 9, 10, 11, 24 et 29 juillet 2016 sont en effet les seules journées où les résidus de l'erreur dépassent  $0,5^{\circ}C$  quand l'angle solaire zénithal est compris entre 0,2 et 0,5 radians. Cette gamme d'angles solaires est probablement critique pour l'abri TRH, correspondant peut-être à une situation où le rayonnement pénètre plus facilement dans l'abri.

De nuit, le modèle proposé est à nouveau le modèle qui semble avoir les résidus les



pointillés rouges : incertitude sur la différence de température associée à la précision des capteurs

(a)



pointillés rouges : incertitude sur la différence de température associée à la précision des capteurs

(b)

FIGURE 2.33 – Résidus de l’erreur (a) diurne et (b) nocturne selon le modèle utilisé et en fonction de chaque variable météorologique  $VM_m$ . Les données météorologiques utilisées sont les données mesurées *localement* sur le site.

plus normalement distribués. Seuls les résidus associés à la vitesse du vent conservent une tendance. Quelques valeurs échappent totalement au modèle proposé, sans que leur recontextualisation ne permette d’expliquer la cause. Elles surviennent après un refroidissement brutal de l’air (taux de réchauffement de  $-1,5^{\circ}\text{C}/\text{h}$ ) pendant une période de stabilisation voire de légère

augmentation de la température (taux de réchauffement de  $0,3^{\circ}\text{C/h}$ ) et associée à une vitesse de vent faible (inférieure à  $0,5\text{ m/s}$ ). Cette combinaison de conditions, rare pour un début de nuit, pourrait expliquer pourquoi ces valeurs échappent à nos modèles. Cependant, l'approximation qui a consisté à simplifier l'équation 2.15 en 2.16 n'est pas à l'origine de la mauvaise prédiction de notre modèle. En effet, cette approximation considère que les constantes de temps sont négligeables devant le temps d'échantillonnage. Dans certaines conditions (comme celle du cas présent), cette hypothèse pourrait ne plus être vérifiée. Mais dans ce cas, l'abri TRH devrait suivre plus facilement les variations de la température de l'air que l'abri URBIO qui les amortirait. Dans le cas du refroidissement brutal, le fait d'utiliser l'équation non simplifiée devrait donc a priori conduire à une sous-estimation encore plus importante de l'erreur.

Les valeurs calculées pour les trois termes du modèle proposé sont toutes significatives (pour  $p_{val} < 0,05$ ). L'analyse de la valeur prise par chacun des coefficients permet d'identifier les variables explicatives qui prédominent dans l'explication de l'erreur ou au contraire celles qui sont négligeables. Afin de réaliser cette comparaison, il est nécessaire que chacune des variables explicatives soit normalisée. Ceci correspond dans notre cas à centrer et réduire chaque variable, c'est-à-dire lui soustraire sa moyenne et diviser par son écart-type afin qu'elle suive une loi normale de moyenne 0 et d'écart-type 1 ( $N(0, 1)$ ). Cette étape repose sur l'hypothèse que les variables utilisées suivent une loi normale. Nous considérerons que c'est le cas pour nos variables. Après centrage et réduction, la régression peut être réalisée et les coefficients comparés. Le tableau 2.7 récapitule la valeur prise par chaque coefficient pour le jour ainsi que pour la nuit selon les données utilisées pour le calcul des variables explicatives (locales ou régionales). De jour, le terme de rayonnement prédominant est celui qui ne dépend pas de la vitesse (l'estimation du coefficient de l'autre terme de rayonnement  $\omega_1$  n'étant pas significatif). La valeur normalisée de son coefficient ( $\omega_3 = 0,41$ ) est plus de deux fois supérieure à celle du dernier terme variable  $\omega_2 = 0,17$ . De nuit, toutes les valeurs des coefficients sont significatives, même si le terme qui dépend du taux de réchauffement est négligeable devant les deux autres (sa valeur - lorsque les variables sont centrées et réduites - est de  $0,0047$  contre  $0,024$  et  $-0,033$  pour les deux autres termes variables). Le signe des deux termes de rayonnement diffère, rendant complexe l'analyse de sensibilité au rayonnement d'un abri par rapport à l'autre.

Pour conclure, aucun terme du modèle proposé ne paraît inutile (si l'on considère de manière globale les modèles de jour comme de nuit). Le modèle proposé est de jour comme de nuit le modèle le plus performant pour expliquer les erreurs de mesure. Deux hypothèses, qui sont ambivalentes, peuvent justifier ces meilleurs résultats.

- Les variables explicatives proposées sont plus appropriées au problème que celles utilisées par les modèles Nakamura et Cheng.
- Seul le nombre de variables explicatives plus important utilisé par le modèle proposé est responsable de ces meilleurs résultats. Si c'est le cas, le modèle surajuste les données. Ses performances devraient donc être limitées dans le cas d'une utilisation avec un autre jeu de données.

Afin de vérifier que le modèle proposé est bien le meilleur (que le jeu de données utilisé pour la calibration n'est pas seulement surajusté), les modèles calibrés dans cette section sont vérifiés à l'aide d'un nouveau jeu de données. Si les estimations qui seront calculées à partir du modèle proposé sont plus éloignées des valeurs observées que les estimations calculées à partir des deux autres modèles, on considérera la deuxième hypothèse comme vérifiée. Dans le cas contraire, les variables explicatives utilisées dans le modèle proposé seront considérées comme les plus appropriées au problème (hypothèse 1).

#### Utilisation de données régionales

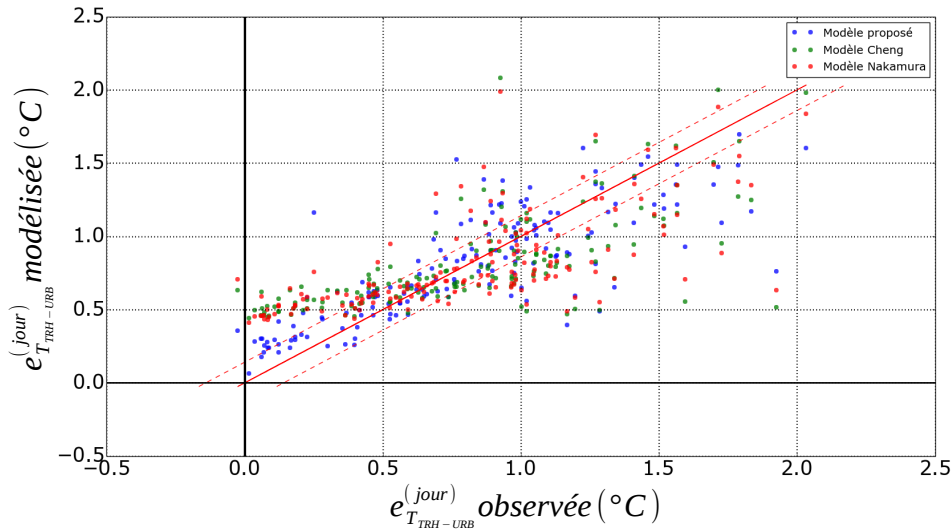
les équations 2.21 et 2.22 sont recalibrés à partir de données valables à l'échelle de l'agglomération. Le rayonnement GLO et le rayonnement solaire montant n'étant pas mesurés par Météo-France, le rayonnement solaire descendant  $K \downarrow$  est utilisé en remplacement de  $Q_{\downarrow+\uparrow}$  dans l'équation 2.21 (erreur de mesure de température diurne). Pour la période nocturne, par simplicité d'utilisation, les données de nébulosité sont utilisées en substitution du rayonnement. En effet, la nébulosité est directement liée au rayonnement GLO (Matzarakis et al. 2010 ; Monteith et Unsworth 2007 ; Oke 1987), même si la relation entre ces deux variables n'est pas linéaire et si elle dépend également d'autres variables météorologiques.

Les estimations issues de la régression des trois modèles en fonction des valeurs observées sont représentées figure 2.34a (pour les données diurnes) et 2.34b (pour les données nocturnes).

Le coefficient de détermination ( $R^2$ ) prend selon la période de la journée et le modèle utilisé les valeurs suivantes :

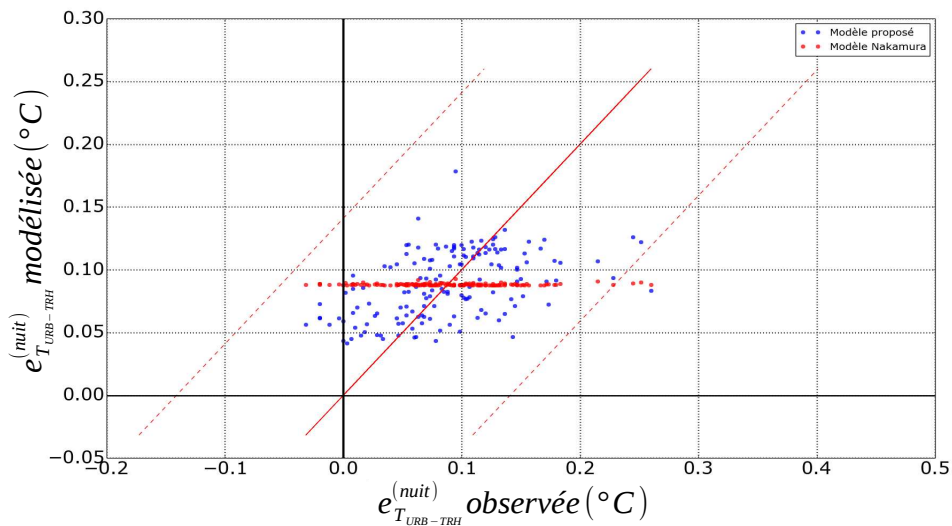
- de jour : 0,65 pour le modèle proposé, 0,48 pour les modèles Cheng et Nakamura ;
- de nuit : 0,22 pour le modèle proposé, nulle pour les modèles Nakamura et Cheng.

Le coefficient de détermination de notre modèle, même s'il reste le plus élevé des trois modèles, a nettement diminué (0,8 à 0,65 de jour et 0,54 à 0,22 de nuit). Au contraire, pour les



pointillés rouges : incertitude sur la différence de température associée à la précision des capteurs

(a)



pointillés rouges : incertitude sur la différence de température associée à la précision des capteurs

(b)

FIGURE 2.34 – Erreur (a) diurne et (b) nocturne estimée à partir des trois modèles étudiés en fonction de l'erreur observée. Les données météorologiques utilisées en entrée des modèles sont les données mesurées sur le site Météo-France

modèles diurnes, celui des deux autres modèles s'est stabilisé voire a nettement augmenté (0,09 à 0,48 pour le modèle Nakamura). Cependant, ces deux modèles n'expliquent pas du tout les variations de l'erreur enregistrée la nuit. De jour comme de nuit, en utilisant les données régionales ou les données locales, le modèle proposé explique donc mieux la variance des données que les deux autres.

L'analyse des résidus de l'erreur diffère très peu de l'analyse réalisée avec les données locales. Seuls les résidus nocturnes en fonction de la nébulosité sont intéressants. L'erreur ne suit

initialement pas une tendance linéaire en fonction de cette variable : elle est presque constante pour les valeurs de nébulosité allant de 0 à 6 octa puis elle diminue linéairement entre 6 et 8 octa. Aucun des modèles utilisé ne permet de distribuer normalement les résidus de l'erreur selon cette variable. Néanmoins, pour le modèle proposé, la valeur associée aux coefficients  $\omega_1$  et  $\omega_3$  (termes dépendants du rayonnement) est toujours significative la nuit (tableau 2.7). Le terme  $\omega_1$  devient dans ce cas un terme prépondérant (contre  $\omega_3$  pour les données locales). De jour, les trois coefficients ont des valeurs significatives et ces valeurs pour les variables centrées et réduites sont toutes trois du même ordre de grandeur. Elles ont donc une importance partagée dans l'explication des valeurs de l'erreur.

Coefficient	Données locales		Données régionales	
	URBIO / TRH (données nocturnes)	TRH / URBIO (données diurnes)	URBIO / TRH (données nocturnes)	TRH / URBIO (données diurnes)
$\omega_1$	0,000022 (0,024)	0,000014 (0,016)*	0,0061 (0,018)	0,00095 (0,11)
$\omega_2$	-0,0033 (0,0047)	0,062 (0,17)	-0,015 (0,0055)*	0,29 (0,16)
$\omega_3$	-0,00097 (-0,033)	0,00098 (0,41)	-0,0095 (0,0055)	0,00099 (0,16)
$\omega_4$	0,79 (0,092)	-0,71 (0,59)	0,11 (0,088)	0,097 (0,80)

\* coefficient pour lequel la statistique  $p_{val} > 0,05$  (significativité faible)

(...) les valeurs entre parenthèses correspondent aux valeurs pour les variables centrées et réduites

TABLE 2.7 – Valeurs des coefficients des équations 2.21 et 2.22 déterminées par régression linéaire après centrage et réduction des variables explicatives

Maintenant que les modèles sont calibrés avec des données régionales, il est intéressant de vérifier s'ils permettent une bonne estimation de l'erreur dans un autre contexte (avec un autre jeu de données).

#### 2.4.4 Vérification du modèle de correction

Les données utilisées pour vérifier le bon fonctionnement des modèles sont acquises à partir de capteurs placés sous quatre abris. Deux de ces abris sont des abris TRH, deux autres sont des abris URBIO. Pour chaque type d'abris, l'un est sale, l'autre est propre. L'objectif de l'expérience est ainsi double :

- vérifier les modèles calibrés précédemment ;



- voir l'influence de la saleté sur les performances de l'abri : si les différences de température mesurées entre deux abris d'un même type (l'un propre et l'autre sale) est supérieure à la différence de température mesurée entre deux abris qui diffèrent (tous les deux propres ou sales), les modèles de correction sont en effet inutiles.

Le site de mesure est toujours celui de la station 06 du réseau TRH et la disposition exacte de l'expérience est présentée sur la figure 2.35 (campagne 3 : vérification).

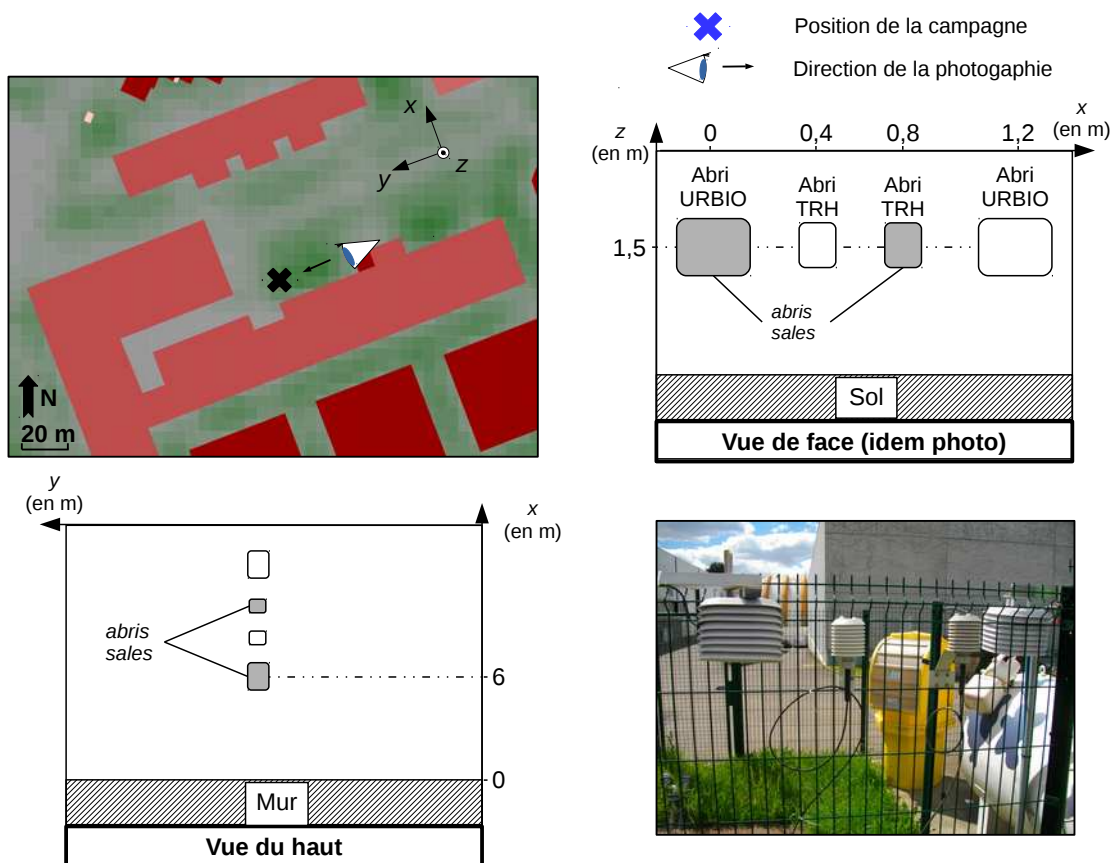


FIGURE 2.35 – Description du site de mesure et du positionnement du matériel utilisé pendant la campagne de mesure 3. Cette campagne est utilisée pour vérifier les modèles d'erreur établis préalablement

#### 2.4.4.1 Étalonnage des capteurs utilisés

Les capteurs sont comparés avant la campagne de mesure (qui dure six jours) afin de déterminer de potentielles erreurs pour chacun d'eux. Ils sont placés à l'air libre (sans abri) dans une pièce ventilée dans laquelle la température est constante à 17,5°C. Aucun capteur n'ayant été calibré par nos soins récemment, on considérera la valeur moyenne des quatre capteurs à un instant  $t$  comme la valeur vraie de la température. L'erreur  $e_{T_i}$  de chaque capteur  $i$  par rapport à cette

valeur peut être calculée par :

$$e_{T_i} = moy(\Delta T_{i-moy})_i \quad (2.25)$$

avec  $moy(\Delta T_{i-moy})_i$  la valeur moyenne de la différence entre la température mesurée par le capteur  $i$  et la température moyenne mesurée par les quatre capteurs.

Par la suite, les valeurs corrigées  $T_{cor_i}$  telles que  $T_{cor_i} = T_i - e_{T_i}$  seront directement utilisées.

#### 2.4.4.2 Évaluation des performances du modèle

Les données de vérification sont enregistrées pendant près d'une semaine au pas de temps de 15 minutes. Avant de vérifier les modèles, il s'agit de vérifier que l'influence de la saleté d'un abri sur l'erreur de mesure n'est pas supérieure à celle liée à l'utilisation de deux abris différents.

L'influence de la saleté est différente selon le type d'abri utilisé :

- URBIO : l'abri sale est celui qui chauffe le plus durant la journée (moyenne : 0,07 et écart type : 0,27 °C) ainsi que durant la nuit (moyenne : 0,1 et écart type : 0,18 °C) ;
- TRH : L'abri sale est l'abri le plus chaud la nuit (moyenne : 0,13 et écart type : 0,09 °C) mais celui qui chauffe le moins en journée (moyenne : -0,10 et écart type : 0,26 °C).

Ces résultats sont à prendre avec précaution car seul un abri propre et un abri sale de chaque type sont utilisés pour cette comparaison. D'autres facteurs tels que les différences minimales qui peuvent exister entre deux abris d'un même type ou la position du capteur dans l'abri pourraient également expliquer les différences observées. Cependant, en considérant l'équation 2.21, un sens physique peut être donné à ces observations : le terme contenant le rayonnement (indépendant de la vitesse du vent) peut être assimilé au rayonnement solaire qui entre par réflexion sur les parois de l'abri. Ce rayonnement réchauffe donc directement le capteur, ou bien seulement les parois intérieures de l'abri qui réchauffent à leur tour par convection l'air et donc le capteur. Si l'abri est propre, cette réflexion est plus marquée. L'abri TRH est celui pour lequel ce terme est prépondérant, c'est donc logiquement l'abri propre qui se réchauffe le plus. Pour l'abri URBIO, nous avons vu auparavant que ce terme est négligeable en journée. Ceci correspond aux nouvelles observations puisque l'abri URBIO sale chauffe plus en journée que l'abri URBIO propre. Le rayonnement absorbé par la surface de l'abri et transmis à l'air entrant dans le système est dans ce cas le mécanisme physique responsable du réchauffement du capteur. La saleté altère ainsi les caractéristiques radiatives de la surface blanche : selon le type d'abri utilisé,

cette altération peut être positive (abri TRH dont les performances sont néanmoins mauvaises dans tous les cas) ou négative (abri URBIO pour lequel les performances restent meilleures que celles de l'abri TRH dans tous les cas).

Pour conclure, même si des différences assimilables au degré de saleté sont observées entre les abris, elles sont faibles et négligeables devant l'ordre de grandeur des erreurs dues au type d'abri utilisé. Les erreurs de mesure propres à chacun des deux abris (sales et propres) seront donc estimées à partir des modèles établis auparavant puis seront comparées aux erreurs mesurées (écarts mesurés entre les types d'abris). La modélisation de l'erreur a été établie avec un abri propre. Les abris de référence pour l'observation de l'erreur seront donc les abris propres. La vérification des modèles d'erreur est ainsi réalisée la nuit pour les abris URBIO (sale et propre) par rapport à l'abri TRH propre et le jour pour les abris TRH (sale et propre) par rapport à l'abri URBIO propre.

Un indicateur (noté  $\widehat{RMSE}$ ) est proposé équation 2.26 pour juger de la pertinence d'utiliser des données corrigées grâce au modèle d'erreur proposé plutôt que les données brutes erronées.

$$\widehat{RMSE} = \frac{RMSE_{res}}{RMSE_{ini}} = \frac{\sqrt{\sum_t e_{T_{res}}(t)}}{\sqrt{\sum_t e_{T_{ini}}(t)}} \quad (2.26)$$

Cet indicateur est défini comme le rapport entre la racine carrée de deux erreurs quadratiques moyennes :

- celle de l'erreur résiduelle  $e_{T_{res}}(t)$  après correction, assimilable aux résidus du modèle (pour chaque instant  $t$ ) : elle sera notée  $RMSE_{res}$  ;
- celle de l'erreur initiale  $e_{T_{ini}}(t)$  dans le cas où aucune correction n'est réalisée : elle sera notée  $RMSE_{ini}$ .

Dans le cas où  $\widehat{RMSE}$  est supérieur à un, les données brutes sont préférables aux données corrigées. Dans le cas contraire, les données corrigées sont préférables : l'utilisation du modèle d'erreur est donc pertinente. C'est le cas de la figure 2.36 où l'indicateur  $\widehat{RMSE}$  est illustré.

Au vu de ce raisonnement, l'utilisation de n'importe quel modèle d'erreur propre à l'abri TRH (le jour) pour corriger les données est pertinente puisque l'indicateur  $\widehat{RMSE}$  est toujours inférieur à 1 (tableau 2.8). Quel que soit le modèle proposé et le degré de saleté de l'abri, il est donc toujours préférable de corriger l'erreur plutôt que d'utiliser la donnée brute. Le modèle le plus performant est toujours le modèle proposé. Pour l'abri le plus sale, seules

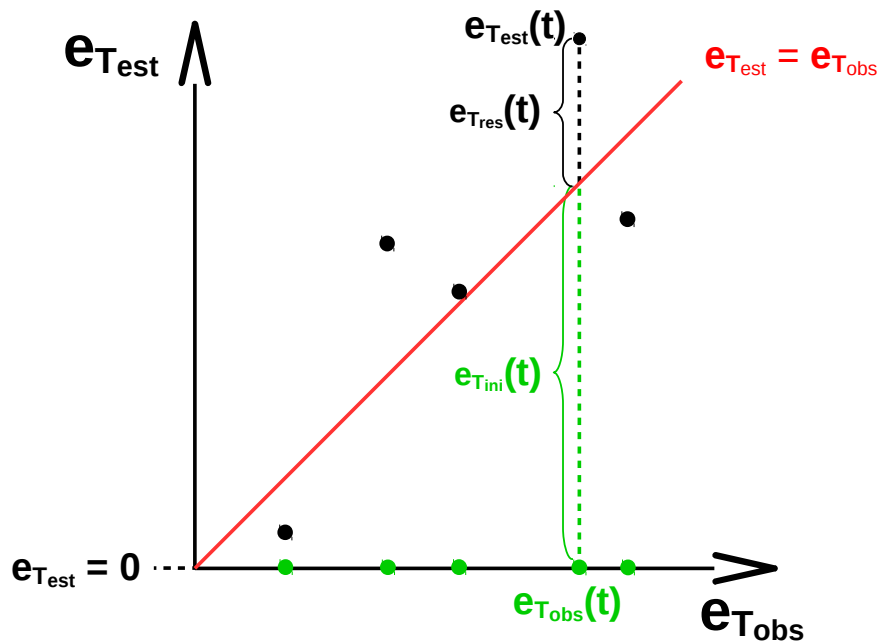


FIGURE 2.36 – Erreur résiduelle  $e_{T_{res}}$  après correction et erreur  $e_{T_{ini}}$  si aucune correction n'est effectuée. Ces erreurs sont utilisées pour juger de la pertinence d'un modèle de correction en calculant l'indicateur  $\widehat{RMSE}$  (défini comme le rapport de ces deux erreurs)

quelques valeurs hautes de biais sont mal modélisées ( $> 2^{\circ}\text{C}$ ), comme l'illustre la figure 2.37. Pour l'abri propre, dès que le biais dépasse  $1^{\circ}\text{C}$ , les valeurs sont sous-estimées par le modèle.

La nuit, aucun des modèles de correction proposés ne fonctionne (les valeurs de la  $\widehat{RMSE}$  sont toutes supérieures à 1). Il est dans ce cas plus sage de garder les données brutes. L'erreur de température faible enregistrée entre les abris est sans doute explicative de l'échec des modèles à la modéliser. Dans ces conditions, il est difficile d'allouer un modèle de correction nocturne à l'abri URBIO. La seule observation qui pourra être réutilisée par la suite est que la température mesurée la nuit par l'abri URBIO est souvent supérieure à celle mesurée par l'abri TRH.

Le tableau 2.8 décrit les performances propres à chaque modèle selon les données de vérification utilisées. Les ordres de grandeurs de la  $\widehat{RMSE}$  augmentent lorsque d'autres jeux de données sont utilisés pour la vérification (on parlera d'extrapolation du modèle). Cependant, sa valeur reste largement inférieure à 1, ce qui signifie que de jour, l'utilisation de modèles pour corriger les données est bien toujours pertinente. La vérification des modèles sur un seul site ne permet pas de conclure sur leur robustesse à l'extrapolation. Nous considérerons néanmoins par la suite qu'utiliser le modèle proposé pour la période diurne pour corriger les données est plus pertinent qu'utiliser l'un des deux autres modèles ou qu'utiliser les données brutes.

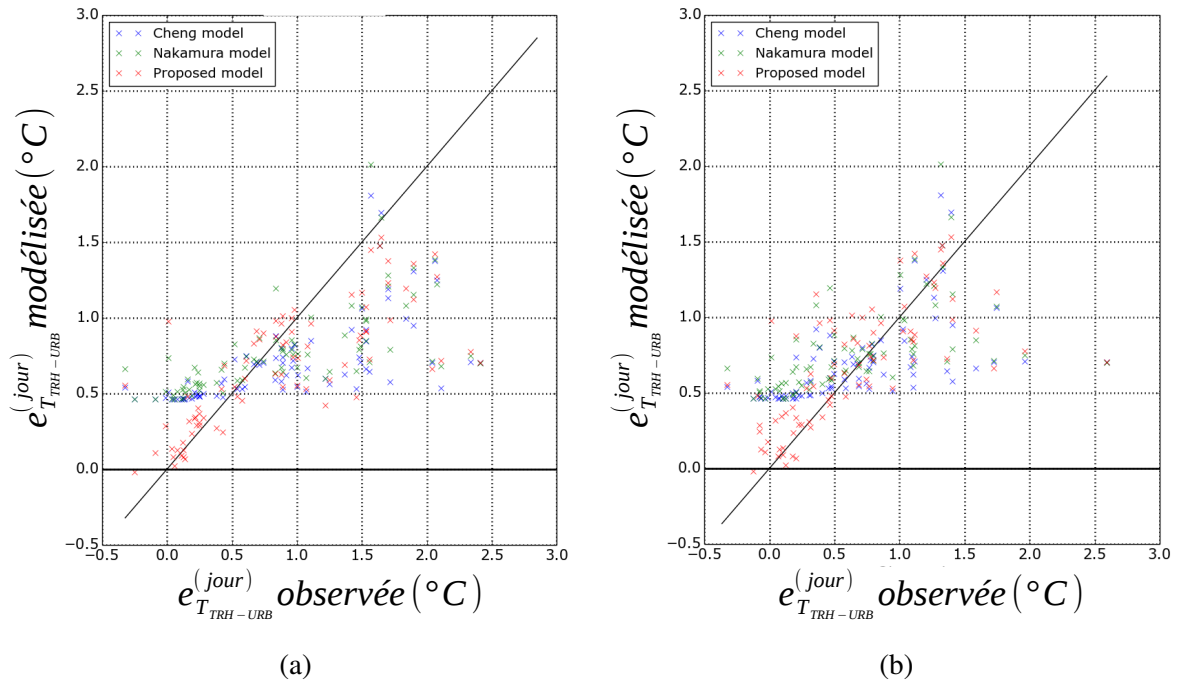


FIGURE 2.37 – Erreur diurne estimée à partir des trois modèles étudiés en fonction de l’erreur observée durant la campagne de vérification 3 : (a) pour les abris propres, (b) pour les abris sales. Les données météorologiques utilisées en entrée des modèles sont les données mesurées sur le site Météo-France.

Période	Modèle	Phase de calibration		Phase de vérification	
		Données locales	Données globales	Données globales	
				Abri propre	Abri sale
Diurne	Nakamura	0,52 (0,62)	0,31 (0,31)	0,54 (0,51)	0,47 (0,54)
	Cheng	0,4 (0,48)	0,31 (0,31)	0,57 (0,53)	0,46 (0,53)
	proposé	0,24 (0,29)	0,26 (0,26)	0,47 (0,47)	0,41 (0,47)
Nocturne	Nakamura	0,053 (0,49)	0,056 (0,54)	0,25 (1,15)	0,28 (1,22)
	Cheng	0,058 (0,54)	0,056 (0,54)	0,25 (1,16)	0,28 (1,22)
	proposé	0,04 (0,37)	0,049 (0,47)	0,25 (1,16)	0,27 (1,19)

(...) les valeurs entre parenthèses correspondent aux valeurs normalisées  $\widehat{RMSE}$

TABLE 2.8 –  $RMSE_{res}$  et  $\widehat{RMSE}$  de l’erreur estimée à partir de chaque modèle : Nakamura, Cheng et celui proposé dans cette thèse

## 2.4.5 Application du modèle : correction des jeux de données

L’objectif initial de cette section 2.4 est de corriger les données des réseaux de capteurs de température qui présentent des erreurs de mesure en utilisant les modèles décrits précédemment.

L'abri TRH est à l'origine de ces erreurs lorsqu'il est soumis au rayonnement solaire. Cet abri est utilisé par les réseaux TRH et FluxSAP. Tous les sites de ces deux réseaux seront donc corrigés.

#### **2.4.5.1 Périodes de correction : rayonnement solaire direct sur l'abri**

Le rayonnement solaire descendant utilisé pour construire les variables explicatives de notre modèle de correction est mesuré dans un espace ouvert dépourvu de masques solaires (site MF Bouguenais). Ce rayonnement est utilisé pour le calcul d'erreur de tous les sites de mesure qui nécessitent une correction. Nos abris, placés en milieu urbain, ne voient en réalité pas tous le même rayonnement du fait des nuages, des particules de pollution et des masques solaires qui peuvent différer d'un site à l'autre.

- Les nuages ne sont pas forcément en même temps présents au-dessus de tous nos sites ni du site où est mesuré le rayonnement. Cette situation est cependant supposée peu fréquente du fait de la distance faible qui sépare les sites (inférieure à 15 kilomètres).
- Les particules de pollution présentes au-dessus de la canopée peuvent réduire légèrement le rayonnement solaire descendant. Les variations spatiales du rayonnement dues à ce phénomène sont également considérées comme négligeables par rapport à la valeur totale du rayonnement.
- Les masques solaires ne sont pas identiques d'un site à l'autre. Ceci implique que seuls les rayonnements diffus et réfléchis touchent un abri à l'ombre tandis qu'un autre site situé au soleil sera touché également par le rayonnement direct. Or, d'après les campagnes de modélisation et de vérification de l'erreur, l'erreur lorsque l'abri était à l'ombre était négligeable devant l'erreur lorsque l'abri était au soleil.

Seules les périodes où un abri sera soumis au rayonnement solaire direct seront donc considérées comme périodes à corriger.

#### **2.4.5.2 Méthode d'identification des périodes de correction**

Pour identifier les périodes pour lesquelles chaque station voit le soleil, une méthode de lancer de rayon sous le logiciel SketchUp est utilisée. Seuls les masques solaires attribuables aux bâtiments sont considérés, ceux créés par les arbres étant trop difficiles à déterminer (ils dépendent de la densité de feuillage, de la vitesse et de la direction du vent, variables au cours du temps). La procédure suivante est appliquée à chacune des stations.

1. Les coordonnées (empreinte au sol et hauteur) des bâtiments alentours à la station de mesure sont extraites de la Base de Données Topographiques (BD Topo version 2011) de l'Institut Géographique National (IGN). Les coordonnées approximatives  $x$ ,  $y$  et  $z$  du capteur sont également déterminées à partir d'un croisement entre les images de Google earth, les coordonnées des bâtiments de la BD Topo et les métadonnées propres à chaque station. Une reconstitution en trois dimensions de la scène (bâtiments et capteur) est ensuite effectuée sous SketchUp pour chaque site. Ce travail a en grande partie été réalisé par Timothée Holla, stagiaire à l'IRSTV en 2015.
2. Un plug-in développé au sein du laboratoire CRENAU est ensuite utilisé pour :
  - (a) déterminer toutes les positions prises par le soleil dans ce ciel au cours d'une année (discrétisation temporelle toutes les cinq minutes) ;
  - (b) projeter un rayon entre le capteur et chacune des positions prises par le soleil dans le ciel. Si le rayon traverse un bâtiment, la valeur 0 est associé à cette position et donc à ce pas de temps (l'abri ne voit pas le soleil). Dans le cas contraire, c'est la valeur un qui est associée (l'abri voit le soleil).

### 2.4.5.3 Confrontation des périodes d'erreur identifiées aux périodes d'erreur observables

La pertinence des périodes d'ensoleillement de l'abri déterminées sous SketchUp est vérifiée par une méthode dite *des dérivées secondes* (vérification illustrée pour les sites TRH11 et TRH06, respectivement figures 2.38 et 2.41). La procédure est décomposée en quatre étapes.

1. Les données de température de chaque station sont utilisées pour construire des journées moyennes correspondant à chaque mois de l'année.
2. Les données booléennes (0 ou 1), produites pour déceler si l'abri de chaque station voit le soleil, sont moyennées selon la même méthode (elles perdent donc leur caractère binaire). Les signaux, calculés pour chaque mois de l'année, sont notés  $bool(t)$ .
3. La dérivée seconde de la température de chacune des 12 journées mensuelles moyennes est calculée pour chaque station de mesure. Les signaux résultants, notés  $d^2T(t)$ , sont triés par mois de l'année.
4. Le comportement (les variations) des signaux  $bool(t)$  et  $d^2T(t)$  est comparé afin d'identifier si les périodes théoriques calculées sous SketchUp correspondent à des périodes de fortes perturbations de nos données de température.

Au final, pour toutes les stations, les périodes d'erreur identifiées sous SketchUp sont cohérentes avec celles identifiables par l'analyse des dérivées secondes. Les sites TRH11 et TRH06, dont la vérification des périodes de correction est respectivement illustrée figure 2.38 et figure 2.41, se distinguent par leur distance à des bâtiments. L'abri TRH11 est environ à 30 centimètres de la face nord-est d'un bâtiment (figure 2.39) tandis que l'abri TRH06 est à plus de quatre mètres de tout bâtiment (figure 2.40). L'abri TRH06 voit donc le soleil quasiment toute la journée pendant les mois estivaux tandis que l'abri TRH11 ne le voit que durant les premières heures du jour. Ceci implique pour le capteur TRH11 une modification assez brutale de la température entre l'instant où l'abri est à l'ombre et celui où il est au soleil, qui est moins perceptible pour le capteur TRH06. La dérivée seconde de la température augmente puis diminue donc très fortement à ces instants pour le TRH11, ce qui permettrait d'identifier les instants de début et de fin d'erreur sans l'analyse faite sous SketchUp. Ce n'est par contre pas le cas pour l'abri TRH06 où les pics de début et de fin d'erreur sont moins visibles. L'utilisation de la méthode de lancer de rayon est dans ce cas primordiale à l'identification de l'erreur.



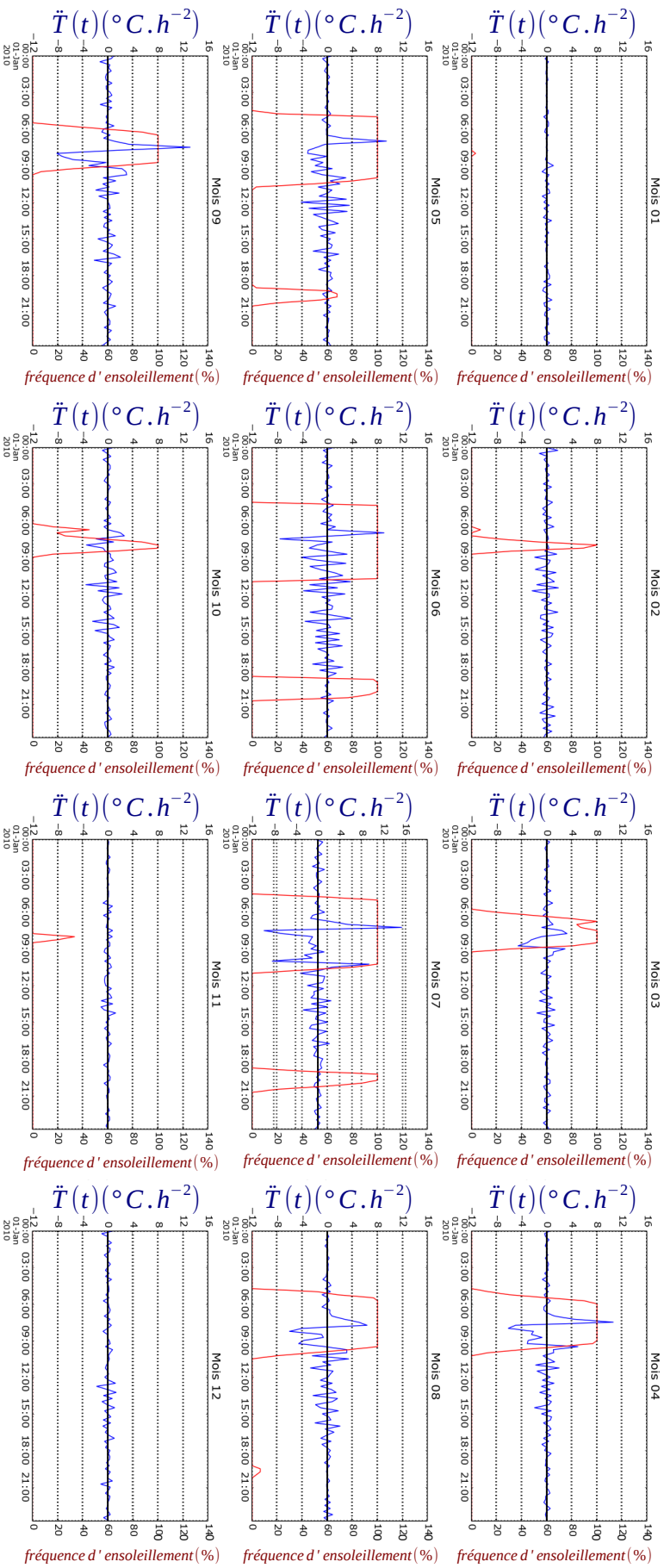


FIGURE 2.38 – Confrontation des périodes d'ensoleillement du capteur TRH11 avec celles de forte variation du taux de réchauffement de l'air

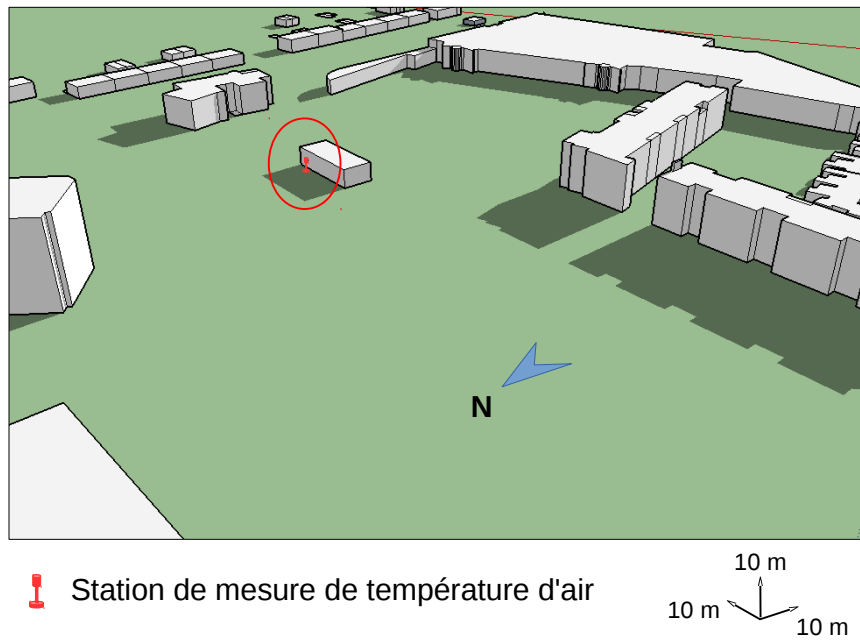


FIGURE 2.39 – Maquette Sketchup du site TRH11 avec ombres portées à 12h (heure solaire)

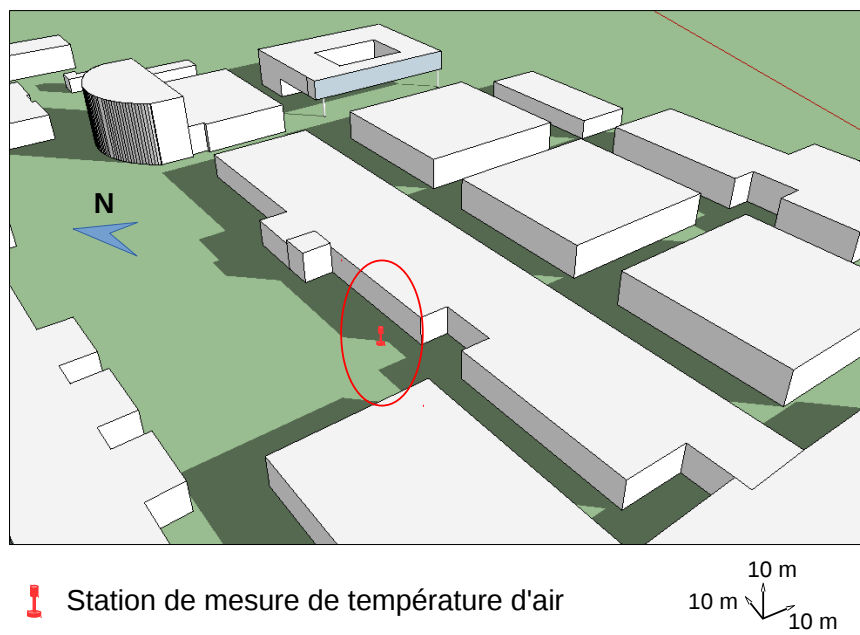


FIGURE 2.40 – Maquette Sketchup du site TRH06 avec ombres portées à 12h (heure solaire)

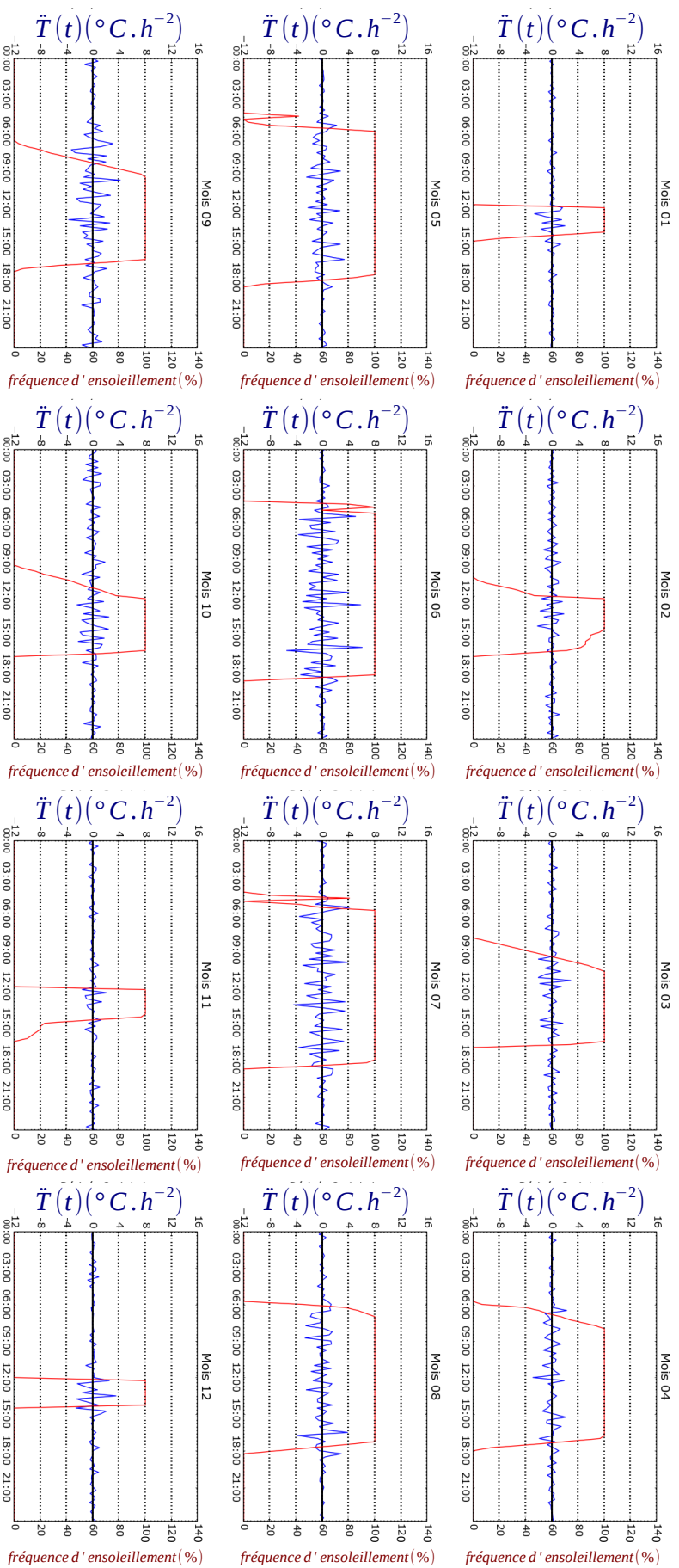


FIGURE 2.41 – Confrontation des périodes d'ensoleillement du capteur TRH06 avec celles de forte variation du taux de réchauffement de l'air

### 2.4.5.4 Correction des jeux de données

Le modèle d'erreur formulé équation 2.21 est maintenant utilisé pour corriger les données de température pendant les périodes identifiées sous SketchUp. Les variables explicatives sont calculées à partir des données météorologiques mesurées à la station Météo-France Bouguenais. La température de chacune des stations des réseaux de mesure de température TRH et FluxSAP est ainsi modifiée selon l'équation 2.27. Celle-ci provient de l'équation 2.21, calibrée avec les données de Météo-France, avec  $K \downarrow$  remplaçant  $Q_{\downarrow+\uparrow}$  ainsi que les coefficients du tableau 2.7.

$$T_{cor_i}(t_{cor_i}) = T_{ini_i}(t_{cor_i}) - 0,000953 \frac{K \downarrow}{U} + 0,294 \frac{\dot{T}}{U} - 0,000989 K \downarrow - 0,969 \quad (2.27)$$

avec  $T_{cor_i}$  la température de la station  $i$  après correction,  $t_{cor_i}$  chaque instant  $t$  pour lequel la station  $i$  voit le soleil (et par conséquent nécessite d'être corrigée) et  $T_{ini_i}$  la température de la station  $i$  avant correction

Finalement, afin de rendre compte de l'ampleur des corrections apportées, trois indicateurs sont calculés pour chaque capteur et chaque saison de l'année :

- le taux de correction, défini comme le rapport entre le nombre d'échantillons pour lesquels une correction a été appliquée et le nombre d'échantillons total ;
- la racine carrée de l'écart quadratique moyen (noté  $RMSE$  et moyenné pour tous les pas de temps) entre les données brutes et les données corrigées ;
- le même indicateur que le précédent mais limité seulement aux périodes pendant lesquelles des corrections ont été appliquées ( $RMSE_{t=t_{cor_i}}$ ).

Le résultat de ces calculs est synthétisé dans le tableau 2.9. Il apparaît que seul l'abri TRH04 ne voit jamais le soleil et ne subit donc aucune correction. Tous les autres abris sauf le TRH11 ont plus de 10% et moins de 62 % de leurs valeurs corrigées quelle que soit la saison. Les  $RMSE_{t=t_{cor_i}}$  varient de 0,16 à 1,1°C selon la saison et l'abri concerné. Les valeurs maximales sont surtout associées aux capteurs du réseau FluxSAP, qui sont souvent à l'écart des façades nord et qui voient donc le soleil lorsque le rayonnement est maximal. Le printemps et l'été sont les saisons où ces corrections sont les plus fréquentes et les plus élevées, ce qui n'est pas surprenant compte tenu de la durée du jour plus longue et des valeurs de rayonnement plus élevées à cette période. Au final, les  $RMSE$  globales sont toutes inférieures à 1°C mais atteignent tout de même 0,80°C l'été pour la plupart des stations du réseau FluxSAP.

Ces statistiques démontrent que les corrections appliquées sont rarement négligeables en regard des valeurs d'ICU observées habituellement pour des villes de la taille de Nantes (qui sont du même ordre de grandeur). Dans la suite de ce travail, les données des réseaux de température TRH et FluxSAP utilisées seront les valeurs corrigées.

Saison	Indicateur	Réseau TRH								Réseau FluxSAP				
		02	04	06	10	11	17	18	19	23	24	25	jardin (jardin)	jardin (rue)
Automne	Taux de correction (%)	27	0	21	39	5,0	43	21	37	-	-	-	-	-
	$RMSE_{t=t_{cor_i}}$ (°C)	0,79	0	0,71	0,70	1,0	0,67	0,80	0,65	-	-	-	-	-
	$RMSE$ (°C)	0,42	0	0,32	0,44	0,24	0,44	0,36	0,39	-	-	-	-	-
Printemps	Taux de correction (%)	44	0	43	53	18	54	28	46	62	56	59	58	55
	$RMSE_{t=t_{cor_i}}$ (°C)	0,81	0	0,80	0,75	0,71	0,74	0,76	0,75	0,63	0,66	0,65	0,66	0,67
	$RMSE$ (°C)	0,53	0	0,52	0,55	0,30	0,55	0,41	0,51	0,50	0,50	0,50	0,50	0,50
Été	Taux de correction (%)	52	0	53	59	26	59	36	52	61	56	58	56	53
	$RMSE_{t=t_{cor_i}}$ (°C)	0,90	0	0,89	0,86	0,79	0,86	0,83	0,86	1,0	1,1	1,0	1,1	1,1
	$RMSE$ (°C)	0,65	0	0,65	0,66	0,41	0,66	0,49	0,62	0,80	0,80	0,76	0,80	0,80
Hiver	Taux de correction (%)	16	0	11	28	0,39	38	17	31	-	-	-	-	-
	$RMSE_{t=t_{cor_i}}$ (°C)	0,61	0	0,43	0,55	0,16	0,49	0,61	0,53	-	-	-	-	-
	$RMSE$ (°C)	0,24	0	0,14	0,29	0,010	0,30	0,25	0,29	-	-	-	-	-

TABLE 2.9 – Fréquence et valeurs moyennes saisonnières des corrections apportées aux valeurs de température mesurées par chaque station des réseaux TRH et FluxSAP

### 2.4.5.5 Évaluation de la méthode de correction

La figure 2.42 présente l'évolution temporelle de la température de l'air pour le site TRH11 (pour quelques jours printaniers ensoleillés).

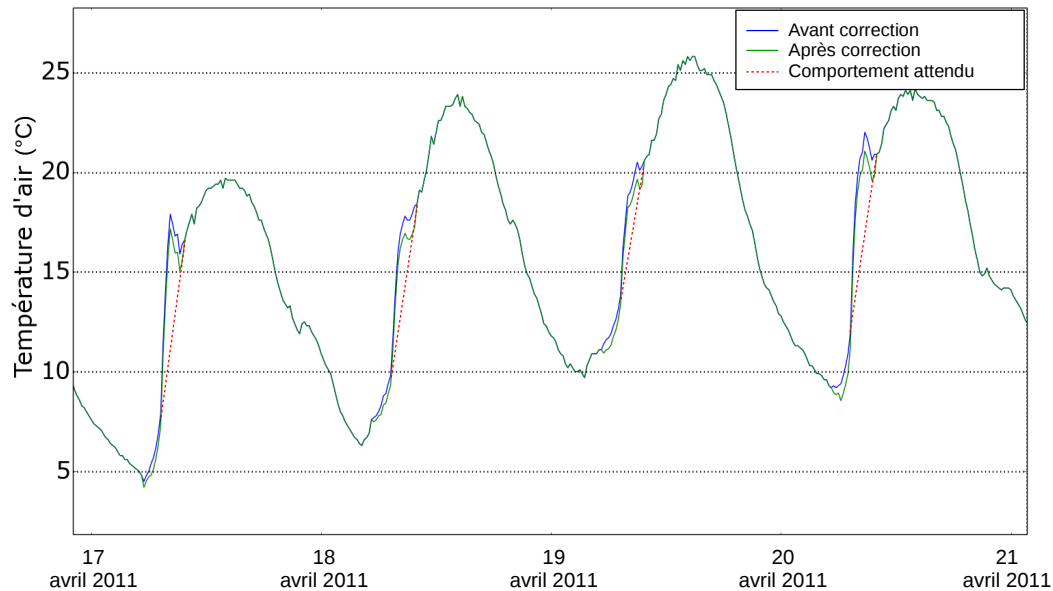


FIGURE 2.42 – Performances de la méthode de correction proposée - Application à la station 11 du réseau TRH

Trois types de données sont tracées : les données brutes (la température effectivement mesurée), les données corrigées (d'après l'équation 2.27), et les données représentant la température à laquelle on aurait pu s'attendre dans le cas d'une évolution linéaire de la température pendant la période identifiée comme erronée. Les corrections sont insuffisantes au vu de l'évolution attendue. Ceci peut s'expliquer par le fait de ne pas considérer l'influence que peut avoir la proximité du mur sur les conditions de rayonnement et de ventilation de l'abri (hypothèse 3 émise en section 2.4.1). Les modèles sont calibrés pour un abri placé de manière particulière dans un site particulier lui aussi. La vitesse du vent et le rayonnement subis par cet abri sont donc propres à l'orientation et à la structure des rues du site mais également à la distance entre l'abri et le mobilier urbain. Ces caractéristiques peuvent impacter le régime de vent et le rayonnement solaire qui s'exercent sur l'abri. Or, la station TRH11 est beaucoup plus proche du mur que l'abri qui a permis la calibration des modèles (TRH06). La vitesse du vent exercée sur cet abri est donc sans doute beaucoup plus faible que celle qui peut s'exercer au même moment sur l'abri qui a permis la calibration de nos modèles. Ceci peut augmenter considérablement l'erreur de mesure de la station TRH11.

Les modèles de correction proposés ne sont donc pas parfaits, ou non adaptés à tous

les sites. Afin de corriger les données de manière plus performante, plusieurs améliorations pourraient être apportées :

- Utiliser le modèle complet (équation 2.15 qui tient compte des constantes de temps) plutôt que sa version simplifiée.
- La nuit comme le jour, le rayonnement GLO n'est pas calculé. Le jour, il n'est pas considéré du tout tandis que la nuit, il est substitué par la nébulosité. Il serait pertinent de calculer ce rayonnement à partir d'une formule empirique telle que proposée par Matzarakis et al. (2010).
- Formuler deux modèles de décroissance de la vitesse du vent :
  - un premier dépendant de la densité frontale de bâtiment,
  - un second dépendant de la distance avec un obstacle majeur (tel un mur).
- Étudier plus en détail l'influence de la propreté de l'abri sur les mesures (en comparant les résultats d'une campagne d'intercomparaison d'abris utilisant plus d'abris sales et plus d'abris propres pour un même modèle).
- Vérifier les modèles finaux à partir de données recueillies sur plusieurs sites de mesure (seul le site 06 utilisé lors de la campagne 3 - illustré figure 2.35 - est utilisé ici).

## 2.5 Conclusion du chapitre

Ce chapitre a permis de présenter les quatre réseaux de capteurs et les six jeux de données qui seront utilisés par la suite. Certaines des données enregistrées sont entachées d'erreur pouvant atteindre jusqu'à 5°C. La différence de température pouvant se développer entre deux points d'une même rue n'explique que partiellement pourquoi la température mesurée sous un abri proche d'un mur est anormalement élevée. L'explication réside en fait dans le fonctionnement de l'abri ventilé protégeant le capteur de température du rayonnement. Les abris TRH, utilisés dans les jeux de données de calibration et certains jeux de données de vérification, surchauffent dès lors qu'ils voient directement le soleil. Les abris URBIO, utilisés dans presque tous les jeux de données de vérification, sont peu sujets à cet échauffement. La température mesurée sous ces derniers abris est considérée comme la température réelle de l'air. Afin que les données obtenues à partir des abris TRH soient comparables à celles obtenues à partir des abris URBIO,



elles sont intercorrignées à partir des conditions météorologiques mesurées à la station Météo-France la plus proche. Les modèles, issus d'un bilan d'énergie simplifié réalisé sur le système {abri + capteur}, dépendent de la vitesse du vent, du rayonnement solaire descendant et de la température de l'air. Les données de chaque station utilisant un abri TRH sont corrigées à partir de ces modèles dès lors que l'abri est au soleil. Même s'il semble que ces modèles devraient être affiner pour intégrer les caractéristiques morphologiques propre à chaque site de mesure, ces données corrigées seront directement utilisées par la suite.



# 3

## Îlot de chaleur urbain : calibration des modèles

Ce chapitre est consacré à la modélisation des variations spatiales et temporelles de l'ICU. Pour rappel, l'objectif de cette thèse est d'estimer la température  $T_i(t)$  de n'importe quel point  $i$  du territoire à n'importe quel instant  $t$ , à partir de deux types de données :

- les conditions météorologiques mesurées à une station de référence  $ref$  ;
- les informations géographiques caractérisant le site  $i$ .

La méthode générale utilisée pour résoudre ce problème spatio-temporel est décrite dans la première section. Les deux sections suivantes traitent respectivement de la modélisation temporelle et de la modélisation spatiale. Seul le jeu de données  $JDD_{calib}$ , décrit section 2.3.1 sera utilisé pour établir ces modèles.

## 3.1 Méthodologie générale de résolution du problème spatial et temporel

### 3.1.1 Îlot de chaleur urbain : calcul aux périodes clés

Nous avons vu au chapitre 1 que la température de l'air variait dans l'espace et dans le temps. Certaines périodes de la journée sont plus propices que d'autres à l'établissement de microclimats. Elles seront identifiées dans cette section.

#### 3.1.1.1 Conditions préalables à l'identification d'une période clé

L'ICU est généralement plus intense lors de journées claires (faible nébulosité) et calmes (vitesse de vent faible). Cependant, notre objectif est de modéliser l'ICU quelle que soit son intensité (i.e. quelles que soient les conditions météorologiques sous lesquelles il se développe). Toutes les journées qui constituent nos données sont donc utilisées pour identifier les périodes d'intérêt. Celles-ci sont triées saisonnièrement puis moyennées afin d'obtenir une journée type de la température de l'air pour chaque saison. Pour une saison donnée, l'opération est donc la suivante :

$$T_{i,k}(j_{moy}) = moy(T_{i,k}(j))_j \quad (3.1)$$

avec  $T_{i,k}(j_{moy})$  la température de l'air à la station  $i$  et pour le pas de temps  $k$ , moyennée entre toutes les journées  $j$  de la saison de façon à avoir une journée moyenne

Le taux de réchauffement  $\dot{T}_k$  est calculé à partir des journées moyennes de chaque saison. Il est défini comme la différence de température entre deux pas de temps successifs  $k - 1$  et  $k$  et divisée par le pas de temps  $\Delta t$ .

$$\dot{T}_{i,k}(j_{moy}) = \frac{T_{i,k}(j_{moy}) - T_{i,k-1}(j_{moy})}{\Delta t} \quad (3.2)$$

Les périodes d'intérêt sont identifiées à partir des journées types de chaque saison et caractérisées par deux critères.

- On privilégie une dispersion maximale de la variable  $T_{i,k}(j_{moy})$  selon la variable  $i$ , représentative des stations de mesure. Ceci implique des différences spatiales maximales de la température.

- On souhaite avoir une dispersion minimale de la variable  $\dot{T}_{i,k}(j_{moy})$  : ceci implique que les stations ont sensiblement le même taux de réchauffement, c'est-à-dire que les écarts moyens entre les stations sont stables. Moyenner les différences de température pendant cette période engendre donc peu de pertes d'informations.

Deux indicateurs de dispersion sont calculés pour chacune de ces variables à chaque pas de temps des journées moyennées :

- l'écart-type  $std()_i$  entre les stations  $i$  ;
- l'intercentile  $\int_{q_1=0.25}^{q_2=0.75} ()_i$  entre les stations : ceci correspond à la différence entre les valeurs minimales et maximales de température après avoir ôté les 25% des stations les plus chaudes et les 25% des stations les plus froides.

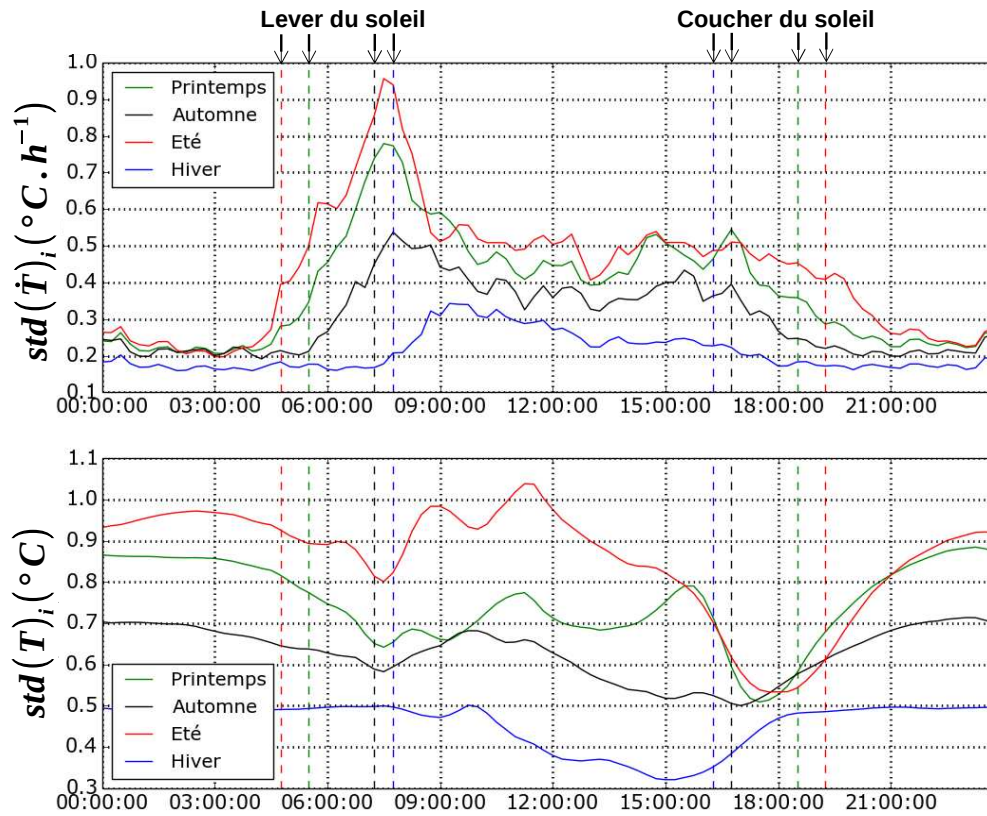
Soit  $x_{i,k} = T_{i,k}(j_{moy})$  ou  $x_{i,k} = \dot{T}_{i,k}(j_{moy})$ . Le premier indicateur ( $std(x_{i,k})_i$ ) peut être écrit sous la forme :

$$std(x_{i,k})_i = \frac{1}{nb_i \sqrt{2}} \sqrt{\sum_{i_1}^{nb_i} \sum_{i_2}^{nb_i} (x_{i_1,k} - x_{i_2,k})^2} \quad (3.3)$$

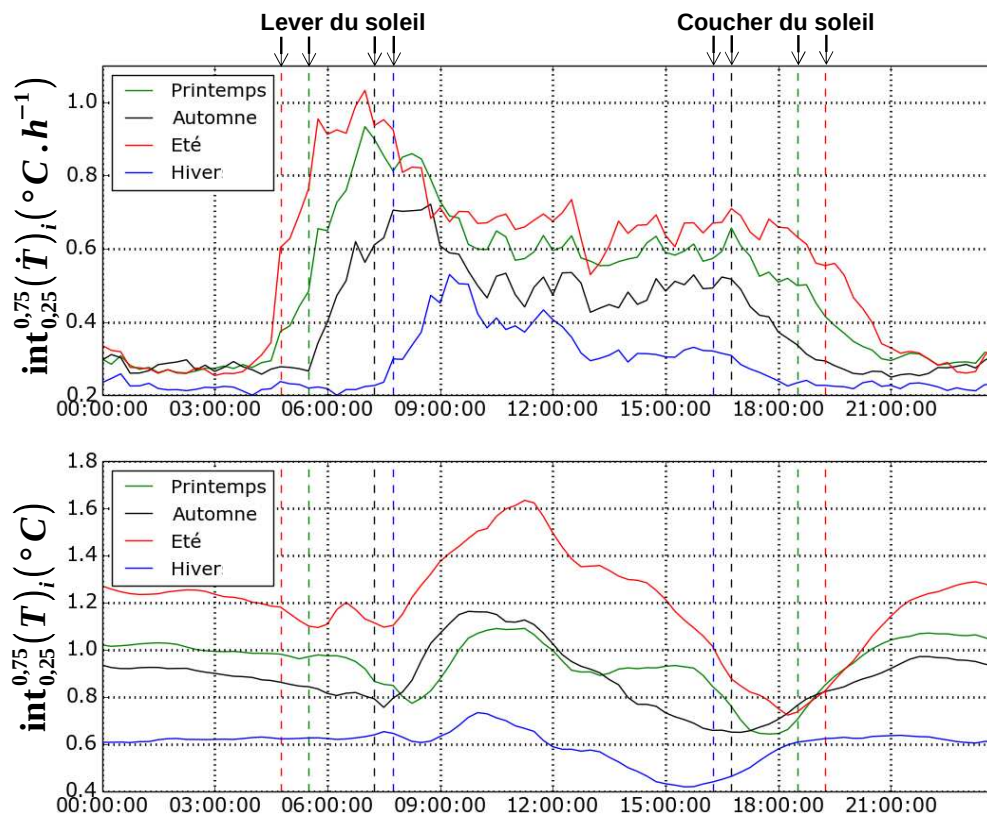
avec  $nb_i$  le nombre de stations,  $std(x_{i,k})_i$  l'écart-type entre les stations de mesure pour l'indicateur  $x$  pour chaque intervalle de temps  $k$ .

D'après cette relation,  $std(x_{i,k})_i$  est donc représentatif de la moyenne des écarts entre toutes les stations  $i$ . Cependant, cet indicateur est très sensible aux valeurs qui s'éloignent de la tendance, notamment lorsque le nombre d'échantillons utilisé  $nb_i$  est faible (ce qui est ici notre cas puisque seulement 10 stations sont utilisées). Nous avons vu au chapitre 2 que les stations utilisées pour la calibration mesuraient des valeurs de température aberrantes lorsqu'elles étaient soumises au rayonnement solaire direct. Ces valeurs ont été corrigées mais les modèles de correction utilisés ne sont pas parfaits. Ces données sont donc toujours erronées. De plus, le nombre de stations utilisé est faible (10). Afin de s'assurer que les valeurs calculées avec l'écart-type sont fiables, une double-vérification est réalisée à partir de  $\int_{q_1=0.25}^{q_2=0.75} ()_i$ , indicateur beaucoup plus robuste que  $std(x_{i,k})_i$  aux valeurs aberrantes engendrées par les erreurs de mesure résiduelles.

L'évolution de la dispersion des grandeurs  $T_{i,k}$  et  $\dot{T}_{i,k}$  est tracée au cours d'une journée moyenne pour chacun des deux indicateurs  $std()_i$  et  $\int_{q_1=0.25}^{q_2=0.75} ()_i$  (respectivement figure 3.1a et figure 3.1b).



(a)



(b)

FIGURE 3.1 – Évolution de la dispersion interstation de la température et de celle du taux de réchauffement au cours d’une journée moyenne pour le jeu de données de calibration  $JDD_{calib}$  : (a) selon l’indicateur *écart-type* et (b) selon l’indicateur *intercentile*.

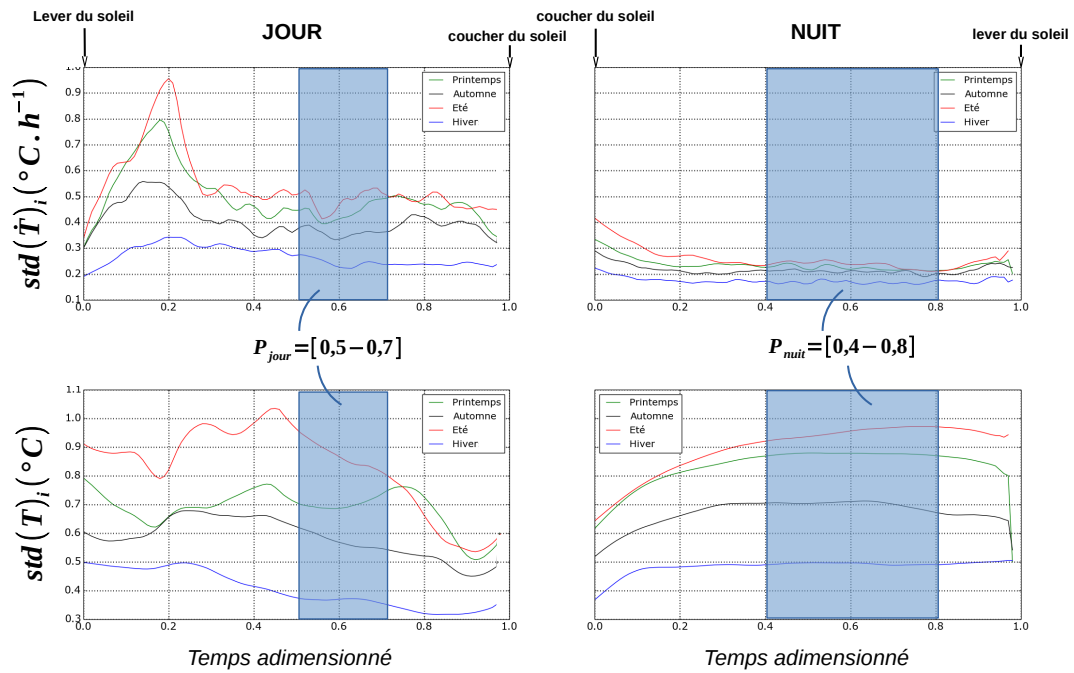
La valeur médiane du lever et du coucher du soleil est calculée pour chaque saison, faisant apparaître clairement la dépendance entre dispersion et cycle solaire journalier quel que soit l'indicateur utilisé. Les heures de lever et de coucher du soleil étant variables selon les saisons, un déphasage de la dispersion au cours des saisons a lieu. Afin de s'en affranchir, un adimensionnement des périodes diurnes et nocturnes par leur durée est donc réalisé. Cette méthode est inspirée de Runnalls et Oke (2000), qui constatent qu'en adimensionnant temporellement la période nocturne, l'ICU atteint généralement sa valeur maximale au même instant quelle que soit la saison.

### 3.1.1.2 Identification des périodes clés en journée adimensionnée

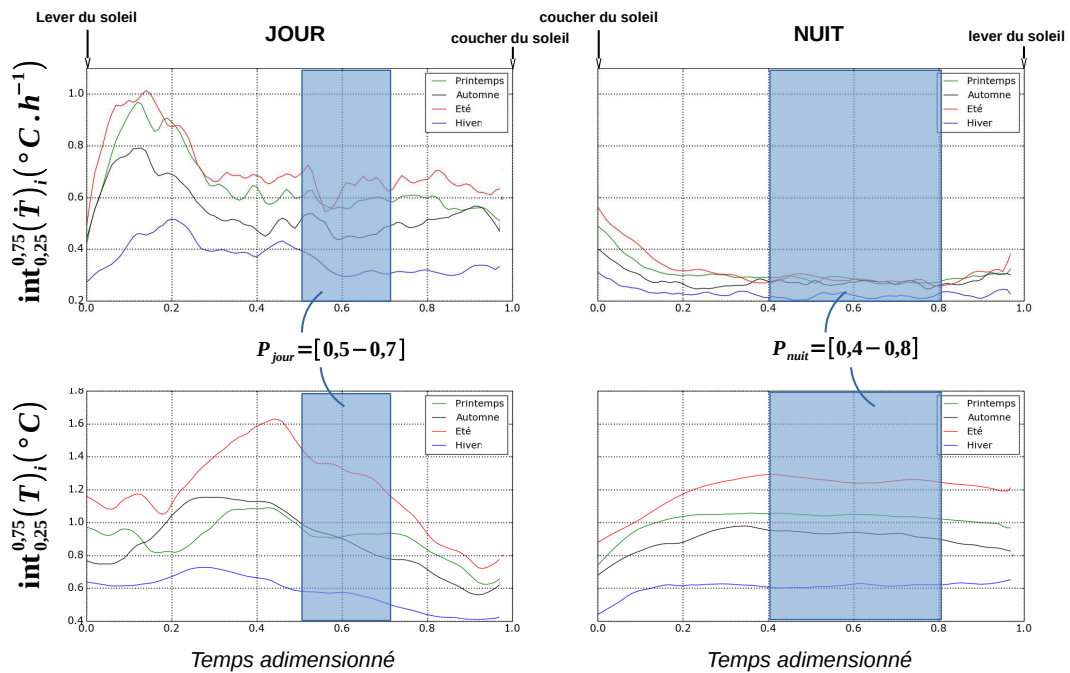
L'adimensionnement temporel consiste à distinguer les périodes diurnes et nocturnes et à considérer qu'elles ont chacune une durée unitaire. Ainsi, le lever du soleil correspond soit au 0 de la période diurne, soit à la valeur 1 de la période nocturne. L'adimensionnement permet d'obtenir une dispersion non déphasée quelle que soit la saison (figure 3.2).

En période diurne, sur l'intervalle  $[0 - 0, 2]^{(jour)}$ , les écarts de température se réduisent : les stations les plus fraîches (correspondant souvent aux sites les plus ouverts) convergent rapidement vers la même température que les stations les plus chaudes. En deuxième moitié de matinée, les écarts de température augmentent (intervalle  $[0, 2 - 0, 45]^{(jour)}$ ) et diminuent ensuite jusqu'au coucher du soleil (intervalle  $0, 45 - 1]^{(jour)}$ ). Lorsque les abris URBIO (jeu de données  $JDD_{spat\&temp}$ ) sont utilisés, l'augmentation des écarts observée en fin de matinée disparaît (figure 3.3). Cette augmentation est donc associée aux erreurs de mesure que subissent certaines stations à cette période de la journée. Au début de la période diurne, le taux de refroidissement est très différent d'une station à une autre. Cet écart se réduit jusqu'au milieu de la nuit (environ  $t^{(nuit)} = 0, 4 - 0, 5$ ) et est stable la deuxième moitié de la nuit. Les écarts de température subissent l'effet inverse : ils sont faibles en fin de journée, se creusent en première moitié de nuit puis se stabilisent en deuxième moitié.

Seule la période nocturne comprise dans l'intervalle  $[0, 4 - 0, 8]^{(nuit)}$  répond clairement aux critères énoncés section 3.1.1.1. Ce résultat est cohérent avec ceux obtenus par d'autres chercheurs, qui observent le maximum de l'ICU quelques heures après le coucher du soleil (Erell et Williamson 2007 ; Fenner et al. 2014 ; Svensson et al. 2002). Oke (1976) considère quant à lui une période plus précoce qui démarre 1h après le coucher du soleil et se termine 2h plus tard. Ces résultats sont difficilement comparables aux nôtres, puisque leurs données n'ont pas été adimensionnées temporellement. Pour ordre de grandeur, la période estivale don-



(a)



(b)

FIGURE 3.2 – Évolution (en temps adimensionné) de la dispersion interstation de la température et de celle du taux de réchauffement au cours d’une journée moyenne pour le jeu de données de calibration  $JDD_{\text{calib}}$  : (a) selon l’indicateur *écart-type* et (b) selon l’indicateur *intercentile*.

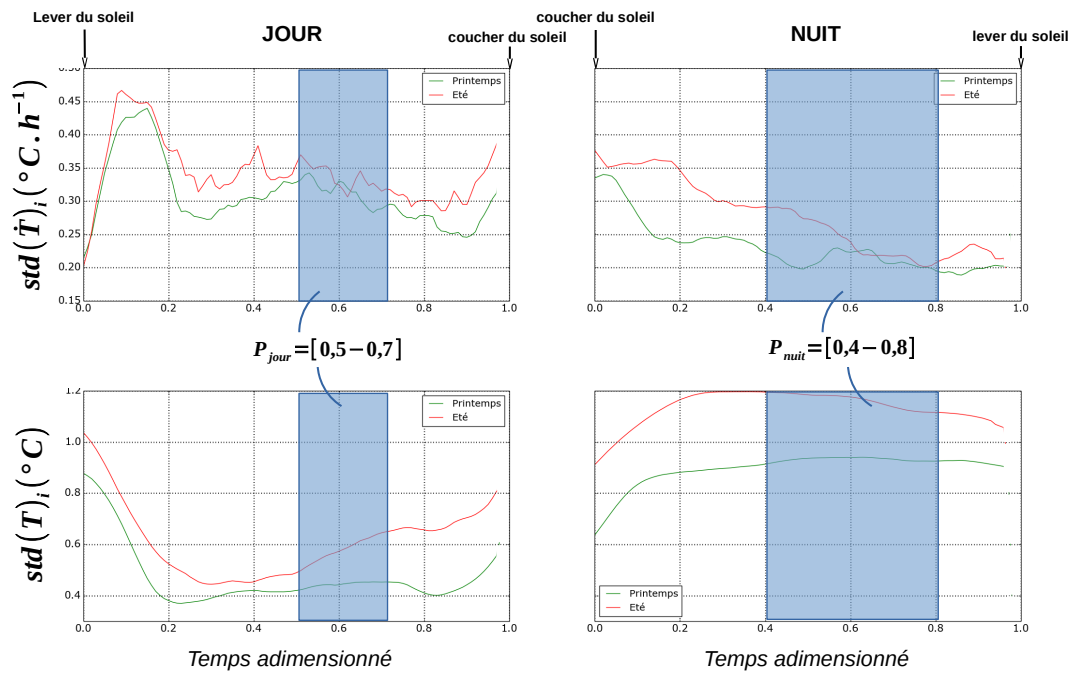


FIGURE 3.3 – Évolution (en temps adimensionné) de la dispersion interstation de la température et de celle du taux de réchauffement au cours d’une journée moyenne pour le jeu de données de vérification  $JDD_{spat\&temp}$  selon l’indicateur *écart-type*

née par Oke (1976) (1h à 3h après le coucher du soleil), correspond grossièrement en temps adimensionné à l’intervalle  $[0, 12 - 0, 25]^{(nuit)}$ . Aucune de ces études n’exprime leurs résultats en temps adimensionné, ce qui permettrait de comparer plus facilement les périodes entre elles, indépendamment de la saison étudiée ou de la latitude du lieu d’observation. Seuls Runnalls et Oke (2000) ont adimensionné temporellement les différences de température qu’ils ont observées la nuit. Ils ont ainsi montré que quelle que soit la saison, l’ICU atteint généralement sa valeur maximale à 50% de la nuit avant de se stabiliser jusqu’au lever du soleil. Ce résultat est cohérent avec la période  $[0, 4 - 0, 8]^{(nuit)}$  que nous avons identifiée.

En journée, il est plus complexe de trouver une période en adéquation avec nos critères. La figure 1.12 (chapitre 1) présente la procédure de calcul des indicateurs de température locale (*ITL*) les plus couramment utilisés dans la littérature. Les périodes diurnes sont beaucoup moins fréquemment utilisées que les périodes nocturnes. Certains auteurs s’y intéressent néanmoins : Chen et al. (2012) et Yan et al. (2014) observent des différences de température en début d’après-midi, même si le milieu d’après-midi (intervalle  $[0, 5 - 0, 7]^{(jour)}$  en temps adimensionné) reste la période la plus étudiée (Hart et Sailor 2009 ; Krüger et Givoni 2007 ; Petralli et al. 2014). Au regard de la figure 3.3 réalisée avec le réseau URBIO, cet intervalle est cohérent avec les critères énoncés section 3.1.1.1 : un maximum local de la dispersion de la température (même faiblement discernable) est observé au printemps. Pour les données TRH, si l’on se re-



porte à la figure 3.2a, le même maximum se trouve plutôt dans l'intervalle  $[0, 6 - 0, 8]^{(jour)}$ . Comme pour l'augmentation matinale, cette période est imputée aux abris TRH et non aux réelles différences de température d'air entre les sites.

Les deux périodes caractéristiques de l'ICU utilisées dans notre analyse seront donc :

- $P_{ICU}^{(jour)} = [0, 5 - 0, 7]^{(jour)}$  : intervalle défini pour la période diurne ;
- $P_{ICU}^{(nuit)} = [0, 4 - 0, 8]^{(nuit)}$  : intervalle défini pour la période nocturne.

### 3.1.1.3 Calcul d'indicateurs de température

Deux indicateurs de température locale ( $ITL^{(jour)}$  et  $ITL^{(nuit)}$ ) sont construits. Chacun d'eux décrit les conditions locales de température respectivement pour les périodes  $P_{ICU}^{(jour)}$  et  $P_{ICU}^{(nuit)}$  du cycle diurne. Ils sont notés génériquement  $ITL$  et sont construits sur le même principe que ceux décrits dans la littérature (figure 1.12). La figure 3.4 schématise la procédure suivie pour leur calcul :

1. la différence de température entre une station  $i$  et une station de référence est calculée à chaque pas de temps  $k$  de chaque cycle diurne  $j$  ;

$$\Delta T_{i-ref,k}(j) = T_{i,k}(j) - T_{ref,k}(j) \quad (3.4)$$

2. les plages journalières d'intérêt (correspondant à  $P_p = P_{ICU}^{(jour)}$  et  $P_p = P_{ICU}^{(nuit)}$ ) sont filtrées ;
3. les écarts de température pendant ces périodes sont moyennés pour former respectivement les deux indicateurs  $ITL^{(jour)}$  et  $ITL^{(nuit)}$ .

$$ITL^{(jour)} = moy(\Delta T_{i-ref,k \in P_{ICU}^{(jour)}}(j))_k \quad (3.5)$$

$$ITL^{(nuit)} = moy(\Delta T_{i-ref,k \in P_{ICU}^{(nuit)}}(j))_k \quad (3.6)$$

### 3.1.1.4 Estimation de la valeur des extremums journaliers

Les températures minimales et maximales journalières sont souvent enregistrées dans les intervalles  $[0 - 0, 1]^{(jour)}$  et  $[0, 65 - 0, 75]^{(jour)}$ . L'intervalle  $[0, 65 - 0, 75]^{(jour)}$  et  $P_{ICU}^{(jour)}$  s'intersectant, les écarts de température mesurés pendant  $P_{ICU}^{(jour)}$  peuvent donc être associés à ceux

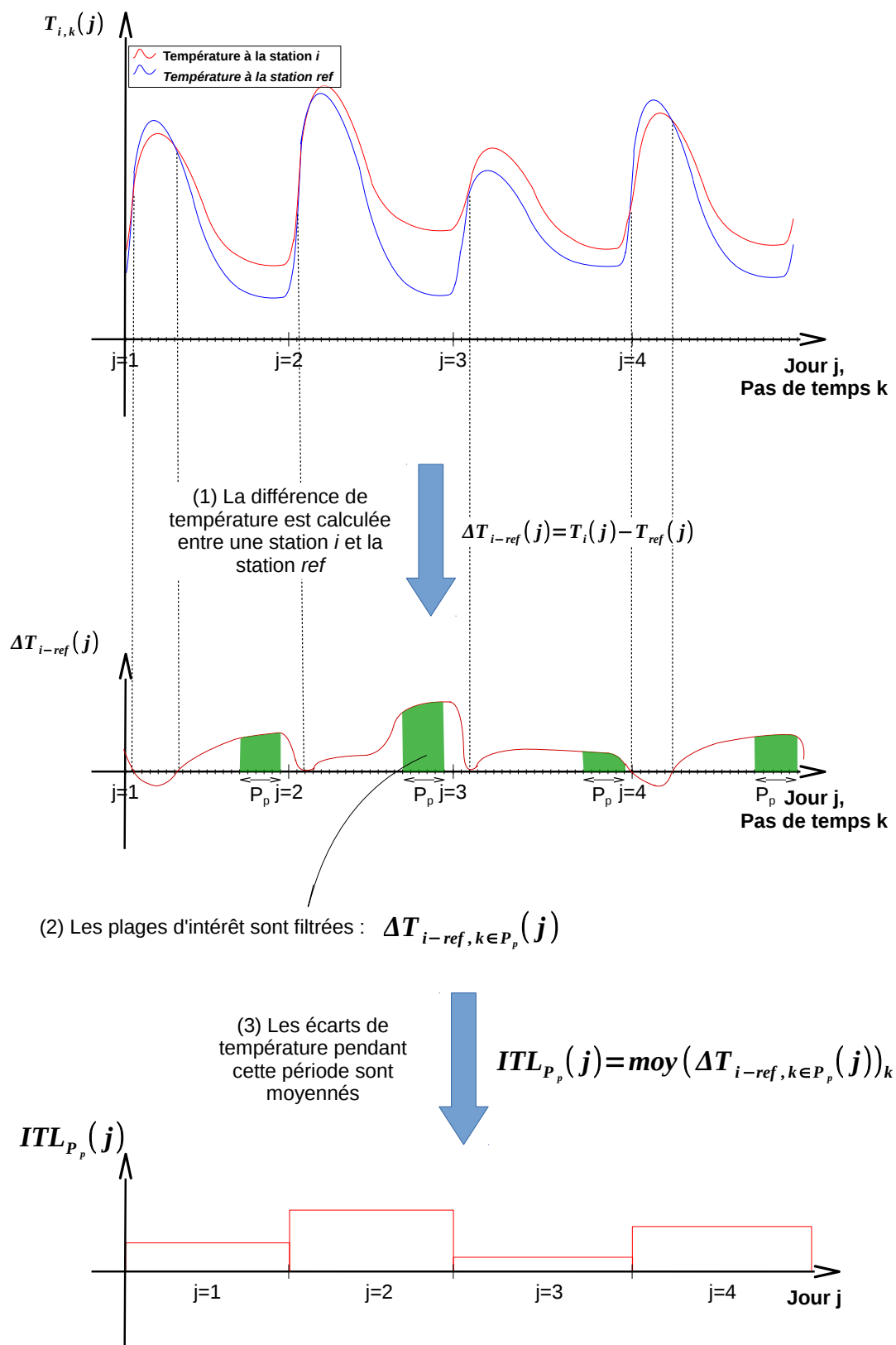


FIGURE 3.4 – Procédure de calcul des indicateurs météorologiques  $ITL$  décrivant les températures locales

enregistrés lorsque la température est maximale. De même, les écarts de température évoluent peu en fin de période nocturne. Ceux mesurés pendant  $P_{ICU}^{(nuit)}$  ont donc toutes les chances d'être identiques à ceux mesurés en début de période diurne, lorsque la température est minimale. À partir de là, nous pourrions calculer un ordre de grandeur des températures minimales journalières  $\min(T_{i,k}(j))_k$  et maximales  $\max(T_{i,k}(j))_k$  de n'importe quel quartier  $i$  de l'agglomération pour un cycle diurne  $j$  donné. L'équation 3.7 sera utilisée :

$$\begin{aligned} \min(T_{i,k}(j))_k &= \min(T_{ref,k}(j))_k + ITL_i^{(nuit)}(j) \\ \max(T_{i,k}(j))_k &= T_{max_{ref}}(j) + ITL_i^{(jour)}(j) \end{aligned} \quad (3.7)$$

avec  $\min(T_{i,k}(j))_k$  et  $\max(T_{i,k}(j))_k$  respectivement les températures minimales et maximales du cycle diurne  $j$  estimées au point  $i$  du territoire (en °C),  $\min(T_{ref,k}(j))_k$  et  $\max(T_{ref,k}(j))_k$  les températures minimales et maximales du cycle diurne  $j$  mesurées à la station de référence,  $ITL_i^{(nuit)}(j)$  et  $ITL_i^{(jour)}(j)$  les indicateurs de température entre un point  $i$  du territoire et la station de référence respectivement pendant les périodes  $P_{ICU}^{(nuit)} = [0, 4 - 0, 8]^{(nuit)}$  et  $P_{ICU}^{(jour)} = [0, 5 - 0, 7]^{(jour)}$  du cycle diurne  $j$ .

### 3.1.2 Îlot de chaleur urbain : procédure de modélisation

Pour chaque  $ITL$  calculé à partir du jeu de données  $JDD_{calib}$ , les variations spatiales et temporelles peuvent être dissociées.

- L'indicateur  $ITL$  a 592 valeurs associées à chacune des 10 stations de mesure : chacune d'elles correspond à un jour du jeu de données  $JDD_{calib}$ . Les différences observables entre deux jours sont par la suite qualifiées de variations temporelles.
- L'indicateur  $ITL$  a 10 valeurs associées à chacun des 592 jours de mesure : chacune d'elles correspond à une station de mesure. Les variations spatiales sont définies comme les différences observables entre les stations de mesure.

Ce sont ces variations spatiales et temporelles que nous devons modéliser par la suite. Ce sera l'objectif de ce chapitre, et une originalité de ce travail, puisqu'aucune des 17 études décrites au chapitre 1 ne proposent de modéliser à la fois les variations spatiales et les variations temporelles de l'ICU. La majeure partie d'entre elles traite des variations spatiales mais des contraintes météorologiques sont appliquées aux jeux de données pour ne conserver

que certaines journées (journées calmes et claires, souvent propices à l'établissement de l'ICU) (Alonso et al. 2007 ; Hart et Sailor 2009). D'autres ne modélisent au contraire que les variations temporelles : l'écart de température entre un point de la ville et la campagne est expliqué à partir des conditions météorologiques propres à chaque journée (Erell et Williamson 2007 ; Oke 1976). Les dernières traitent les deux aspects mais séparément : il n'est donc pas possible de connaître la valeur de l' $ITL$  pour n'importe quel point du territoire et n'importe quel jour  $j$  (Svensson et al. 2002). L'objectif de la méthode proposée ici est de résoudre le problème spatialement et temporellement. Pour cela, un diagnostic des différences spatiales et temporelles est réalisé section 3.1.2.1. La méthode de résolution est ensuite décrite section 3.1.2.2 et les variantes de cette méthode sont présentées section 3.1.2.3.

### 3.1.2.1 Observation de la distribution spatiale et temporelle des indicateurs de température

Les valeurs des indicateurs  $ITL^{(jour)}$  et  $ITL^{(nuit)}$  sont triées par saison, puis pour chaque station, un histogramme de ces valeurs est tracé. La figure 3.5 présente les histogrammes des stations TRH04 et TRH11 pour  $ITL^{(nuit)}$ . Ces deux stations représentent les extrêmes de nos jeux de données pour l'indicateur nocturne : les valeurs de toutes les autres stations, si elles devaient être représentées sur cette figure, seraient globalement distribuées à gauche de celles de la station TRH04 et à droite de celles de la station TRH11.

En-dehors de ces différences spatiales, trois points communs à tous les histogrammes, schématisés sur la figure 3.6, sont identifiés pour  $ITL^{(nuit)}$  (figure 3.5) et également pour  $ITL^{(jour)}$  (non montré) :

- les distributions peuvent toutes être associées à des lois normales (figures 3.5 et 3.6) ;
- ces lois normales semblent toutes avoir une moyenne et un écart-type plus élevés au printemps et en été qu'en automne ou en hiver (figure 3.5) ;
- les journées claires et calmes ainsi que les journées nuageuses et ventées correspondent toujours à des extrêmes des lois normales (figure 3.6). Pour les stations qui sont globalement plus fraîches que la station de référence, les journées claires et calmes correspondent au minimum de leur histogramme. Pour les stations globalement plus chaudes que la station de référence, ces mêmes journées correspondent au maximum de l'histogramme. Ainsi, la dispersion (écart moyen) entre toutes les stations est maximisée sous ces conditions météorologiques.

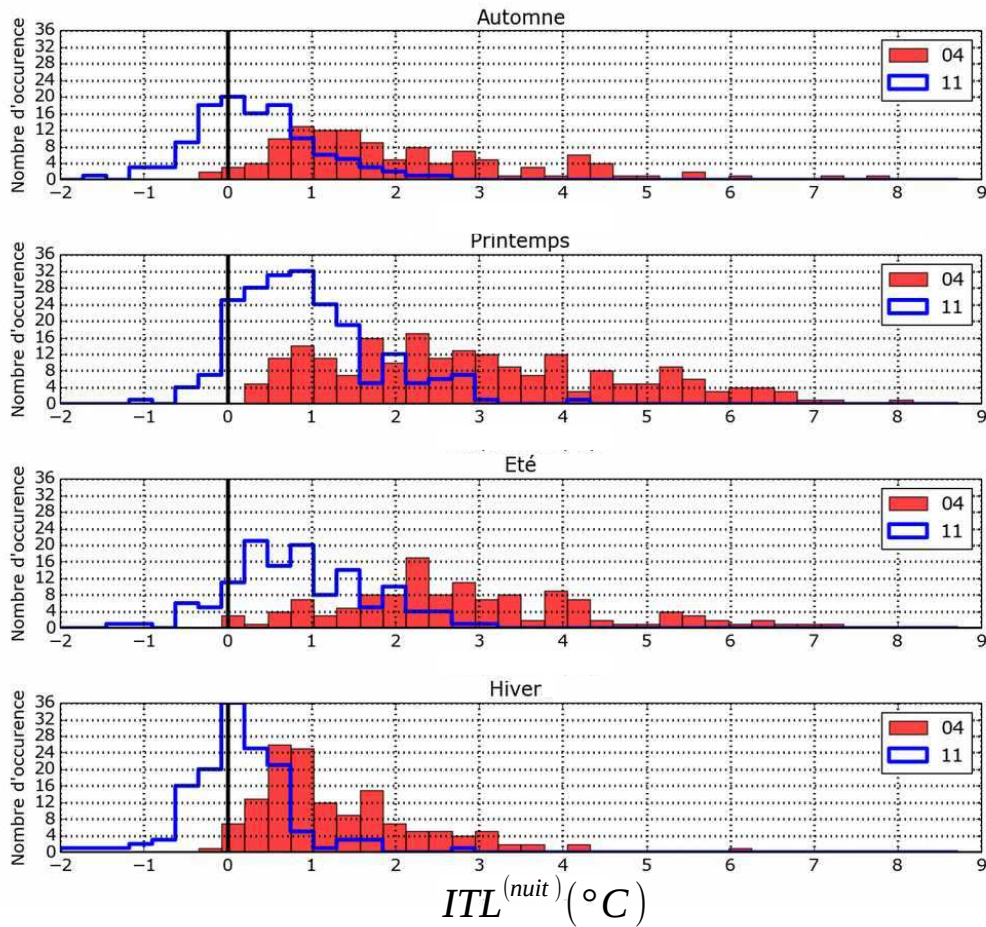


FIGURE 3.5 – Distribution saisonnière des valeurs de l'indicateur de température locale nocturne  $ITL^{(nuit)}$  pour les stations TRH04 et TRH11

### 3.1.2.2 Formalisation d'une méthode de résolution du problème spatial et temporel : notion d'ensemble de modèles

D'après les observations faites à la fin de la section précédente, deux hypothèses simplificatrices sont émises, toutes deux valables pour les deux indicateurs  $ITL$ .

- Pour un indicateur  $ITL$  donné, les valeurs prises par chaque station  $i$  sont distribuées selon une loi normale de moyenne  $moy(ITL_i(j))_j$  et d'écart-type  $std(ITL_i(j))_j$ . Ces deux paramètres sont donc représentatifs seulement des variations spatiales de  $ITL$  (la partie temporelle  $j$  ayant été moyennée). Ils seront donc par la suite notés respectivement  $ITLS_{moy_i}$  et  $ITLS_{std_i}$ .
- Les variations spatiales de l'indicateur  $ITL$  sont éliminées par centrage et réduction. Les lois normales correspondant aux variations temporelles de chaque station  $i$  sont donc transformées selon l'équation 3.8 :

$$ITLT_i(j) = \sigma_i \frac{ITL_i(j) - ITLS_{moy_i}}{ITLS_{std_i}} \quad (3.8)$$

avec  $ITL_i(j)$  et  $ITLT_i(j)$  les valeurs prises par l'indicateur de température locale respectivement avant et après centrage et réduction pour la station  $i$  et le jour  $j$ ,  $ITLS_{moy_i}$  la valeur moyenne de l'indicateur de température locale pour la station  $i$ ,  $ITLS_{std_i}$  l'écart-type de l'indicateur de température locale pour la station  $i$ ,  $\sigma_i$  un coefficient qui est égal à un si  $ITLS_{moy_i} \geq 0$  et égal à -1 si  $ITLS_{moy_i} < 0$ .

L'objectif est d'établir des modèles simples permettant de calculer les indicateurs  $ITLT_i(j)$ ,  $ITLS_{moy_i}$  et  $ITLS_{std_i}$ , afin d'être en mesure de calculer l'indicateur  $ITL_i(j)$ , représentatif des différences de température de n'importe quel point  $i$  pour n'importe quel jour  $j$ . Pour cela, l'équation 3.8 est transformée en l'équation 3.9

$$ITL_i(j) = \frac{ITLT_i(j) \cdot ITLS_{std_i} + ITLS_{moy_i}}{\sigma_i} \quad (3.9)$$

Les valeurs de  $ITLT_i(j)$  pour un jour  $j$  donné doivent être identiques, quelle que soit la station  $i$  (principe illustré figure 3.6). Son écriture sera donc par la suite allégée à  $ITLT(j)$  et cet indicateur correspondra aux variations temporelles de l'ICU.

L'indicateur  $ITLT(j)$  rassemble les valeurs des 10 stations utilisées au sein d'un même vecteur qui ne dépend plus que du temps. Par la suite, la variation de ces valeurs est expliquée à partir des conditions météorologiques de chaque journée  $t$ . Les indicateurs  $ITLS_{std_i}$  et  $ITLS_{moy_i}$ , qui représentent les variations spatiales, seront quant à eux modélisés à partir des informations géographiques qualifiant chaque site  $i$ . La procédure générale de résolution du problème est schématisée figure 3.7.

Elle comprend des étapes déjà décrites : correction des données (section 2.4.5), calcul des  $ITL$ , centrage réduction et concaténation (début de section 3.1.2). Elle inclut également la modélisation temporelle et la modélisation spatiale, deux sous-procédures qui seront décrites respectivement section 3.2 et 3.3. Enfin, des variantes à la procédure par défaut apparaissent sur le schéma. Elles sont décrites section 3.1.2.3.

### 3.1.2.3 Des alternatives à la méthode principale de résolution

À chaque étape de la procédure, plusieurs manières de procéder sont proposées. En effet, aucun élément de connaissance ne permet de justifier l'utilisation d'une manière plutôt que d'une

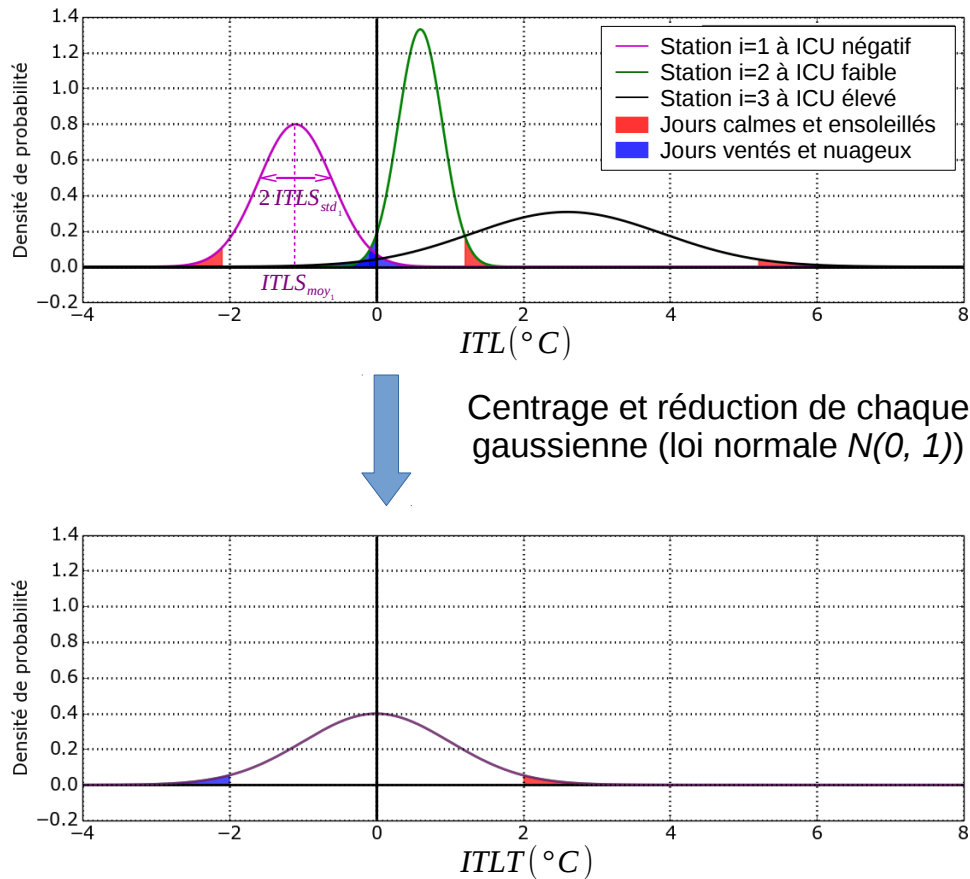


FIGURE 3.6 – Distribution des valeurs de l'indicateur  $ITL$  selon le type de temps avant et après centrage et réduction

autre. Ces variantes, au nombre de cinq et indiquées sur la figure 3.7, utilisent ou non :

1. un tri des données météorologiques selon la direction du vent ;
2. un tri des données météorologiques par saison ;
3. une méthode de centrage réduction qualifiée de robuste ;
4. un nombre limité de stations pour le calibrage des modèles temporels ;
5. une méthode de régression qui soit robuste aux valeurs hors tendance.

À chaque variante est attribuée un booléen : soit elle est utilisée, soit elle n'est pas utilisée. Toutes les combinaisons de variantes sont testées, donnant lieu à la calibration de  $2^4 = 16$  modèles spatiaux (16 modèles pour  $ITLS_{moy}$ , 16 pour  $ITLS_{std}$ ) et de  $2^5 = 32$  modèles temporels  $ITLT$ . Au final, on obtient donc 32 ensembles de modèles (noté EM par la suite), un EM étant lui-même constitué d'un modèle pour  $ITLS_{moy}$ , d'un modèle pour  $ITLS_{std}$  et d'un modèle pour  $ITLT$ . Chaque EM permet d'estimer les valeurs  $ITL$  d'un jeu de données.

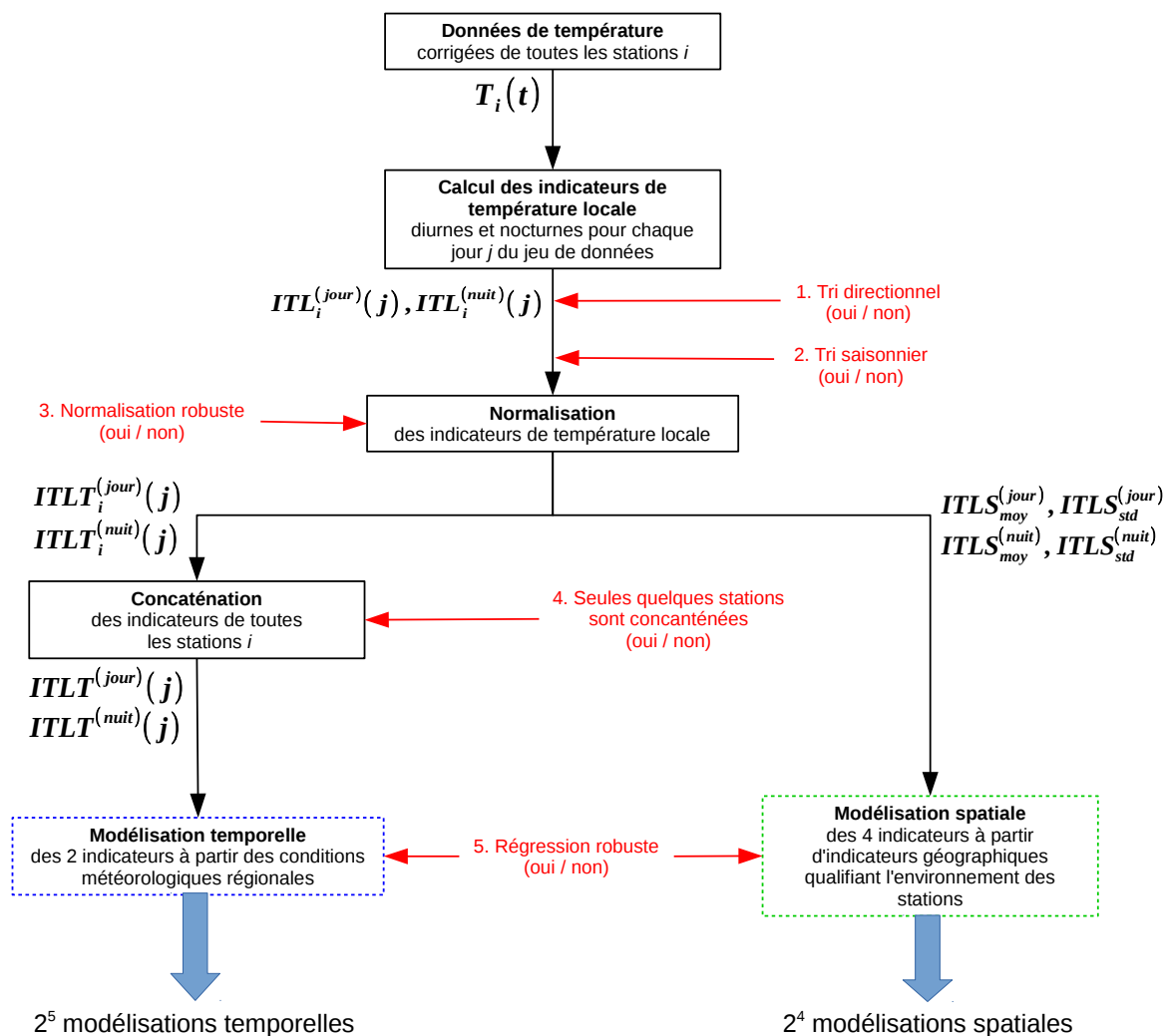


FIGURE 3.7 – Méthodologie générale utilisée pour modéliser les variations spatiales et temporelles de l'ICU

Pour comparer les performances des 32 EM, chacun d'eux sera utilisé pour estimer les valeurs  $ITL$  de tous les jeux de données présentés au chapitre 2. L'EM conservé sera celui dont les estimations seront les plus proches des observations.

### Tri des données météorologiques selon la direction du vent

Le tri des données météorologiques selon la direction du vent est motivé par l'anisotropie géographique des sites de mesure. Jusqu'à présent, les sites sont considérés comme isotropes, c'est-à-dire que les formes urbaines et les types de sols qui les entourent sont les mêmes dans toutes les directions. En réalité, ce n'est jamais parfaitement le cas. Le degré de perfection de cette isotropie est donc étudié en triant les données par direction de vent. La rose des vents est découpée en six parts égales de  $60^\circ$ , et on nomme  $A_{g_i, dir_{pr}}$  la zone géographique



explicative de la température mesurée à la station  $i$  lorsque le vent vient de la direction  $dir_{pr}$ . La figure 3.8 illustre la relation entre direction de vent et zone géographique explicative de la température mesurée par le capteur. La zone géographique est volontairement prise plus large que l'angle correspondant à la direction du vent. Ceci permet de considérer la possible déviation des vents à l'approche de la station de mesure.

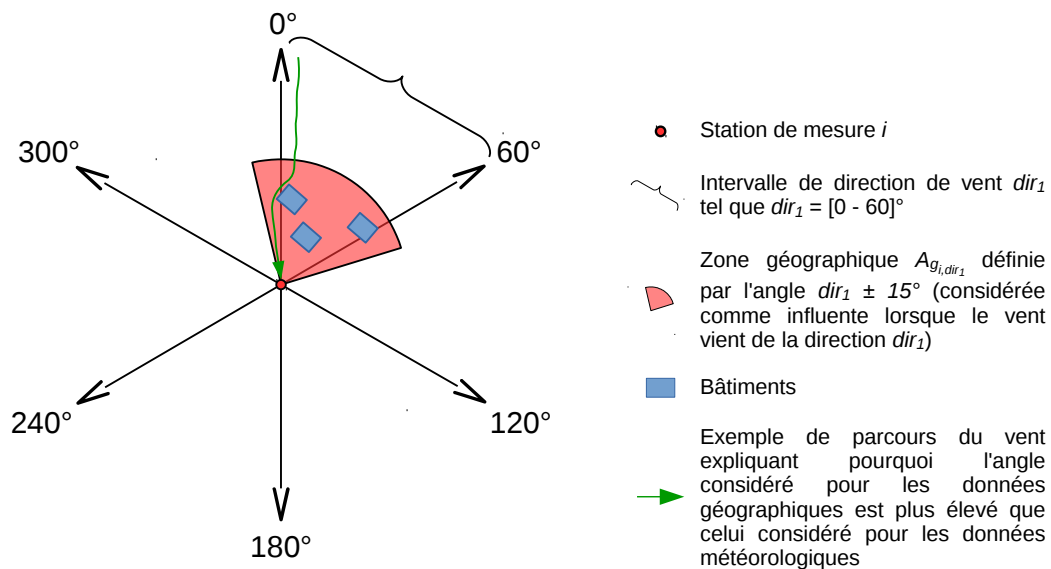


FIGURE 3.8 – Division microclimatique et géographique d'une station de mesure en plusieurs sites : exemple pour la direction de vent  $dir_1 = [0 - 60]^\circ$

La figure 3.9 illustre les effets engendrés par l'application de notre méthode de centrage et réduction à deux sites parfaitement isotropes puis à deux sites anisotropes. La valeur de l'indicateur  $ITLL_{i,dir_{pr}}$  peut être calculée pour une direction  $dir_{pr}$  donnée en modifiant légèrement l'équation 3.8 :

$$ITLL_{i,dir_{pr}}(j) = \sigma_i \frac{ITL_{i,dir_{pr}}(j) - ITLS_{moy_{i,dir_{pr}}}}{ITLS_{std_{i,dir_{pr}}}} \quad (3.10)$$

La valeur de l'indicateur  $ITLL_{i,dir_{pr}}$  calculée pour une station  $i$  peut ensuite être ramenée à la valeur moyenne de toutes les stations :

$$\Delta ITLL_{i,dir_{pr}} = ITLL_{i,dir_{pr}} - \frac{1}{nb_i - 1} \sum_{i_i \neq i}^{nb_i} ITLL_{i_i,dir_{pr}} \quad (3.11)$$

Dans l'exemple de la figure 3.9, lorsque les deux sites sont isotropes, pour n'importe quelle direction  $dir_{pr}$ , nécessairement  $ITLL_{i_1,dir_{pr}} = ITLL_{i_2,dir_{pr}}$ . Dans le cas contraire cela implique que soit la station  $i_1$ , soit la station  $i_2$  ou soit les deux stations présentent une ani-

sotropie. Cependant, utilisée seule, la figure 3.9 ne permet pas de trancher : le problème vient soit de la station  $i_1$  qui serait anormalement chaude sous le vent  $dir_1$ , soit de la station  $i_2$  qui serait anormalement froide sous le vent  $dir_1$ . La rose des vents de  $\Delta ITLT_{i,dir_{pr}}$  (défini équation 3.11), permet d'identifier si la station  $i_1$  est la source du problème.

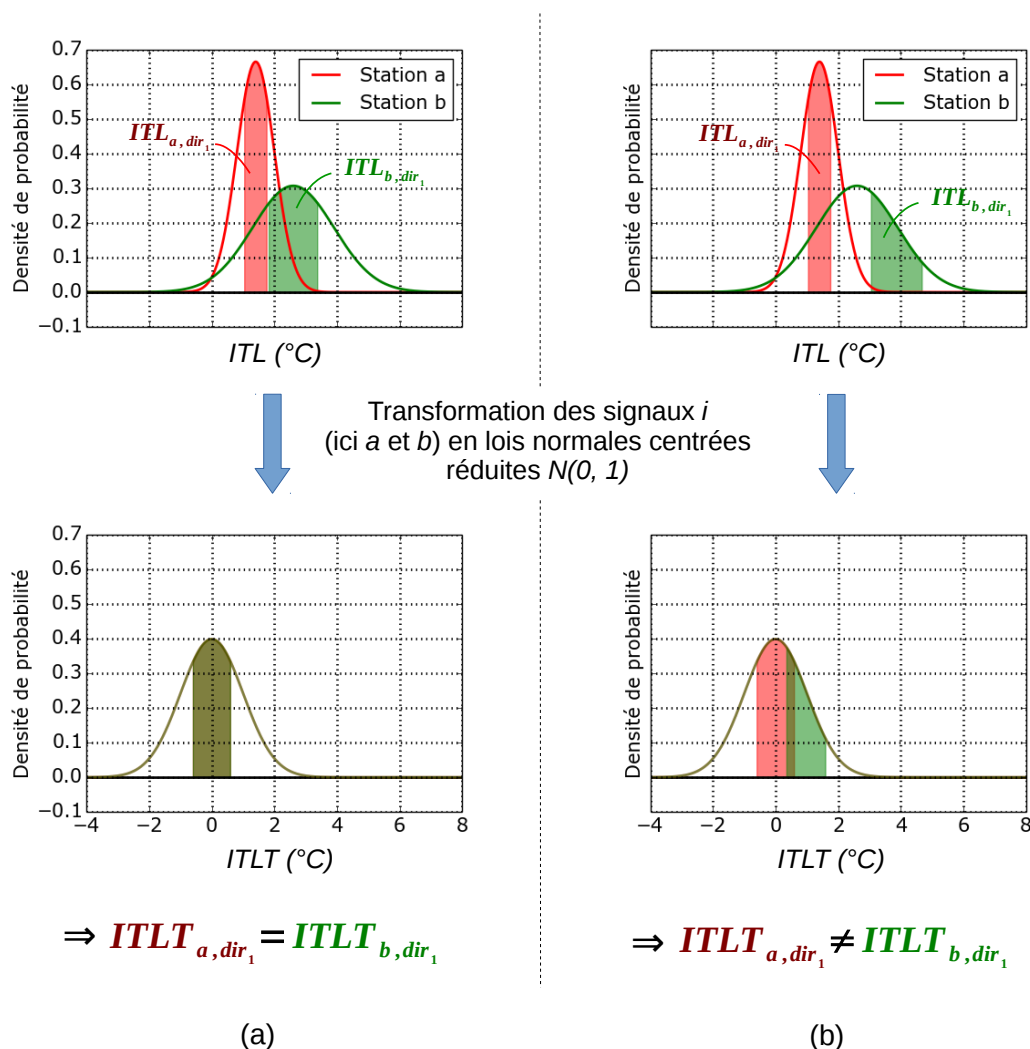


FIGURE 3.9 – Centrage et réduction appliqués aux données de deux stations de mesure (a) lorsque les deux sites sont isotropes, (b) lorsqu'au moins un des deux sites est anisotrope

Le diagnostic de l'anisotropie du site  $i_1$  se fait ensuite en comparant les valeurs de cet indicateur obtenues pour toutes les directions de la rose des vents. Si  $\Delta ITLT_{i,dir_{pr}}$  varie fortement en fonction de la direction, le site peut être considéré comme anisotrope. La figure 3.10 illustre les résultats obtenus à partir de cette méthode d'analyse appliquée aux stations plutôt isotropes (TRH10) ou plutôt anisotropes (TRH11 et MF\_Boug) pour  $ITL^{(nuit)}$ .

- La station TRH11 est anisotrope : sa température est beaucoup plus élevée la nuit lorsque le vent provient de sa moitié sud-ouest que lorsqu'il vient de sa moitié nord-est. Cette analyse microclimatique est cohérente avec le contexte géographique du site : la partie

sud est très minéralisée (parking) et est occupée par des bâtiments de taille élevée. La partie nord contient quant à elle plus d'espaces verts. Ses sols accumulent probablement moins de chaleur que ceux du sud en journée, assurant un refroidissement plus rapide de l'air après le coucher du soleil. Les vents issus du nord sont donc logiquement plus froids que ceux qui viennent de la partie sud du site.

- La station MF\_Boug est elle aussi fortement anisotrope, et cette anisotropie évolue au cours de l'année : la chaleur (ou la fraîcheur) ne vient pas de la même direction en fonction des saisons. La moitié nord-est est plus chaude pendant l'été et le printemps alors que c'est la moitié ouest qui est plus chaude en automne ou en hiver. Le tiers sud est généralement le plus frais quelle que soit la saison. L'analyse du contexte géographique permet d'expliquer ce constat : toutes les directions excepté le tiers sud comportent des espaces minéralisés ou des bâtiments susceptibles de restituer de la chaleur pendant la nuit. Ces éléments géographiques correspondent en grande majorité à des bâtiments ou des pistes de décollage de l'aéroport.
- La station TRH10 semble plus isotrope que les deux autres stations : l'indicateur de température  $\Delta ITLT_{i,dir_{pr}}$  est globalement constant quelle que soit la direction du vent. Ceci est cohérent avec l'isotropie des éléments géographiques observable sur la figure 3.10.

Au final, toutes les stations n'étant pas isotropes, il est possible de considérer que les propriétés d'un site de mesure sont différentes selon la direction du vent. Ceci est cohérent avec la proposition de Oke (2004) de considérer deux sites distincts selon la direction du vent lorsqu'une station est située à l'interface entre deux LCZ. Cela permet d'éviter de perdre de l'information spatiale en ne moyennant pas un indicateur géographique qui prendrait une valeur différente selon la direction analysée. Cela permet également de multiplier *artificiellement* le nombre de stations (initialement faible - 10) afin d'apporter plus de fiabilité aux analyses statistiques qui seront réalisées.

Ce gain d'information spatiale se fait néanmoins au détriment de l'information temporelle puisque le nombre d'échantillons par site sera dépendant du nombre de jours où le vent soufflera dans chaque direction. Afin de donner autant d'importance aux variations spatiales qu'aux variations temporelles, il est nécessaire de trouver un compromis entre quantité d'échantillons spatiaux et quantité d'échantillons temporels. Dans le cas où la variante de tri directionnel est utilisée, la rose des vents est ainsi coupée en six et l'équation 3.10 remplace l'équation 3.8.

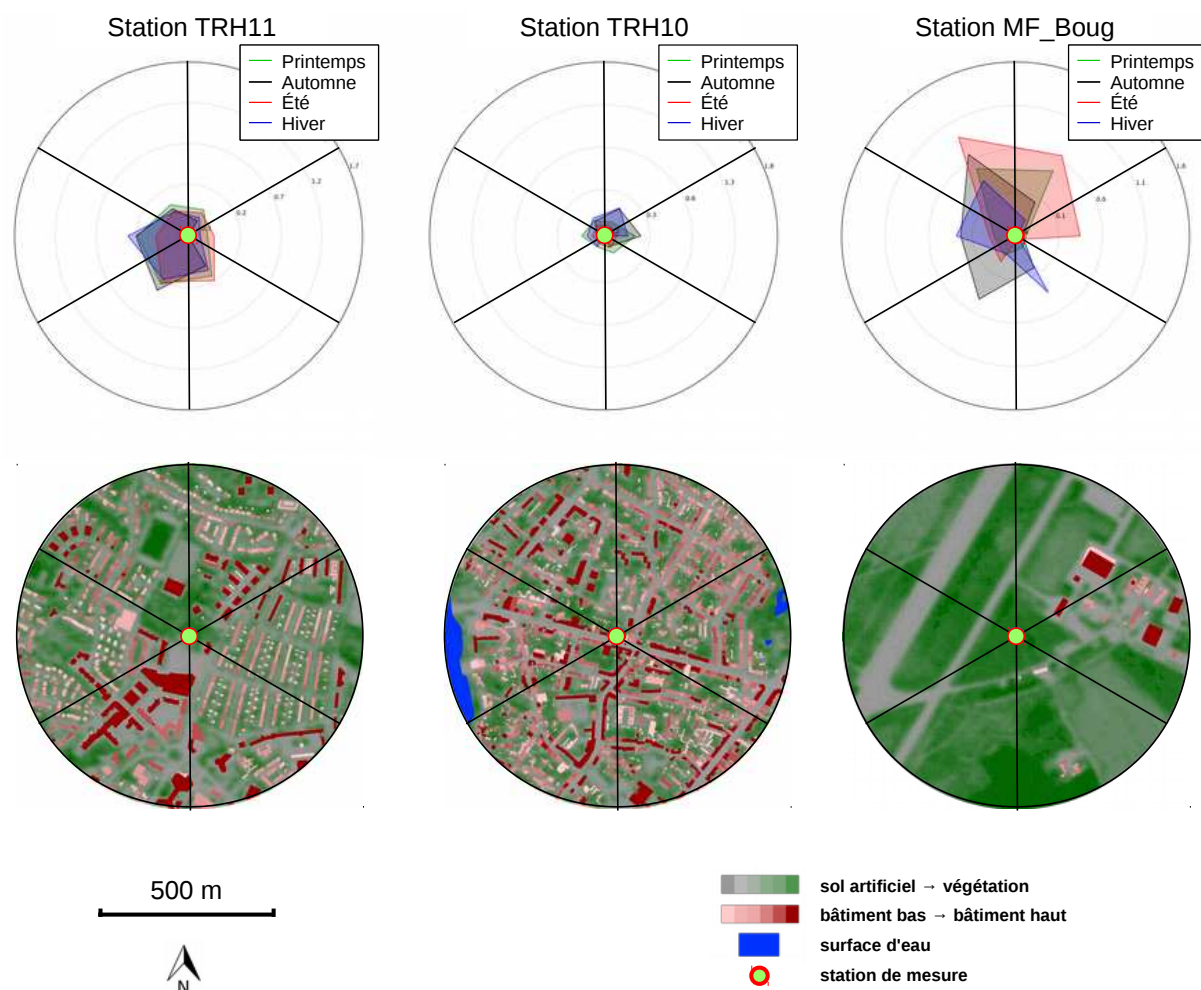


FIGURE 3.10 –  $\Delta ITLLT_{i,dir_{pr}}$  et contexte géographique des stations TRH11, TRH10 et MF\_Boug selon la direction du vent. Cette figure permet de mettre en évidence les directions de vent pour lesquelles un site de mesure se comporte différemment des autres directions. L'indicateur  $\Delta ITLLT_{i,dir_{pr}}$  est ainsi calculé à partir de l'équation 3.11 pour chacune des directions de vent

La qualité de l'estimation des indicateurs spatiaux  $ITLS_{moy_{i,dir_{pr}}}$  et  $ITLS_{std_{i,dir_{pr}}}$  dépend ainsi :

- du nombre de journées utilisées pour les calculer : plus ce nombre est important, meilleure devrait être l'estimation ;
- du caractère aléatoire du tirage des journées : si ce n'est pas le cas, cela peut nuire aux modèles qui seront élaborés.

Dans le cas où un tri directionnel a lieu, ces deux conditions sont sans doute moins respectées puisque le nombre de journées utilisées pour qualifier chaque site est plus faible. Souvent, chaque direction de vent est également associée à un type de temps particulier. Si c'est le cas, la valeur des indicateurs  $ITLS_{moy_{i,dir_{pr}}}$  et  $ITLS_{std_{i,dir_{pr}}}$  est fortement dépendante

des conditions météorologiques associées à chaque direction de vent, alors qu'elles devraient seulement dépendre du contexte géographique de chaque site. Afin de diagnostiquer si certaines directions de vent sont propices à biaiser les résultats de nos futures analyses, les valeurs de chacun des indicateurs  $ITLS_{moy_i,dir_{pr}}$  et  $ITLS_{std_i,dir_{pr}}$  sont comparées entre elles. La figure 3.11 présente les résultats de l'analyse pour  $ITL^{(nuit)}$ .

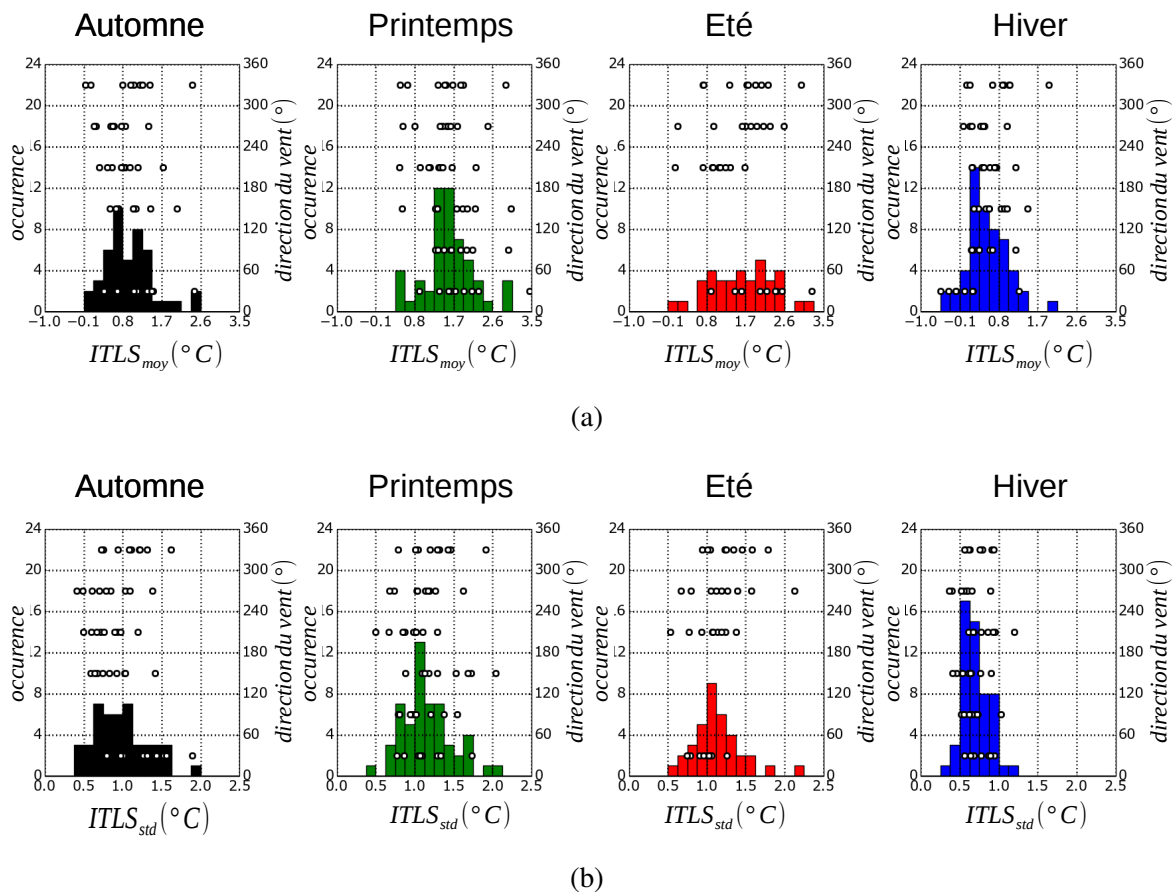


FIGURE 3.11 – Distribution (histogramme) des valeurs et valeurs pour chaque station et chaque direction de vent (a) de l'indicateur  $ITLS_{moy}$  et (b) de l'indicateur  $ITLS_{std}$  lorsque la variante *tri directionnel* est utilisée

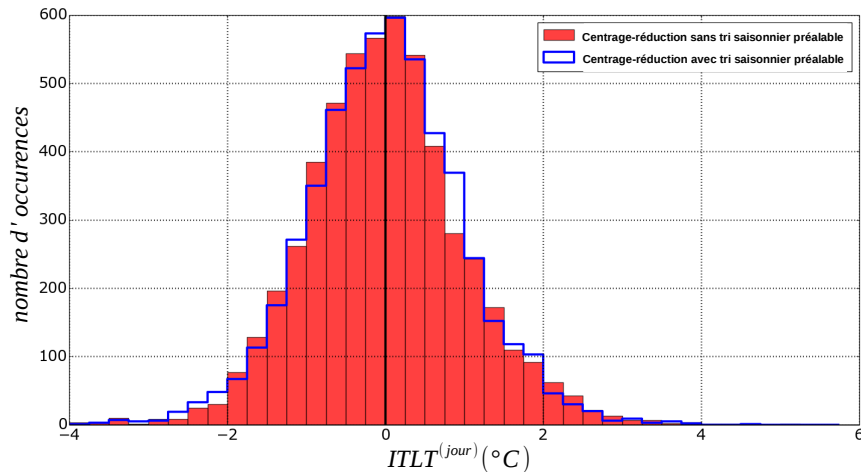
Chaque point correspond à la valeur prise par une station pour une direction de vent donnée (axe des ordonnées situé à droite). Une direction de vent est considérée comme potentiellement problématique dès lors que la valeur des indicateurs diffère largement de celles calculées pour les autres directions. Des valeurs particulières apparaissent ainsi en été pour des vents qui proviennent de directions comprises entre 0 et 60°. En effet, la valeur de l'indicateurs  $ITLS_{std_i,dir_{pr} \in [0-60]^\circ}$  est faible quelle que soit la station  $i$ . Ceci correspond à une faible variation des conditions météorologiques. Même si les températures et la vitesse du vent sont assez variables pour cette direction de vent, ces journées sont souvent peu nuageuses (cf figure 2.8). Ceci explique pourquoi les valeurs moyennes  $ITLS_{moy_i,dir_{pr} \in [0-60]^\circ}$  sont relativement

élevées par rapport à celles obtenues pour les autres directions de vent  $dir_{pr}$ . Au printemps, ce sont les directions de vent comprises entre 180 et 240° qui sont associées à des variabilités faibles. Une fois de plus, ce constat est cohérent avec les conditions météorologiques généralement nuageuses (les quartiles hauts et bas de nébulosité sont respectivement 7.2 et 5.4 octa). Cette fois-ci, les moyennes sont donc faibles comparées à celles des autres directions. En hiver, le même raisonnement et les mêmes conclusions sont tirées pour des vents en provenance de l'ouest (vents compris entre 240 et 300°). En automne, toutes les directions comprises entre 120 et 300°, associées à des types de temps presque exclusivement nuageux, entraînent des valeurs  $ITLS_{std, dir_{pr} \in [120-300]^\circ}$  faibles. Lors de l'analyse des résultats de modélisation, une attention particulière sera portée à ces directions de vent lorsque le tri directionnel aura été utilisé. La figure 3.11 permet également de vérifier que la répartition des indicateurs  $ITLS_{moy, dir_{pr}}$  et  $ITLS_{std, dir_{pr}}$  est bien assimilable à une loi normale. Pour cela, l'histogramme des valeurs est tracé pour chaque indicateur et chaque saison. La taille de l'échantillon n'est pas suffisante pour affirmer avec certitude la normalité de chaque distribution mais les histogrammes ont presque tous la forme d'une gaussienne. Au final, le tri directionnel permettra d'utiliser l'équivalent de 60 stations de mesure au lieu de 10 par défaut. Pour chacune des 10 stations, les directions de vent permettent de différencier géographiquement et microclimatiquement six sites : chacun est défini selon la méthode présentée figure 3.8.

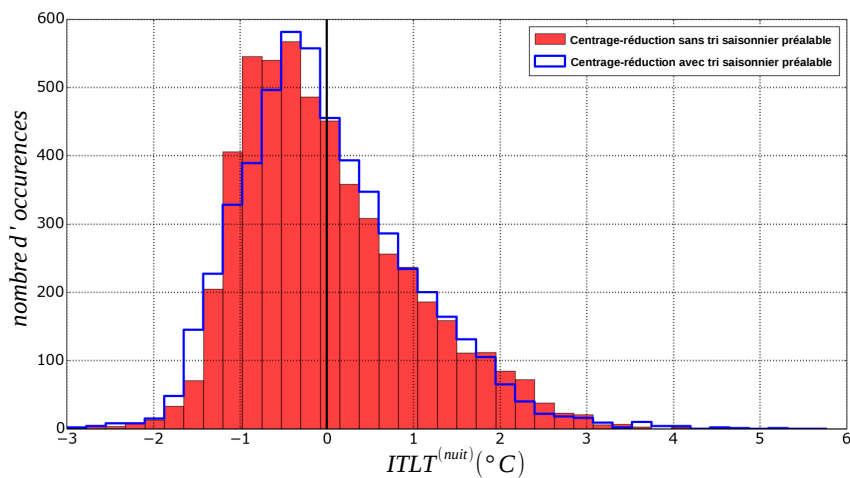
## 2. Tri des données météorologiques par saison

Un tri saisonnier est presque toujours réalisé dans les études de microclimatologie urbaine. Ceci permet d'identifier les éléments à l'origine du développement de l'ICU pour chaque saison, ceux-ci n'étant pas forcément les mêmes tout au long de l'année. Sous le climat de nos zones d'études, la chaleur dégagée par les bâtiments en hiver peut par exemple être à l'origine des températures plus chaudes observées dans les quartiers les plus densément bâtis. Au printemps et en été, ce sont les zones les plus végétalisées qui sont souvent les plus froides car elles sont sources d'ombrage ou d'évapotranspiration. Or, même si c'est souvent le cas, les zones les plus densément bâties ne sont pas nécessairement les moins végétalisées. Établir un EM commun à toutes les saisons pour expliquer les différences de température à partir d'éléments géographiques n'est donc pas très pertinent pour certaines saisons. Dans le même temps, développer l'équivalent d'un EM par saison soulève le problème de continuité temporelle. Deux jours consécutifs météorologiquement comparables mais appartenant tous les deux à une saison différente pourront conduire à des estimations de température complètement différentes. Les

deux méthodes ont donc leurs avantages et leurs inconvénients, elles seront testées toutes les deux : quatre EM seront établis dans le cas de l'utilisation du tri saisonnier, un seul sera établi à partir de toutes les données le cas échéant. La normalité de la distribution des valeurs de  $ITLT(j)$  est testée dans les deux cas. La figure 3.12 permet de comparer les distributions.



(a)



(b)

FIGURE 3.12 – Distribution des valeurs (a) de l'indicateur  $ITLT^{(jour)}(j)$  et (b) de l'indicateur  $ITLT^{(nuit)}(j)$  selon si la variante *tri saisonnier* est utilisée

Dans le cas du tri saisonnier, les valeurs  $ITLT(j)$  sont calculées seulement après que les données ont été triées par saison. Ceci explique la différence des deux histogrammes dans le cas de l'indicateur nocturne : la forme obtenue dans le cas du tri saisonnier est plus proche d'une gaussienne que lorsque toutes les données sont centrées et réduites sans tri saisonnier préalable.

### 3. Méthode de centrage et réduction qualifiée de robuste

Le centrage-réduction des indicateurs  $ITL$  est réalisé à partir de la valeur moyenne et de l'écart-type, qui sont des indicateurs peu robustes aux valeurs aberrantes. Dans le cas d'une population normalement distribuée, la médiane permet d'estimer la moyenne de manière robuste. L'écart-type doit être estimé à partir de l'écart absolu médian à la médiane (EAM) et d'un facteur multiplicatif  $\sigma = 1,483$ . Ce facteur permet d'ajuster l'échelle afin que l'EAM soit un bon estimateur de l'écart-type. Cette valeur n'est valide que dans le cas d'une distribution gaussienne. L'EAM, pour  $x = ITL^{(jour)}$  ou  $x = ITL^{(nuit)}$ , est défini par :

$$eam(x_i(j))_j = \text{mediane}(|x_i(j) - \text{mediane}(x_i(j))_j|)_j \quad (3.12)$$

avec  $i$  la station de mesure et  $j$  un cycle diurne

Lorsque le centrage et la réduction utilise la variante robuste, la médiane et le produit  $\sigma \cdot eam(x)$  sont substitués respectivement à la moyenne et à l'écart-type.

### 4. Utilisation d'un nombre limité de stations pour le calibrage des modèles temporels

L'objectif de la modélisation temporelle est d'expliquer les variations de l'indicateur  $ITL_i(j)$  à partir des conditions météorologiques. Toutes les stations sont utilisées par défaut pour l'élaboration de ce modèle. Or, toutes les stations  $i$  ne sont pas impactées dans les mêmes proportions par les conditions météorologiques : la figure 3.5 permet de constater que la variance de la station TRH04 est beaucoup plus élevée que celle de la station TRH11. La variance selon  $j$  d'une station  $i$  peut être déclinée comme une somme de variances :

$$var_{tot}(ITL_i(j))_j = var_{met}(ITL_i(j))_j + var_{mes}(ITL_i(j))_j + var_{res}(ITL_i(j))_j \quad (3.13)$$

avec  $var_{tot}(ITL_i(j))_j$  la variance totale selon  $j$  de l'indicateur  $ITL$  pour la station  $i$ ,  $var_{met}(ITL_i(j))_j$  la variance attribuable à la variation des conditions météorologiques,  $var_{mes}(ITL_i(j))_j$  la variance attribuable à l'incertitude des appareils de mesure,  $var_{res}(ITL_i(j))_j$  la variance résiduelle dont on ne connaît pas la cause.

On considère que la variance de mesure et la variance résiduelle sont identiques pour toutes les stations. La variance totale prend des valeurs disparates selon la station considérée. On considérera donc que cette variation de la variance est seulement imputable aux conditions météorologiques. Dans ces conditions, plus la variance totale d'une station est importante, plus



sa variance météorologique en représente une part importante. Les modèles seront donc a priori mieux estimés s'ils sont établis à partir des stations qui ont les plus fortes variances que s'ils sont estimés à partir de toutes les stations. Seules les trois stations ayant les variances les plus élevées sont donc utilisées pour calibrer les modèles temporels. Selon si certaines variantes (tri saisonnier, tri directionnel, etc.) sont ou non utilisées, ces trois stations sont susceptibles de différer. Notons que la station TRH04, située dans l'hyper-centre, est la station la plus fréquemment présente dans ce trio.

### 5. Utilisation de méthodes de régression robustes aux valeurs hors tendance

Les modèles spatiaux et temporels seront établis à partir de régressions linéaires. La méthode des moindres carrés est traditionnellement utilisée pour la détermination des coefficients de cette régression. Néanmoins, elle n'est pas robuste aux valeurs qui sortent de la tendance. D'autres méthodes, appelées méthodes de M-estimation permettent de remédier à ce problème. L'estimateur *de Huber* est ainsi utilisé.

## 3.2 Modélisation temporelle par régression linéaire multiple

Les variations temporelles  $ITLT(j)$  sont modélisées à partir des conditions météorologiques méso-échelles par régression linéaire multiple.

$$ITLT(j) = \alpha_{t_0} + \sum_m \alpha_{t_m} x_m \quad (3.14)$$

avec  $\alpha_{t_0}$  une constante de régression,  $\alpha_{t_m}$  le coefficient de régression de la variable explicative  $x_m$ .

Chaque variable explicative  $x_m$  est une fonction  $f_m$  d'une variable météorologique  $VM_m$  moyennée pendant une période  $P_m$ . Chaque variable explicative s'exprime donc selon la relation :

$$x_m = f_m(\text{moy}(x_{m_{k \in P_m}}(j)))_k \quad (3.15)$$

Pour simplifier les notations et alléger les équations, on définit l'indicateur météorologique  $IMR_m$  tel que :

$$IMR_m = \text{moy}(x_{m_{k \in P_m}}(j)) \quad (3.16)$$

Cet indicateur décrit les conditions météorologiques (d'une variable  $VM_m$ ) à l'échelle régio-

nale.

Cette section est construite en deux temps. Elle consiste d'abord à identifier :

- les variables météorologiques qui peuvent être utilisées pour construire les variables explicatives  $x_m$  (et qui ont donc potentiellement de l'influence sur les valeurs de  $ITLT(j)$ );
- la période  $P_m$  pendant laquelle il est préférable de moyenner chacune de ces variables météorologiques pour maximiser l'intensité de leur lien avec  $ITLT(j)$ ;
- la fonction usuelle la plus appropriée pour expliquer la forme de la relation qui lie les variables  $ITLT(j)$  et  $x_m$  pour chaque variable météorologique  $VM_m$ ;

Il s'agit ensuite de déterminer dans l'équation 3.14 :

- la combinaison des variables explicatives  $x_m$  qui permet d'expliquer au mieux les variations de  $ITLT(j)$ ;
- la valeur des coefficients de régression  $\alpha_{t_0}$  et  $\alpha_{t_m}$ .

### 3.2.1 Élaboration des variables explicatives

#### 3.2.1.1 Proposition de variables météorologiques

Les variables météorologiques les plus souvent utilisées pour expliquer les variations temporelles de l'ICU sont la vitesse du vent et la nébulosité. En effet, de jour comme de nuit, un vent fort est souvent associé à un brassage de l'air qui conduit à l'homogénéisation des températures à l'échelle de l'agglomération. En journée, si la nébulosité est élevée, le rayonnement solaire net reçu par la ville est généralement faible. Or, le piégeage radiatif engendré par les rues canyons est un des phénomènes physiques à l'origine de l'ICU. Un rayonnement solaire faible est donc souvent synonyme d'ICU faible. En présence de nuages nocturnes, la température de ciel est plus élevée, ce qui augmente le rayonnement GLO descendant, réduisant ainsi le refroidissement de la ville. Or, le refroidissement plus rapide des surfaces rurales est à l'origine de l'ICU. Si le refroidissement des surfaces de tous les sites est limité, les différences de température entre les sites le sont également. En résumé, le développement de l'ICU a lieu lorsque la nébulosité et le vent sont faibles. Ces deux variables seront utilisées dans notre étude.

D'autres variables météorologiques peuvent également expliquer pourquoi les valeurs de l'ICU varient : Le Bras (2015) utilise l'humidité relative et la température de l'air ainsi que

les rayonnements solaire et infra-rouge descendants. La pression atmosphérique, l'intensité de l'inversion thermique verticale, et la quantité d'eau contenue dans la couche de surface du sol sont utilisées par Lauwaet et al. (2016). Toutes nos zones d'étude ne disposent pas de données de rayonnement IR, le gradient de l'inversion thermique ou de contenu en eau des sols. Ces variables ne sont donc pas considérées dans notre étude.

Les journées qui comportent des périodes de précipitation ne sont généralement pas utilisées dans les études qui traitent de l'ICU car l'eau issue des précipitations est difficile à intégrer dans une modélisation simplifiée de l'ICU. Ceci peut s'expliquer par plusieurs raisons.

- La hauteur de précipitation  $HP$  a une distribution très différente d'une loi normale, qui est une condition pour utiliser une variable dans une régression linéaire.
- Les champs de pluie peuvent être très disparates sous certains événements pluvieux : Les quantités d'eau reçues par chaque quartier de notre zone d'étude pourraient fortement varier spatialement.
- Suite aux précipitations, la dynamique des températures peut parfois être complexe à modéliser : selon la saison, la période du jour et le type de pluie, les températures de l'air lors de précipitations sont affectées différemment. Au milieu d'un après-midi estival, la température diminue ainsi brusquement tandis qu'elle augmente dans le cas où l'atmosphère est chaude et où les surfaces sont froides.
- D'un sol à l'autre, le bilan énergétique peut différer fortement pendant et après des précipitations : un sol perméable absorbe l'eau, ce qui peut lui permettre d'évaporer ou d'évapotranspirer plusieurs heures, voire plusieurs jours après la pluie. Un sol imperméable évapore sur une durée moins longue après l'évènement pluvieux car l'eau est très souvent drainée par les égouts.

Seules les journées où la pluie tombe sans interruption sur l'ensemble du territoire étudié sont faciles à appréhender puisqu'elles correspondent souvent à des périodes où l'ICU est nul. Dans cette situation, la couverture nuageuse est souvent importante. Une forte nébulosité étant déjà associée à un ICU faible, dire que des chutes de pluie et une nébulosité élevées sont synonymes d'ICU faible seraient donc redondants. Pour toutes ces raisons, les précipitations ne seront pas considérées comme variable explicative de l'ICU.

Au final, celles conservées sont : la vitesse du vent ( $U(t)$ ), la nébulosité ( $N(t)$ ), le rayonnement solaire descendant ( $K \downarrow (t)$ ), la température de l'air ( $T(t)$ ), l'humidité relative de

l'air ( $RH(t)$ ) et la pression atmosphérique ( $PA(t)$ ).

### 3.2.1.2 Proposition de périodes d'intérêt

Pour rappel, les variables météorologiques sont mesurées au pas horaire aux stations Météo-France décrites section 2.2.2 puis interpolées au pas quart-horaire. Ces données sont utilisées pour expliquer les indicateurs  $ITLT(j)$  calculés pendant les périodes  $P_{ICU}^{(jour)}$  et  $P_{ICU}^{(nuit)}$ . Il est commun d'utiliser les variables météorologiques moyennées pendant ces mêmes périodes, même si certaines variables se prêtent parfois mal à l'exercice. C'est par exemple le cas du rayonnement solaire si l'ICU nocturne est étudié : il sera plus pertinent de considérer sa valeur en journée plutôt que sa valeur la nuit pour deux raisons. La première est que le rayonnement solaire est nul la nuit. La seconde est que l'énergie solaire reçue et stockée dans les matériaux pendant la journée est responsable de l'ICU nocturne (restitution de la chaleur accumulée pendant la nuit). Ainsi, Le Bras (2015) calcule le rayonnement solaire moyen durant les 12 heures qui précèdent la mesure des différences de température. Afin d'identifier la période  $P_m$  la plus appropriée pour expliquer les variations de  $ITLT(j)$ , nos six variables météorologiques sont moyennées pendant quatre intervalles de temps différents (illustrés figure 3.13) :

- période A : période similaire à celle ayant permis le calcul de nos indicateurs  $ITL^{(nuit)}$  et  $ITL^{(jour)}$  (respectivement  $P_A = [0, 4 - 0, 8]^{(nuit)}$  et  $P_A = [0, 5 - 0, 7]^{(jour)}$ );
- période B : avant la période de calcul de nos indicateurs mais dans la même partie de la journée ( $P_B = [0 - 0, 4]^{(nuit)}$  pour  $P_{ICU}^{(nuit)}$  et  $P_B = [0 - 0, 5]^{(jour)}$  pour  $P_{ICU}^{(jour)}$ );
- période C : pendant la partie précédente de la journée mais après la période de calcul de l' $ITL$  associé ( $P_C = [0, 7 - 1]^{(jour)}$  pour  $P_{ICU}^{(nuit)}$  et  $P_C = [0, 8 - 1]^{(nuit)}$  pour  $P_{ICU}^{(jour)}$ );
- période D : pendant la partie précédente de la journée intégrant la période de calcul de l' $ITL$  associé ( $P_D = [0 - 1]^{(jour)}$  pour  $P_{ICU}^{(nuit)}$  et  $P_D = [0 - 1]^{(nuit)}$  pour  $P_{ICU}^{(jour)}$ ).

La méthodologie de choix de la période d'intérêt sera décrite section 3.2.1.4.

### 3.2.1.3 Proposition de fonctions usuelles

Les indicateurs de température  $ITLT(j)$  ne varient pas nécessairement linéairement avec les indicateurs météorologiques  $IMR_m$  (définis équation 3.16). Svensson et al. (2002) utilisent par exemple le logarithme népérien du vent et créent pour la nébulosité trois modèles linéaires différents pour expliquer les variations de l'ICU. Erell et Williamson (2007) utilisent des polynômes

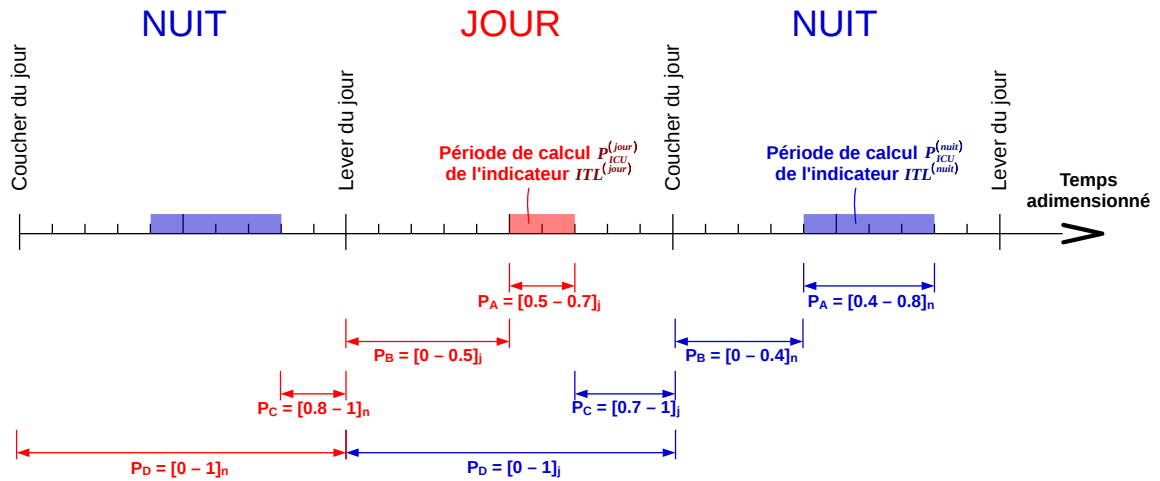


FIGURE 3.13 – Périodes  $P_m$  utilisées pour calculer les indicateurs météorologiques  $IMR$  associés à chacun des indicateurs  $ITL^{(nuit)}$  et  $ITL^{(jour)}$

du deuxième ordre pour ces deux mêmes variables tandis que Oke (1976) applique la fonction  $\frac{1}{V^{0.46}}$ . Afin d'identifier la forme la plus appropriée à chaque indicateur  $IMR_m$ , les nuages de points  $ITLT(j) = f(IMR_m)$  sont tracés. La figure 3.14 illustre les résultats obtenus pour l'indicateur nocturne.

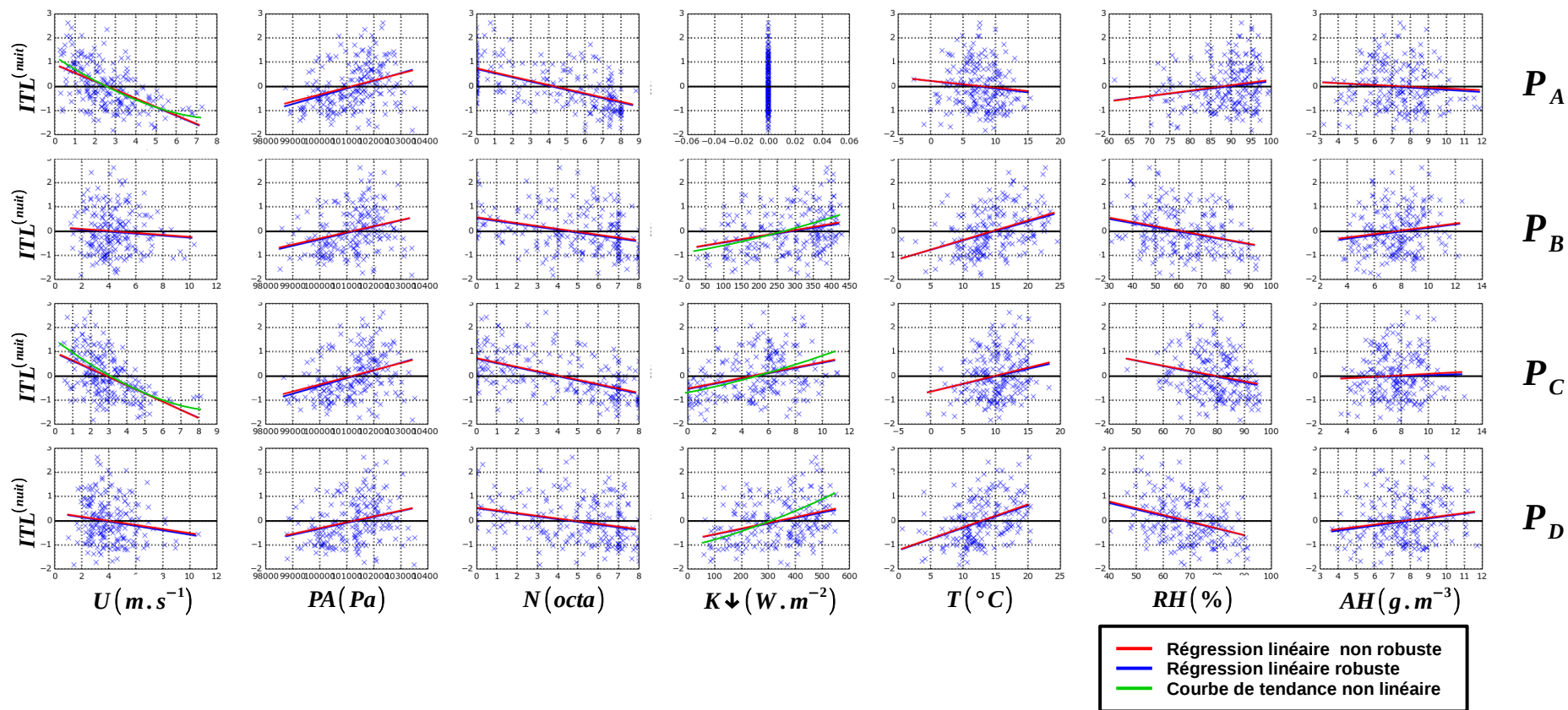


FIGURE 3.14 – Valeurs prises par l'indicateur nocturne  $ITLT^{(nuit)}(j)$  en fonction de chaque indicateur météorologique  $IMR_m$  calculé pour chacune des périodes  $P_m$

Aucune fonction usuelle et aucune période ne s'impose clairement pour chaque variable. Plusieurs fonctions sont donc proposées pour linéariser la relation  $ITLT(j) = f(IMR_m)$ . Par exemple, l'ICU nocturne varie ainsi :

- linéairement ou quadratiquement avec le rayonnement solaire quelle que soit la période  $P_p$  utilisée (sauf  $P_A$ ) ;
- linéairement, sous la forme d'une fonction inverse ou d'une exponentielle décroissante pour la vitesse de vent (les tendances étant les plus nettes pour les périodes  $P_A$  et  $P_C$ ).

La méthodologie de choix des fonctions et des périodes les plus appropriées à chaque variable sera décrite dans la section suivante.

### 3.2.1.4 Procédure utilisée pour choisir les variables explicatives les plus pertinentes

Afin de former chaque terme  $x_m$  de l'équation 3.14, une seule combinaison {période  $P_m$  + fonction  $f_m$ } doit finalement être conservée pour chaque variable  $VM_m$ . La procédure suivie pour choisir cette combinaison est schématisée figure 3.15.

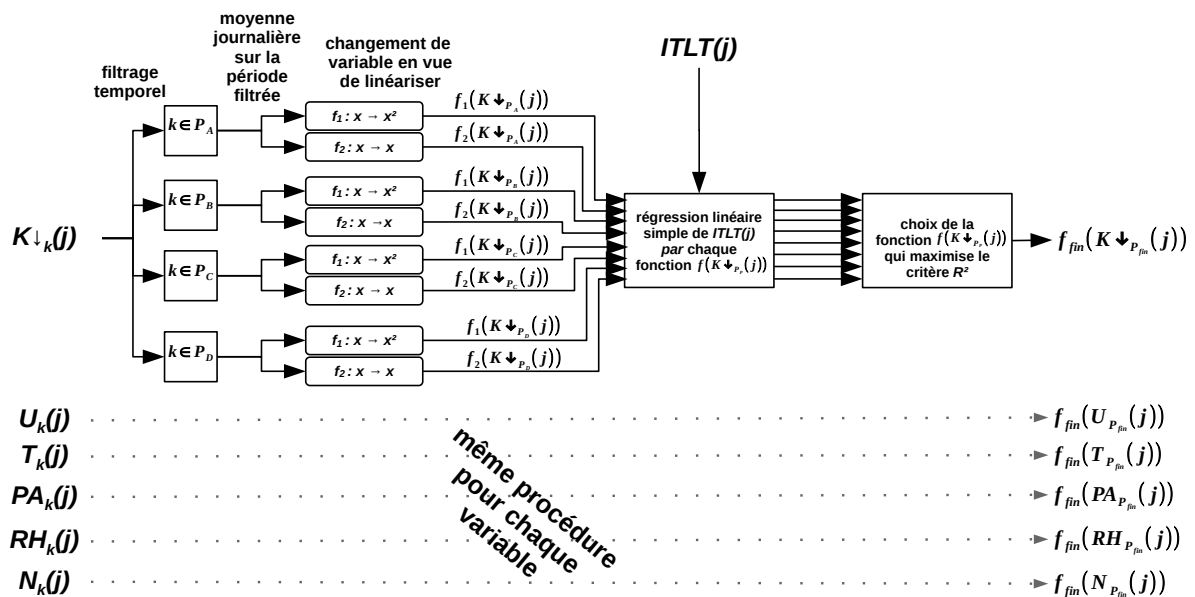


FIGURE 3.15 – Procédure utilisée pour déterminer le terme  $x_m$  (propre à chaque variable météorologique  $VM_m$ ) le plus approprié pour expliquer les variations de l'indicateur  $ITLT(j)$

Pour chaque variable  $VM_m$ , toutes les combinaisons  $\{P_m, f_m\}$  sont testées par régression linéaire simple de  $ITLT(j)$  par  $x_m$ . La meilleure combinaison est choisie lorsque le coefficient de détermination  $R^2$  est maximal. À noter que selon les variantes présentées section 3.1.2.3, la combinaison choisie peut différer.

### 3.2.2 Procédure d'identification du modèle temporel le plus pertinent

Tous les termes  $x_m$  finalement identifiés n'ont pas le même effet sur les variations de  $ITLT(j)$  : certains peuvent être négligeables, d'autres peuvent être redondants. La redondance est obtenue lorsqu'un terme  $x_m$  est une combinaison linéaire d'un ou plusieurs autres termes. Si le  $R^2$  de cette combinaison linéaire est élevé, les variables utilisées sont qualifiées de fortement colinéaires. Plus le nombre de termes utilisé est important, plus le risque de colinéarité est élevé. De plus, l'utilisation d'un nombre trop grand de termes conduit souvent à une imprécision sur la valeur des coefficients de régression  $\alpha_{t_0}$  et  $\alpha_{t_m}$ . Le modèle établi dans ces conditions est souvent peu performant lorsqu'il est utilisé pour estimer les observations d'un nouveau jeu de données : on dit qu'il sur-ajuste le jeu de données pour lequel il a été calibré.

La procédure suivie pour choisir la meilleure combinaison est schématisée figure 3.16.

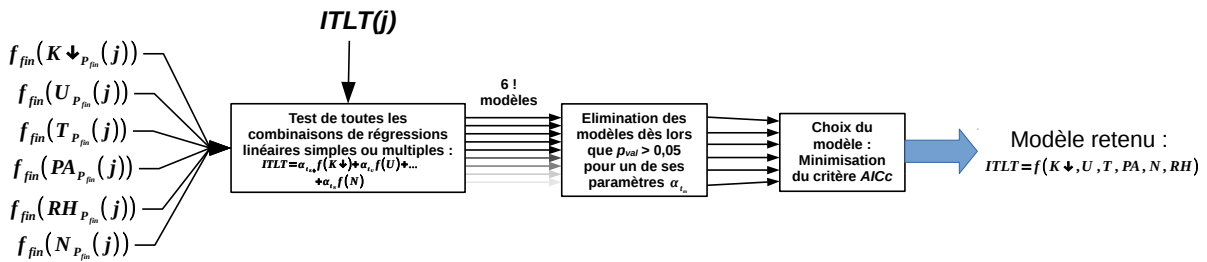


FIGURE 3.16 – Procédure utilisée pour déterminer la combinaison des termes  $x_m$  qui explique le mieux les variations de  $ITLT(j)$

Pour estimer  $ITLT(j)$ , le modèle linéaire proposé comportera de un à six termes (plus un terme constant) correspondant chacun à une variable météorologique. Toutes les combinaisons de variables sont testées et la meilleure est celle qui minimisera le Critère d'Information corrigé d' Akaike (AICc). Le critère d' Akaike (AIC) permet d' éviter le sur-ajustement en faisant un compromis entre les performances du modèle et le nombre de variables utilisées. Il est défini par la relation :

$$AIC = -2\ln(\max(f_L)) + 2nb_{x_n} \tag{3.17}$$

avec  $AIC$  le critère d'information d' Akaike,  $\max(f_L)$  le maximum de la fonction de vraisemblance  $f_L$  et  $nb_{x_n}$  le nombre de variables explicatives de la régression linéaire

Sa version corrigée, notée  $AICc$ , est utilisée dans le cas où l'échantillon de données est de taille faible. Dans ce cas, le sur-ajustement est plus fréquent. La correction consiste donc à pénaliser les modèles qui utilisent un nombre de termes important en augmentant la valeur du



critère AICc :

$$AIC_c = AIC + \frac{2nb_{x_n}(nb_{x_n} + 1)}{nb_{ech} - nb_{x_n} - 1} \quad (3.18)$$

avec  $AIC_c$  le critère d'information d'Akaike corrigé et  $nb_{ech}$  le nombre d'échantillons utilisés pour la régression linéaire

Une autre contrainte est finalement ajoutée au modèle à choisir : le coefficient de régression  $\alpha_{t_m}$  de tous ses termes  $\alpha_{t_m}x_m$  doit être significatif. Cette contrainte impose de respecter l'inégalité suivante :

$$p_{val} < 0.05 \quad (3.19)$$

avec  $p_{val}$  la valeur p représentant la probabilité que l'hypothèse de non-significativité du coefficient d'une relation linéaire soit vraie.

### 3.3 Modélisation spatiale par régression linéaire multiple

$ITLS_{moy_i}$  et  $ITLS_{std_i}$  sont modélisés pour les périodes  $P_{ICU}^{(jour)}$  et  $P_{ICU}^{(nuit)}$  par régression linéaire multiple.

$$\begin{aligned} ITLS_{moy} &= \alpha_{sm_0} + \sum_g \alpha_{sm_g} \Delta x_g \\ ITLS_{std} &= \alpha_{ss_0} + \sum_g \alpha_{ss_g} x_g \end{aligned} \quad (3.20)$$

avec  $ITLS_{moy}$  et  $ITLS_{std}$  les indicateurs de température locale qui correspondent aux variables à expliquer,  $\alpha_{sm_0}$  et  $\alpha_{ss_0}$  deux constantes de régression,  $\alpha_{sm_g}$  et  $\alpha_{ss_g}$  les coefficients de régression qui correspondent aux variables explicatives  $x_g$ .

Les variables explicatives utilisées sont formées à partir d'indicateurs géographiques  $IG_g$ , caractérisant chaque site de mesure. L'objectif de cette section est d'identifier et de calculer les indicateurs géographiques les plus pertinents à utiliser puis de déterminer la valeur des coefficients  $\alpha_{sm_0}$ ,  $\alpha_{ss_0}$ ,  $\alpha_{sm_g}$  et  $\alpha_{ss_g}$  des équations 3.20. Une première partie permettra d'identifier les indicateurs géographiques qui sont a priori les plus intéressants pour notre application. La deuxième partie décrira la méthode de calcul de ces indicateurs ainsi que les sources de données utilisées. Une dernière partie sera enfin consacrée à l'identification de la combinaison

de paramètres géographiques la plus pertinente pour modéliser  $ITLS_{moy}$  et  $ITLS_{std}$  pour les périodes diurnes et nocturnes.

### 3.3.1 Identification d'indicateurs géographiques pertinents

Pour rappel, selon la définition de Maby (2003), les indicateurs géographiques sont répartis en deux catégories.

- Les indicateurs thématiques : ils correspondent à une information moyennée à l'échelle d'un espace donné. Dans le cadre d'études microclimatiques, ils sont souvent divisés en deux sous-catégories : les indicateurs qui décrivent la morphologie (des objets ancrés au sol mais aussi du sol lui-même) et ceux qui décrivent l'occupation des sols (par leur nature et par leur type d'occupation).
- Les indicateurs spatiaux : ils sont représentatifs de la distance de l'objet étudié à un autre objet géographique.

La figure 1.15 présentée au chapitre 1 dressait une liste non-exhaustive des indicateurs géographiques utilisés dans les études de RG. Ceux utilisés pour ce travail sont catégorisés *indicateurs thématiques* ou *indicateurs spatiaux* (respectivement section 3.3.1.1 et 3.3.1.2). Les bases de données utilisées ainsi que les méthodes de calcul sont décrites 3.3.2.

#### 3.3.1.1 Indicateurs thématiques

Les indicateurs thématiques sont moyennés à l'échelle d'une unité spatiale de référence qui doit être définie. Sa taille est loin de faire l'unanimité dans la littérature, mais sa forme est très souvent un cercle centré sur la station où est mesurée la température. Le calcul de nos indicateurs géographiques thématiques est réalisé au sein de cercles de rayon 500 mètres, rayon choisi comme le compromis entre les valeurs observées les plus élevées (1000 m - Suomi et al. (2012)) et les valeurs les plus basses (20 m - Petralli et al. (2014)). La surface de ce cercle sera par la suite appelée surface de référence ( $A_{ref}$ ). Pour rappel, lorsque la variante de *tri directionnel* est utilisée, la surface de référence devient environ un sixième de ce cercle (en réalité légèrement plus - cf. figure 3.8), le sixième choisi étant différent selon la direction du vent.

Les surfaces urbaines élémentaires qui seront utilisées dans cette section sont représentées figure 3.17. Elles serviront à calculer certains indicateurs géographiques dont nous aurons besoin pour notre travail.

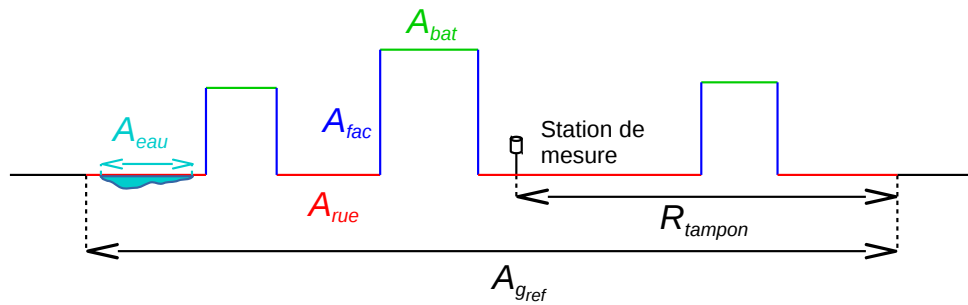


FIGURE 3.17 – Surfaces urbaines élémentaires utilisées pour calculer les indicateurs géographiques

Les indicateurs géographiques thématiques proposés sont ceux qui, de notre point de vue, ont le plus de lien avec les flux de chaleur échangés à l'échelle du quartier. Les interactions qui existent entre chaque flux contribuant au bilan d'énergie (équation 1.1) et ses indicateurs géographiques sont décrites ci-dessous.

- $Q^*$  : le rayonnement global net reçu par un quartier dépend du rayonnement global descendant et de la morphologie du quartier. En effet, les rues sont des espaces de piégeage radiatif : le jour, le rayonnement solaire va subir des multiples réflexions sur les façades de bâtiments ainsi que sur le sol des rues. De l'énergie sera absorbée à chaque réflexion, réduisant considérablement l'énergie réfléchie après multi-réflexion. Une bonne estimation de l'énergie solaire absorbée  $K^*$  par un quartier peut être réalisée en considérant les propriétés radiatives des matériaux comme homogènes à l'ensemble des surfaces. Bernabé et al. (2015) montrent que  $K^*$  est dans ce cas proportionnel à un terme représentatif des propriétés radiatives des matériaux, au rayonnement solaire descendant ( $K \downarrow$ ) et au facteur de vue du ciel ( $SVF$ ). Ils établissent également une relation géométrique entre le  $SVF$  et la densité de façade  $D_{fac}$  ( $D_{fac}$  telle que définie tableau 3.1) :  $SVF = 1 - D_{fac}$ . La nuit, les surfaces situées dans la canopée ont un facteur de vue du ciel réduit du fait de la présence des bâtiments. L'évacuation de chaleur par rayonnement GLO est donc limitée. De jour comme de nuit, l'indicateur géographique  $D_{fac}$  représente donc bien les différences de comportement radiatif qui peuvent exister entre deux quartiers morphologiquement différents.
- $\Delta Q_S$  : terme lié au stockage de l'énergie dans les sols urbains, dans les façades et dans les toits des bâtiments. Cette énergie est stockée en journée et restituée la nuit par convection avec l'air situé dans la canopée ou évacuée par rayonnement GLO. Ce terme dépend de

la quantité d'énergie absorbée  $Q^*$  mais également de l'effusivité thermique équivalente d'un quartier. L'effusivité thermique d'un matériau, proportionnelle à la racine carrée de sa capacité thermique volumique, est caractéristique de l'augmentation de sa température lorsqu'il accumule une quantité d'énergie donnée. Un matériau à effusivité élevée est associé à une augmentation faible de la température pour une quantité d'énergie reçue importante. La présence de ce type de matériau est propice au développement de l'ICU. La capacité thermique volumique équivalente d'un quartier entièrement constitué d'un même matériau est directement proportionnelle à la quantité de surfaces urbaines en contact avec l'air. La densité de surfaces urbaines  $D_{surf}$  telle que définie tableau 3.1 est donc utilisée comme étant représentative du terme  $\Delta Q_S$ .

- $Q_H$  et  $Q_E$  sont respectivement les flux de chaleur sensible et latente qui correspondent aux échanges de chaleur et d'humidité entre les surfaces urbaines et l'air. Le flux total d'énergie échangé entre la surface et l'air est la somme de ces deux termes. Or, une surface qui produit un flux de chaleur latente en proportion importante est souvent associée à une surface couverte de végétation. Pour un même flux de chaleur, une surface de végétation va donc moins réchauffer l'air qu'une autre surface. Afin de déterminer les surfaces couvertes de végétation, des images satellitaires multi-spectrales sont utilisées. Chaque pixel de l'image représente un carré de territoire dont la taille dépend de la résolution  $r_{sat}$  du satellite utilisé (une résolution de cinq mètres permet par exemple d'associer un pixel à un carré de cinq mètres de côté). L'intensité du rayonnement solaire réfléchi par le sol terrestre est enregistrée pour un nombre de plages de longueurs d'ondes du spectre solaire propre à chaque satellite. Chacune de ces plages est appelée canal et la valeur de chaque canal est enregistrée dans chacun des pixels de l'image. C'est cette information qui permet de discriminer les matériaux entre eux (voir illustration figure 3.18). Les couverts végétaux ont par exemple des valeurs de réflectance plus élevées dans le canal du proche infra-rouge ( $RE_{pir}$ ) que dans celui du rouge ( $RE_{rou}$ ). C'est le contraire pour tous les autres matériaux sauf les sols nus qui ont une réflectance aussi importante pour les deux canaux. Plusieurs indicateurs tirent parti de cette caractéristique pour identifier la présence d'une végétation plus ou moins dense dans un pixel de l'image. L'indicateur le plus couramment utilisé est l'indice de végétation par différence normalisé (NDVI) ou indice de Tucker (Rouse Jr et al. 1974 ; Tucker 1979). Il est calculé en divisant la différence des réflectances enregistrées entre les canaux du proche infra-rouge et du rouge par leur somme. Son application permet souvent de classifier les surfaces : les valeurs de NDVI

supérieures à 0,1 sont associées à des couverts végétaux, les valeurs proches de zéro à des sols nus et les valeurs négatives à d'autres sols. Cependant, cette classification n'a que très peu d'intérêt dans le cadre de notre application car elle entraîne une perte d'information : avec cette technique, les couverts végétaux les plus denses sont confondus avec les moins denses. L'information continue (les valeurs de NDVI) sera donc directement utilisée et associée au flux  $Q_E$ . Cette méthode est originale en climatologie urbaine mais a déjà été utilisée pour estimer le flux de chaleur latente émis par des couverts végétaux : à rayonnement solaire donné, une relation quasi linéaire entre l'évapotranspiration et le NDVI est proposée par Pelgrum et Bastiaanssen (1996). Le NDVI ne considère néanmoins pas l'évaporation qui peut avoir lieu entre une surface d'eau (lac ou rivière) et l'atmosphère. La valeur pour une surface d'eau est proche de celle d'un béton (valeur très faible) alors qu'ils n'ont pas du tout le même potentiel d'évaporation. La densité de surface d'eau  $D_{eau}$  est donc utilisée en complément du NDVI pour représenter les termes  $Q_H$  et  $Q_E$ .

- $Q_F$  représente les émissions de chaleur anthropique générées par l'activité humaine. En ville, ces flux sont principalement assimilés à l'énergie consommée dans les bâtiments (chauffage et pertes par effet Joules) et à l'énergie consommée par le transport. Les émissions dues au transport sont difficilement modélisables (à l'échelle fine) à partir d'informations géographiques simples qui soient disponibles pour l'ensemble du territoire. Par contre, celles dues aux bâtiments peuvent être simplement reliées à la quantité de surfaces en contact avec l'air extérieur (façades et toits). L'indicateur géographique associé à  $Q_F$  peut donc être calculé selon :

$$\frac{A_{bat} + A_{fac}}{A_{gref}} = \frac{A_{bat}}{A_{gref}} + \frac{A_{fac}}{A_{gref}} = D_{bat} + \frac{A_{surf} - A_{gref}}{A_{gref}} = D_{bat} + D_{surf} - 1 \quad (3.21)$$

Cet indicateur correspond donc à une combinaison linéaire des indicateurs  $D_{surf}$  et  $D_{bat}$ . L'indicateur  $D_{surf}$  étant déjà associé à un terme de notre bilan de chaleur, seul l'indicateur  $D_{bat}$  sera finalement ajouté à l'analyse.

Certains flux de chaleur auront probablement une influence plus importante que d'autres sur le développement de l'ICU selon les saisons. Le flux de chaleur d'origine anthropique pourra par exemple avoir une importance prépondérante en hiver mais être négligeable en été. Rattacher un indicateur géographique à un flux permettra ainsi d'interpréter avec une approche plus

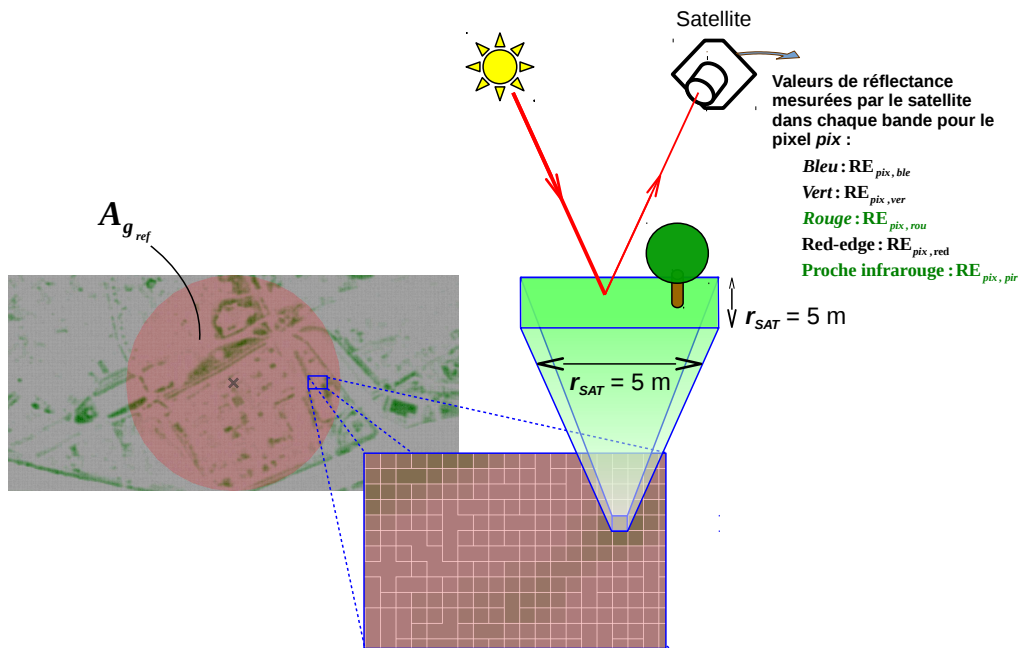


FIGURE 3.18 – Représentation schématique de l’acquisition d’une image satellitaire multi-spectrale à partir d’un satellite de résolution  $r_{sat} = 5m$  enregistrant cinq canaux. Les canaux du rouge  $RE_{pix,rou}$  et du proche infrarouge  $RE_{pix,pir}$  sont utilisés pour distinguer le potentiel évaporatif des surfaces.

physique le fait qu’un indicateur géographique soit prépondérant dans l’explication de l’ICU à une saison donnée.

Le tableau 3.1 décrit chacun des indicateurs thématiques et spécifie les données utilisées pour leur calcul tandis que la figure 3.19 synthétise les interactions entre flux de chaleur et indicateurs géographiques thématiques.

Sigle	Nom	Description	Formule	Type de données utilisée pour le calcul
$D_{bat}$	Densité de bâtiments	Rapport de la surface au sol totale couverte par des bâtiments ( $A_{bat}$ ) et de la surface de référence ( $A_{gref}$ )	$\frac{A_{bat}}{A_{gref}}$	BDD Topographique de l'IGN (version 2011)
$NDVI$	Indice de végétation par différences normalisées (NDVI)	Indice de végétation défini comme le NDVI moyen de tous les pixels $NDVI_{pix}$ contenus dans $A_{gref}$ pondérée par leur surface $A_{pix}$ - on notera $RE_{pix,rou}$ la valeur de réflectance associée au pixel $pix$ dans la bande spectrale rouge et $RE_{pix,pir}$ la valeur dans le proche infra-rouge	$\frac{\sum_{pix} \frac{RE_{pix,pir} - RE_{pix,rou}}{RE_{pix,pir} + RE_{pix,rou}} A_{pix}}{\sum_{pix} A_{pix}}$	Image multispectrale RAPIDEYE (résolution : 5 m - acquisition : mars 2011)
$D_{fac}$	Densité de façade	Rapport des surfaces de façades $A_{fac}$ par la somme des surfaces de façades et de la surface de référence	$\frac{A_{fac}}{A_{fac} + A_{gref}}$	BDD Topographique de l'IGN (version 2011)
$D_{surf}$	Densité de surface totale	Rapport entre la somme des surfaces urbaines en contact avec l'air et la surface de référence	$\frac{A_{fac} + A_{gref}}{A_{fac}}$	BDD Topographique de l'IGN (version 2011)
$D_{eau}$	Densité de surface en eau	Rapport entre les surfaces recouvertes d'eau $A_{eau}$ et la surface de référence	$\frac{A_{eau}}{A_{gref}}$	BDD Topographique de l'IGN (version 2011)

TABLE 3.1 – Indicateurs géographiques thématiques et bases de données utilisées pour leur calcul

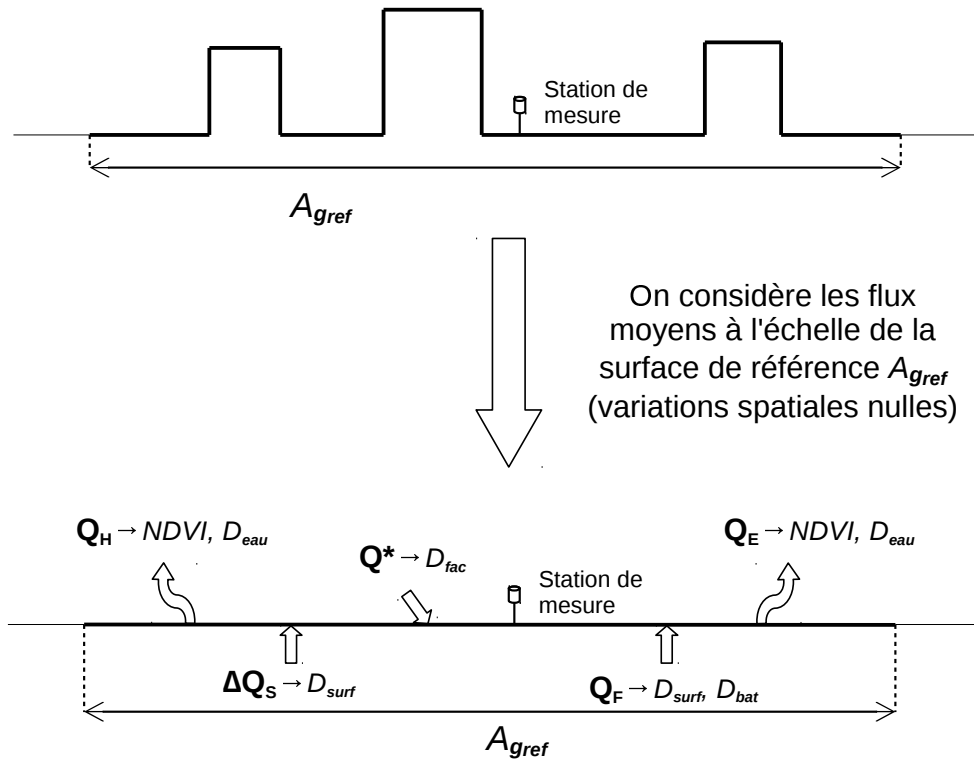


FIGURE 3.19 – Associations d'un ou de plusieurs indicateurs géographiques à chacun des termes du bilan d'énergie d'une surface urbaine

### 3.3.1.2 Indicateurs spatiaux

Un indicateur spatial est utilisé en complément des indicateurs thématiques présentés auparavant. Il s'agit du logarithme népérien de la distance séparant chaque station à la périphérie de la zone urbaine  $\ln(d_{periph} + 1)$ . L'ajout de la constante 1 permet d'avoir des valeurs exclusivement positives (la distance à la périphérie d'une station située en dehors de la zone urbaine étant considérée comme nulle). Cet indicateur est utilisé entre autres par Balázs et al. (2009) et donne de bons résultats. Dans leur cas, la zone d'étude est découpée en mailles carrées de 500 m de côté dans lesquelles la densité de bâtiments  $D_{bat}$  est calculée. La périphérie est atteinte lorsqu'une maille voit sa valeur  $D_{bat}$  descendre sous la valeur seuil de 0.05. Dans notre cas, la limite morphologique de la ville est calculée à partir de l'outil MorphoLim (Tannier et al. 2011).

La signification physique de cet indicateur est la suivante : le bâtiment est l'élément principal qui constitue la ville. Il est également l'élément de la ville qui altère localement le comportement radiatif. Or, le rayonnement est le mécanisme physique principal à l'origine des variations temporelles de température de l'air. Des différences de température peuvent appa-



raître entre deux zones de l'agglomération qui n'ont pas les mêmes caractéristiques morphologiques. La nuit, ces différences de température peuvent donner lieu au développement de vents faibles frais issus de la périphérie de la ville (souvent plus ouverte) vers le centre (quartiers les plus denses). En effet, les zones les plus ouvertes sont généralement les plus froides la nuit et sont connues pour rafraîchir l'air des quartiers environnants. Jauregui (1991) ainsi que Spronken-Smith et Oke (1998) ont par exemple montré que les parcs urbains étaient plus froids que les zones densément bâties et que cette fraîcheur pouvaient se propager à l'extérieur du parc. Spronken-Smith et Oke (1998) montrent que cette fraîcheur est d'autant plus intense lorsque le parc est dénué d'arbres et que les sols ont une effusivité thermique faible (particulièrement vrai en période de sécheresse). Ils expliquent également le déplacement de l'air frais vers l'extérieur du parc généré par des forces de flottabilité : en l'absence de vent, les différences de température et donc de masse volumique générées par le refroidissement de l'air du parc entraînent un mouvement de convection naturelle entre le parc et l'extérieur. Pour toutes ces raisons, il paraît nécessaire d'identifier les zones du territoire dont le facteur de vue du ciel est le plus fort (souvent les parcs, les étendues d'eau ou les zones périphériques de la ville). Ces zones ont en effet une influence sur la mesure de température des stations situées à proximité. Plus la station sera proche de cette zone, plus cet effet sera important. Ce comportement n'est pas pris en compte par les indicateurs thématiques qui moyennent les informations au sein de la surface de référence  $A_{gref}$ . L'indicateur  $\ln(d_{periph} + 1)$  permet donc de remédier à ce manque.

### 3.3.2 Procédure et données utilisées pour le calcul des indicateurs géographiques

Toutes les données géographiques utilisées datent de 2011, ce qui est cohérent avec les périodes d'acquisition des données météorologiques utilisées pour la calibration (2010 - 2014). Certaines données météorologiques utilisées pour la vérification sont néanmoins enregistrées en 2015. On fera l'hypothèse qu'en termes de morphologie et de typologie des sols, les modifications subies entre 2011 et 2015 par les zones d'étude sont mineures et qu'elles n'affecteront pas nos résultats.

#### 3.3.2.1 Indicateurs thématiques

La totalité des indicateurs thématiques est calculée à partir de la BD Topo (sauf le NDVI). L'avantage de cette base de données par rapport à celles produites par les villes ou les agglomérations réside dans la disponibilité et l'homogénéité de ses données à l'échelle du territoire

français. Les objets utilisés sont les bâtiments et les surfaces en eau, chaque type correspondant à une base de données du même nom. Les coordonnées  $x$  et  $y$  des polygones correspondant aux empreintes au sol des bâtiments ainsi qu'à celles des surfaces d'eau sont stockées sous un champ géométrique. D'après l'IGN, «*la hauteur du bâtiment [correspond] à la différence entre le  $z$  le plus élevé du pourtour du bâtiment et un point situé au pied du bâtiment. La hauteur est arrondie au mètre*». Cette information est renseignée dans un autre champ que le champ géométrique. Si elle n'est pas renseignée, une valeur par défaut de trois mètres est utilisée. D'autres champs définissant le type de bâtiments ou de surfaces en eau sont également disponibles mais ils ne seront pas utilisés ici. Le calcul des indicateurs décrits tableau 3.1 est réalisé sous le logiciel libre OrbisGIS (Bocher et Petit 2012). Les opérations suivantes sont effectuées.

1. Une zone tampon concentrique de rayon 500 m est générée autour de chaque station de mesure (voir le cercle représenté par la surface  $A_{ref}$  figure 3.18) : les coordonnées de chaque cercle (représenté par un polygone) sont ainsi enregistrées dans le champ géométrique de la base de données nommée *Sites*.
2. L'intersection des objets des bases de données *Sites* avec ceux des bases de données *Bâtiments* et *Surfaces en eau* est réalisée. Lorsqu'un bâtiment ou une surface d'eau n'est pas entièrement contenue dans  $A_{ref}$ , seule la partie située à l'intérieur du cercle est conservée. Dans le cas d'un bâtiment, la nouvelle limite issue du découpage est assimilée à une façade. L'erreur de calcul du périmètre des bâtiments qui découle de cette approximation est négligeable devant la somme des périmètres de tous les bâtiments du site. Il est intéressant d'observer que dans le cas de zones tampons de taille plus faible, cette simplification ne serait pas valable.
3. La somme de toutes les surfaces de bâtiments est calculée pour chaque site et enregistrée sous l'indicateur  $A_{bat}$ . Le même calcul est mené pour les surfaces d'eau  $A_{eau}$ .
4. La surface totale de façades  $A_{fac}$  est calculée en deux étapes. La première consiste à additionner toutes les surfaces de façades, en considérant dans un premier temps les murs partagés entre deux bâtiments comme des façades. Pour cela, le périmètre de chaque bâtiment est simplement multiplié par sa hauteur. La seconde étape consiste à soustraire au résultat précédant les surfaces de façades partagées entre deux bâtiments accolés, soit la surface de façade du bâtiment le plus bas.

Les données satellitaires respectent également les conditions de disponibilité et d'ho-

mogénéité à l'échelle du territoire français. Elles sont issues d'images RAPIDEYE, disponibles au téléchargement sur le portail GEOSUD<sup>1</sup>. Elles sont fournies gratuitement pour les adhérents par dalles de 20 kilomètres de côté à une résolution de cinq mètres. Les photos ont été prises en mai 2011 pour chacune des trois zones d'études, pendant des périodes dénuées de nuages. Afin de pouvoir comparer les images entre elles, il est nécessaire de convertir les comptes numériques associés à chaque pixel (intensité dans chaque canal) en réflectance exo-atmosphérique. La réflectance correspond à l'énergie réfléchiée par la Terre. L'atmosphère absorbe et diffuse les rayons solaires lors de leur parcours dans l'atmosphère et modifie les valeurs de réflectance reçues par le satellite. Ces modifications sont plus ou moins importantes selon :

- la concentration d'ozone et de vapeur d'eau dans l'atmosphère ;
- le type d'aérosols et leur concentration respective.

Ces paramètres varient spatialement et temporellement, rendant incomparables deux photos prises à deux instants ou à deux endroits différents. Ceci pose problème puisque l'objectif de cette thèse est de pouvoir extrapoler les modèles proposés. Les données géographiques d'entrées doivent donc être homogènes entre nos zones d'études. Il est heureusement possible de calculer l'ordre de grandeur des modifications atmosphériques de la réflectance afin de corriger les données enregistrées. La réflectance obtenue après ces corrections est qualifiée d'exo-atmosphérique du fait d'avoir supprimé l'influence de l'atmosphère sur le rayonnement. L'image finale est construite en calculant le NDVI de chaque pixel  $pix$  à partir de l'équation 3.22, en utilisant les valeurs de réflectance du canal *proche infra-rouge*  $RE_{pix,pir}$  et de celles du canal *rouge*  $RE_{pix,r}$ .

$$NDVI_{pix} = \frac{RE_{pix,pir} - RE_{pix,r}}{RE_{pix,pir} + RE_{pix,r}} \quad (3.22)$$

La procédure de calcul du NDVI moyen au sein de la zone tampon (surface de référence  $A_{gref}$ ) de chaque site est rapide à mettre en place mais coûteuse en temps de calcul. Elle est constituée de trois étapes.

1. Étape 1 : l'image raster est transformée en base de données vectorielle. Cela consiste à considérer chaque pixel comme un polygone (champ géométrique) auquel est associée la valeur du champ NDVI. Cette étape est réalisée sous le logiciel libre QGIS, après avoir préalablement actualisé les valeurs de NDVI en appliquant la relation 3.23 et en ne

<sup>1</sup> disponible sur la page <http://ids.equipe-geosud.fr/> [Accédé le 08/07/2016]

conservant que la partie entière du résultat.

$$NDVI_{fin} = 100(NDVI + 1) \quad (3.23)$$

En effet, seuls les entiers naturels sont traités par la fonction de polygonisation de QGIS. Les valeurs de NDVI sont finalement distribuées entre 0 et 200. Ces valeurs étant plus commodes à traiter par la suite, elles seront conservées. Afin de réduire le nombre de polygones de la base de données nouvellement créée, QGIS agrège automatiquement les polygones adjacents qui ont la même valeur de NDVI. Cette nouvelle table est nommée *NDVI*.

2. Étape 2 : la table *NDVI* est ensuite intersectée avec le cercle qui correspond à chaque site de mesure. Cette opération entraîne le scindage de quelques polygones de la table *NDVI*. Leur surface est donc modifiée mais leur valeur est cependant conservée.
3. Étape 3 : le NDVI moyen à l'échelle de chaque zone tampon est calculé en pondérant le NDVI de chaque pixel par sa surface (cf tableau 3.1).

### 3.3.2.2 Indicateurs spatiaux

Le logiciel MorphoLim calcule les limites d'une zone urbaine à partir de l'empreinte au sol des bâtiments. Les données utilisées sont donc celles de la BD Topo. Une méthode de dilatation est utilisée : un nombre  $nb_{bat}$  de bâtiments (ou agrégats de bâtiments) isolés est initialement décompté sur le territoire. Les empreintes de ces bâtiments sont élargies d'une certaine distance  $d_{dil_0}$ , puis agrégées dès lors qu'elles s'intersectent. Le nombre d'agrégats après cette manipulation est noté  $nb_{bat_1}$ , tel que  $nb_{bat_1} \leq nb_{bat_0}$ . Les bâtiments sont ensuite élargis d'une distance  $d_{dil_1} > d_{dil_0}$  puis agrégés, et ainsi de suite jusqu'à ce qu'un seul agrégat soit identifiable. L'hypothèse est faite que la partie urbaine de l'agglomération est bâtie de manière fractale. Dans ce cas, l'approche proposée devrait mettre en évidence que le nombre d'agrégats  $nb_{bat}$  diminue sous la forme :

$$nb_{bat} = \chi_0 d_{dil}^{\chi_1} \quad (3.24)$$

avec  $\chi_0$  et  $\chi_1$  deux constantes propres à la structure de la ville étudiée et  $d_{dil}$  la distance de dilatation des bâtiments (en m)

Cette équation peut être linéarisée en calculant le logarithme :

$$\log(nb_{bat}) = \chi_1 \cdot \log(d_{dil}) + \log(\chi_0) \quad (3.25)$$

Lorsque la méthode de dilatation est appliquée à une agglomération, tant que le nuage de points suit cette loi, l'agrégat principal de bâtiments ne comprend que des surfaces qualifiées d'urbaines. La distance critique  $d_{dil,seuil}$  correspond à la distance pour laquelle le nuage de points s'éloigne de la tendance théorique régie par l'équation 3.25. Pour cette distance, les surfaces couvertes par les bâtiments dilatés sont considérées comme appartenant à la zone urbaine. La limite de la ville est donc définie comme la frontière entre les surfaces correspondant à des bâtiments dilatés et les surfaces vierges de tout bâtiment. Cette méthode est appliquée aux agglomérations de Nantes, Angers et La-Roche-sur-Yon. Les distances critiques obtenues sont respectivement 56, 71 et 81 m. La distinction entre zone urbanisée et zone non-urbanisée de l'agglomération de Nantes est présentée figure 3.20.

L'image laisse apparaître certaines zones du centre-ville comme zones non urbaines (espaces ouverts). Ces surfaces correspondent souvent à des parcs, des cimetières ou des étendues d'eau mais elles peuvent également correspondre à des rues très larges ou à des places minéralisées.

La distance  $d_{dil,seuil}$  est très différente entre nos trois agglomérations. Ce résultat peut s'avérer cohérent avec les observations faites par Tannier et al. (2011) lorsqu'ils appliquent leur méthode à différentes villes européennes.

- La valeur  $d_{dil,seuil}$  augmente lorsque la différence d'espacement entre les bâtiments situés à l'intérieur et à l'extérieur de la ville diminue.
- Elle augmente également lorsque les tissus urbains sont très peu connectés entre eux spatialement, c'est-à-dire lorsque des tâches hétérogènes sont identifiables.

Les tissus nantais sont peut-être mieux connectés entre eux et les différences d'espacement entre bâtiments diffèrent peut-être plus largement entre la ville et sa périphérie que pour les agglomérations d'Angers et de La Roche-sur-Yon. L'indicateur de distance à la périphérie  $\ln(d_{periph} + 1)$  est en tous cas calculé à partir de la frontière déterminée par cette méthode (illustration de la distance  $d_{periph}$  figure 3.20).

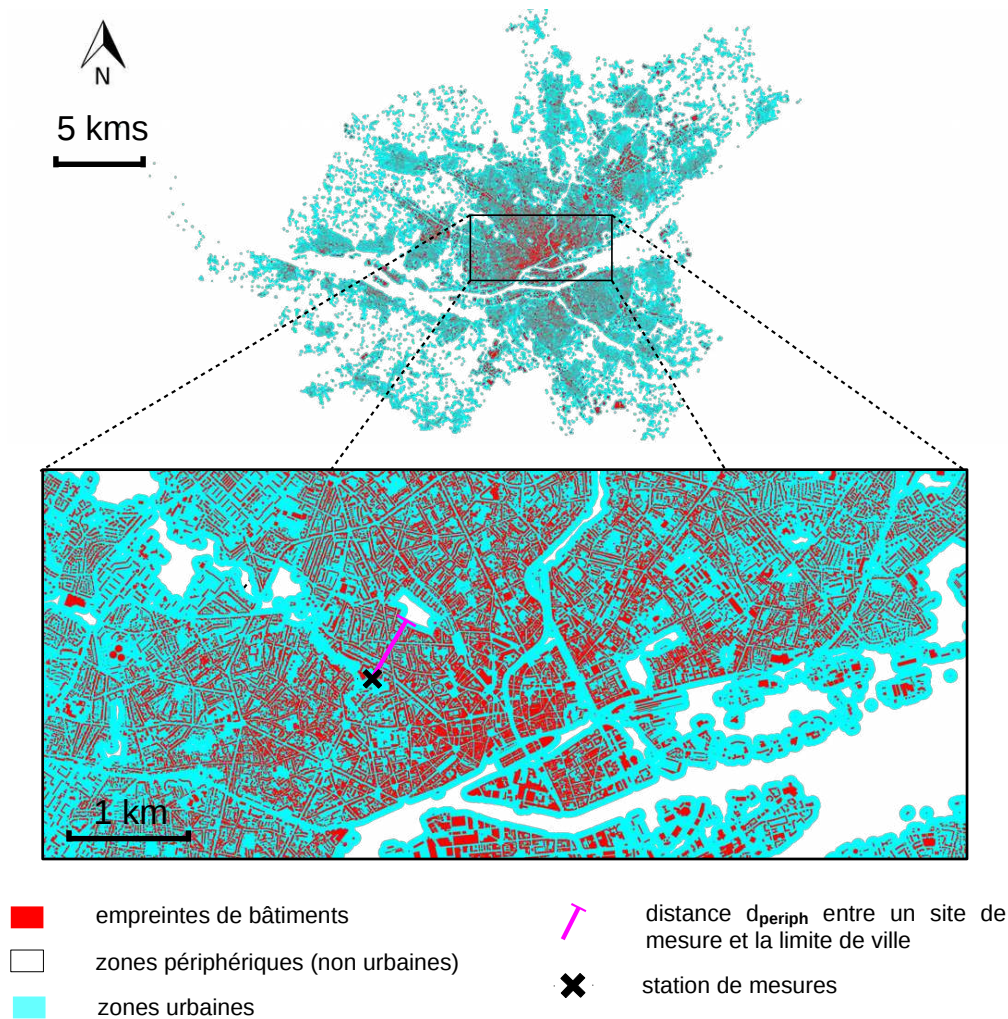


FIGURE 3.20 – Différenciation des surfaces considérées comme urbaines de celles considérées comme non-urbaines pour le centre-ville de Nantes. Les limites de zone urbaine sont calculées à partir du logiciel MorphoLim et de l’empreinte au sol des bâtiments de la BD Topographique de l’IGN.

### 3.3.3 Sites d’observation du microclimat urbain : quelle représentativité géographique des zones d’étude ?

Un des objectifs de cette thèse est de construire un EM qui permettent d’estimer l’ICU de n’importe quel point de la zone d’étude. Il est donc nécessaire que les sites utilisés pour vérifier les performances des EM soient géographiquement représentatifs de l’ensemble de la zone d’étude. Il est préférable que ceux utilisés pour la calibration des modèles spatiaux le soient également. Pour vérifier que c’est bien le cas, les six indicateurs géographiques décrits auparavant ( $NDVI$ ,  $D_{\text{eau}}$ ,  $D_{\text{bat}}$ ,  $d_{\text{periph}}$ ,  $D_{\text{surf}}$  et  $D_{\text{fac}}$ ) sont calculés pour 200 sites tirés aléatoirement sur le territoire de chaque zone d’étude. Un site est considéré comme invalide lorsqu’il est situé dans

l'eau ou sur un bâtiment. Si c'est le cas, on choisit point valide le plus proche. Les indicateurs géographiques sont calculés au sein de cercles (de 500 m de rayon) centrés sur chacun de ces 200 points. Cet échantillon est supposé représentatif de la variabilité géographique du territoire étudié. Le travail est réalisé pour les trois zones d'étude (les agglomérations de Nantes, Angers et LRY). La répartition spatiale de cet échantillon pour l'agglomération de Nantes est présentée figure 3.21.

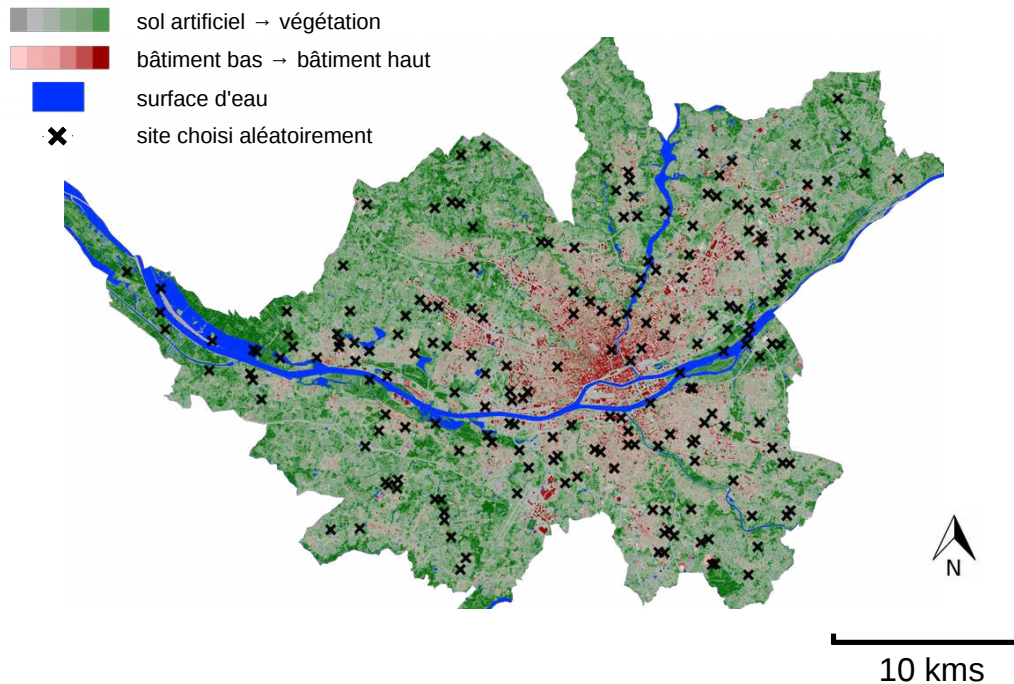


FIGURE 3.21 – Répartition des 200 points tirés aléatoirement sur le territoire nantais

Pour chaque indicateur géographique, la distribution des valeurs de chaque jeu de données est confrontée à celle des valeurs de l'échantillon représentatif (figure 3.22). Pour plus de lisibilité, seule la distribution de l'échantillon de 200 points est représentée sous forme d'un histogramme. Chaque jeu de données comporte deux distributions. La première, qualifiée de *sans direction*, représente un point par site de mesure. La seconde, nommée *direction*, représente un point pour un site et une direction  $dir_{pr}$  donnée (dans ce cas, la variante de *tri directionnel* a été utilisée).

Les résultats pour la zone d'étude de Nantes sont représentés figure 3.22a. Pour les trois indicateurs morphologiques ( $D_{bat}$ ,  $D_{surf}$  et  $D_{fac}$ ), tous les jeux de données sont représentatifs de la zone d'étude excepté  $JDD_{spat}$ . En effet, les sites du jeu de données  $JDD_{spat}$  sont tous situés dans un même quartier (le quartier du *Pin Sec*), relativement homogène en termes de morphologie. Pour l'indicateur géographique spatial  $d_{periph}$ , aucun jeu de donnée n'est re-

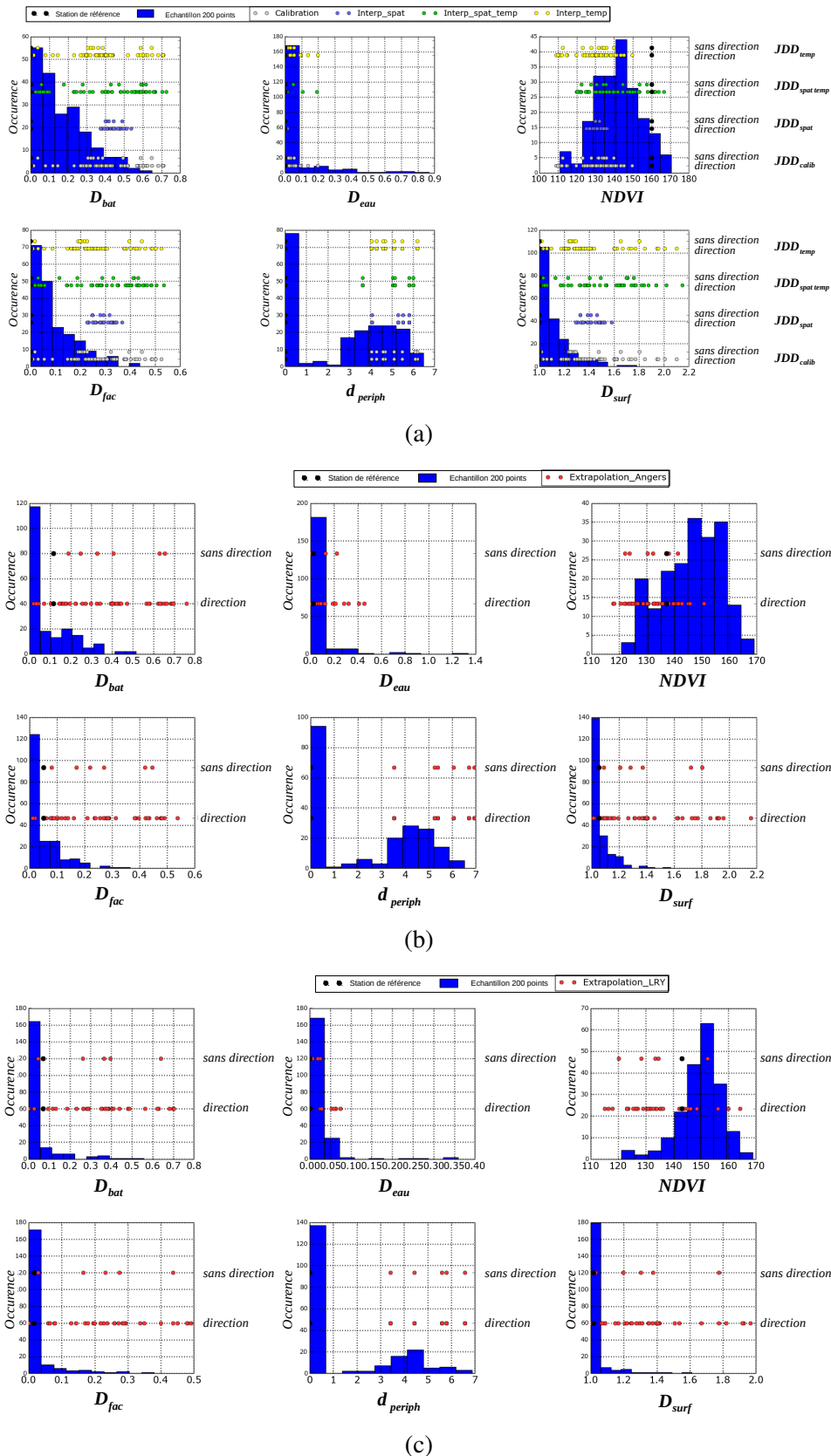


FIGURE 3.22 – Distribution des valeurs de chaque indicateur géographique  $IG_g$  calculées pour 200 points aléatoirement choisis sur la zone d'étude et pour les stations de chaque jeu de données utilisé. (a) agglomération de Nantes, (b) agglomération d'Angers ( $JDD_{extrapol}^{(Angers)}$ ), (c) agglomération de LRY ( $JDD_{extrapol}^{(LRY)}$ ). Cette figure permet de vérifier que chaque jeu de données est bien géographiquement représentatif (au vu des indicateurs géographiques calculés) de la zone d'étude dans laquelle il a été produit



présentatif de l'intervalle  $[0, 5 - 3, 5]$ . Cet intervalle concerne moins de 10% du territoire et est compris entre deux intervalles bien représentés ( $[0 - 0, 5]$  et  $[3, 5 - 6, 5]$ ). Si la condition de linéarité entre les indicateurs géographiques et les indicateurs *ITLS* est bien respectée, cet intervalle ne devrait pas poser de problème. En revanche, aucun jeu de données ne représente bien les valeurs extrêmes hautes de densité de surface d'eau puisqu'aucun site des quatre jeux de données nantais n'est couvert d'eau à plus de 20% (tandis que certains sites de l'échantillon aléatoire atteignent les 80%). De même, pour le *NDVI*, seul le jeu de données  $JDD_{spat\&temp}$  a plus d'un site ayant un *NDVI* supérieur à 150. Les autres jeux de données (dont celui de calibration) n'ont que la station de référence ayant un *NDVI* au-dessus de 150. Les mêmes observations peuvent être réalisées pour les indicateurs géographiques morphologiques et spatiaux des zones d'études d'Angers et de La-Roche-sur-Yon (respectivement figures 3.22b et 3.22c). Concernant  $D_{eau}$ , même si aucune station angevine ne permet de représenter les valeurs les plus hautes de l'échantillon aléatoire, deux ont des valeurs au-dessus de 40%. Ceci correspond à plus du double des valeurs maximales observées pour les jeux de données nantais. Les valeurs de *NDVI* sont quant à elles peu représentatives des valeurs hautes de l'échantillon aléatoire. Ceci n'est pas le cas des stations yonnaises, où seule la représentativité de  $D_{eau}$  peut être mise en cause.

Pour conclure, si  $D_{eau}$  est utilisée dans les modèles, il se peut que ceux-ci ne soient pas très performants pour estimer la température d'air de sites contenant de grandes surfaces d'eau. Les jeux de vérification sont également dépourvus de sites à forte  $D_{eau}$  (excepté le jeu de données angevin). Ils permettront donc difficilement de mettre ce comportement en évidence. De plus, le jeu de données  $JDD_{calib}$  n'est pas représentatif de la variabilité du *NDVI* à l'échelle de l'agglomération nantaise. Le coefficient  $\alpha_{s_{NDVI}}$ , supposé représenter l'influence du *NDVI* sur l'ICU, sera donc potentiellement surestimé ou sous-estimé par la régression. Cette erreur pourra être vérifiée à partir des jeux de données de vérification  $JDD_{spat\&temp}$  et  $JDD_{extrapol}^{(Angers)}$ , qui ont tous plusieurs stations à fort *NDVI*.

### 3.3.4 Procédure d'identification du modèle spatial le plus pertinent

Les relations permettant de calculer les indicateurs  $ITLS_{moy}$  et  $ITLS_{std}$  ont été définis par les équations 3.20. Chaque variable explicative  $x_g$  est définie par la différence entre les valeurs de l'indicateur géographique  $IG_{g_i}$  de chaque station  $i$  par rapport à celle de la station de référence  $IG_{g_{ref}}$  :

$$\Delta x_g = IG_{g_i} - IG_{g_{ref}} \quad (3.26)$$

Afin que la manière de calculer les indicateurs géographiques  $IG_g$  soit cohérente avec celle utilisée pour le calcul des  $ITL$ , on soustrait à chaque station  $i$  et pour chaque indicateur  $IG_g$  la valeur de l'indicateur prise par la station de référence. De même que pour le modèle temporel, les relations entre  $ITL$  et  $IG_g$  ne sont pas toutes linéaires. La figure 3.23 illustre ce constat : la relation entre  $ITLS_{moy_i}$  et l'indicateur  $D_{surf}$  pourrait être assimilée à une fonction non-linéaire (quadratique ou exponentielle).

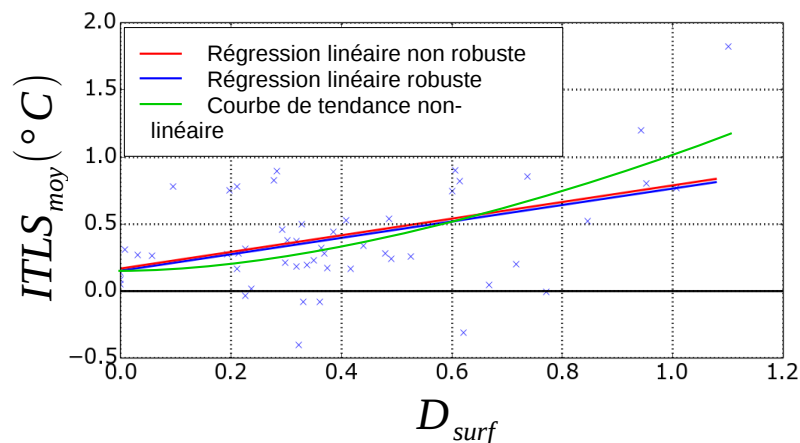


FIGURE 3.23 – Évolution de l'indicateur  $ITLS_{moy_i}$  en fonction de l'indicateur géographique  $D_{surf}$  pour la saison hivernale lorsque la variante *tri directionnel* est utilisée

Cependant, comme les indicateurs géographiques ont été calculés par rapport aux valeurs prises par la station de référence, il est impossible d'utiliser des fonctions de linéarisation. Celles-ci sont beaucoup trop dépendantes des valeurs prises par la station de référence. Leur utilisation engendreraient donc des erreurs d'estimation si les caractéristiques de la station de référence étaient différentes des caractéristiques de celle utilisée pour la calibration. Ceci est illustré par la figure 3.24, sur laquelle il apparaît clairement que le modèle linéaire est plus robuste à l'extrapolation que le modèle non-linéaire.

Seules des relations linéaires sont donc utilisées pour expliquer les valeurs de  $ITL$  à partir de celles prises par  $IG_g$  (équation 3.20).

Afin de trouver la combinaison de variables explicatives qui maximise les performances du modèle en évitant tout sur-ajustement, la procédure utilisée pour calibrer le modèle temporel est appliquée au modèle spatial. Celle-ci est illustrée figure 3.25.

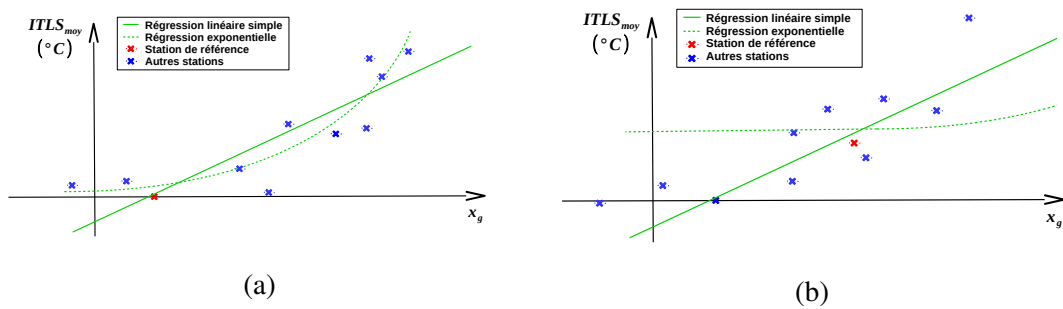


FIGURE 3.24 – Représentation schématique des phases (a) de calibration et (b) de vérification (à partir d'un nouveau jeu de données) d'un modèle linéaire et d'un modèle non-linéaire. Cette figure illustre les limites de l'extrapolation d'un modèle non-linéaire par rapport à un modèle linéaire lorsque les caractéristiques géographiques de la station de référence diffèrent de celles de la station utilisée pour la calibration.

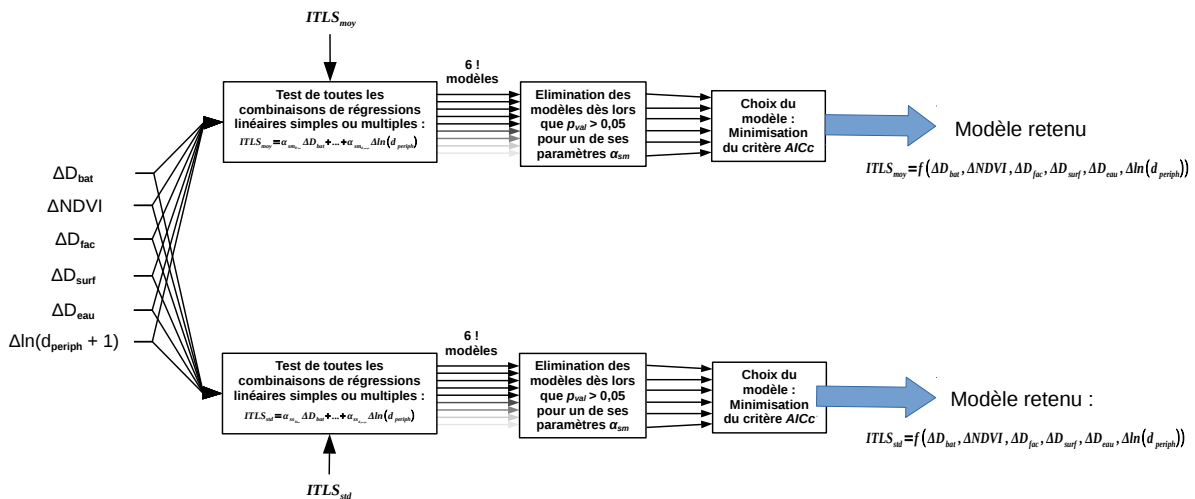


FIGURE 3.25 – Procédure utilisée pour déterminer la combinaison de termes  $x_g$  qui explique le mieux les variations des indicateurs  $ITLS_{moy}$  et  $ITLS_{std}$ .

### 3.4 Conclusion du chapitre

La méthode de modélisation de l'ICU a été décrite dans ce chapitre. Deux périodes du cycle diurne ont été identifiées comme intéressantes pour l'étude de l'ICU : l'une diurne ( $P_{ICU}^{(jour)}$ ), l'autre nocturne ( $P_{ICU}^{(nuit)}$ ). La différence de température enregistrée pendant ces périodes entre une station  $i$  et une station de référence diffère selon la journée considérée. Ces variations, qualifiées de variations temporelles de l'ICU, sont calculées à partir d'un indicateur noté  $ITLT$ .  $ITLT$  est modélisé à partir de régressions linéaires multiples avec comme variables explicatives les conditions météorologiques mesurées à la station Météo-France : vitesse du vent, humidité relative de l'air, température de l'air, rayonnement solaire descendant et pression atmosphérique. Toutes les stations urbaines ne mesurent pas la même valeur de température d'air pendant

les périodes  $P_{ICU}^{(jour)}$  et  $P_{ICU}^{(nuit)}$ . Ces différences, assimilées aux variations spatiales de l'ICU, sont calculées à partir de deux indicateurs :  $ITLS_{std}$  et  $ITLS_{moy}$ . Ceux-ci sont modélisés à partir de régressions linéaires multiples. Les variables explicatives sont des indicateurs géographiques représentatifs des flux contribuant au bilan d'énergie (équation 1.1) : densité de façades, densité de bâtiments, densité de surface, densité de surface d'eau,  $NDVI$ . Ces indicateurs géographiques sont calculés au sein de cercles de rayon 500 m centrés sur chacune des stations de mesure. Un dernier indicateur est calculé, la distance de la station de mesure à la périphérie, qui correspond en réalité à n'importe quelle zone dépourvue de bâtiments dans un rayon donné. Les modèles spatiaux et temporels sont calibrés à partir du jeu de données  $JDD_{calib}$ . Cinq variantes à la méthode de modélisation initiale sont ensuite utilisées, donnant lieu à la création de 16 modèles spatiaux et 32 modèles temporels (32 ensembles de modèles). Deux de ces variantes sont liées à la robustesse des modèles (estimation robuste des indicateurs ou régression linéaire robuste). Une variante consiste à découper l'année en EM saisonniers, une autre à découper un site en six sites selon la direction du vent. La dernière variante consiste à n'utiliser pour le calcul de l'indicateur temporel  $ITLT$  que les sites de mesure de plus forte variance. Aucun des EM créé à partir d'une variante spécifique n'est a priori meilleur et le but du chapitre suivant est d'identifier les meilleurs EM au sein des 32 qui ont été calibrés.





# 4

## Vérification et choix de l'ensemble de modèles le plus performant

Ce chapitre est consacré à l'évaluation des EM présentés au chapitre 3. Trente deux modèles temporels et 16 modèles spatiaux ont été construits, soit 32 EM. Chacun d'eux est associé à une combinaison de cinq variantes. La méthode générique d'évaluation d'un EM sera décrite en section 4.1. Les performances générales de tous les EM seront ensuite considérées pour évaluer l'intérêt de chaque variante (section 4.2). Les équations des modèles spatiaux ( $ITLS_{moy}$  et  $ITLS_{std}$ ) et du modèle temporel  $ITLT$  qui appartiennent à l'EM le plus performant seront données en section 4.4. La présence de chaque variable météorologique dans le modèle temporel et de chaque indicateur géographique dans les modèles spatiaux seront discutés dans cette section. Enfin, une analyse détaillée des estimations obtenues avec l'EM le plus performant sera effectuée section 4.5.

### 4.1 Indicateur retenu pour juger les performances d'un ensemble de modèles

Les indicateurs  $ITL^{(nuit)}$  et  $ITL^{(jour)}$  sont estimés à partir de l'équation 3.9. Les valeurs obtenues sont comparées aux observations des jeux de données présentés au chapitre 2 selon la

procédure décrite figure 4.1.

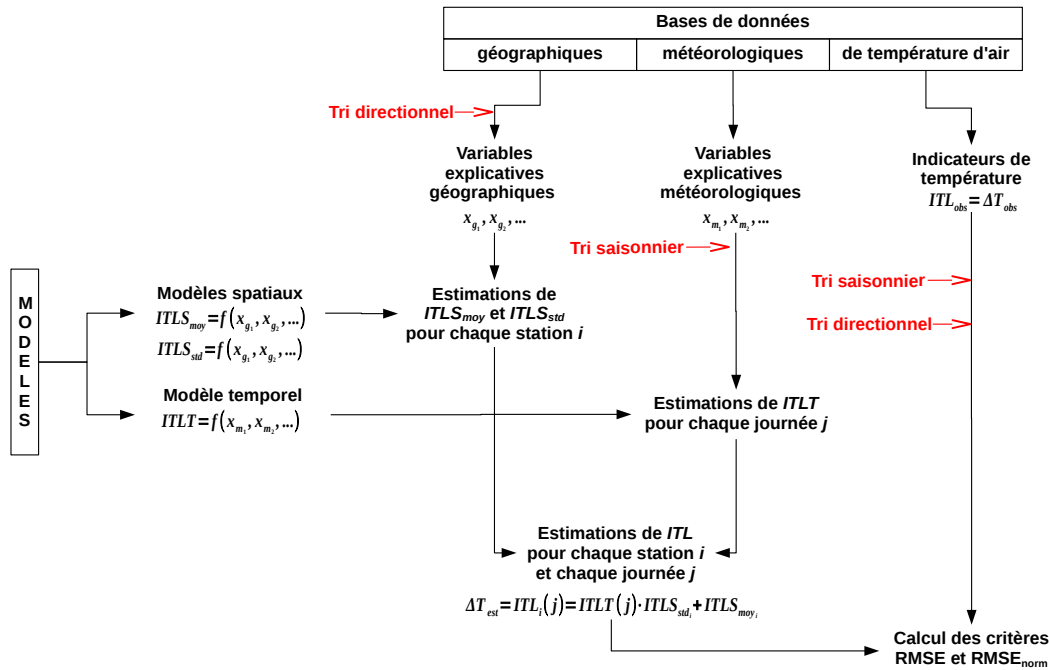


FIGURE 4.1 – Procédure utilisée pour estimer les indicateurs  $ITL^{(jour)}$  et  $ITL^{(nuit)}$ . Ces estimations permettent ainsi de vérifier la performance des EM en comparant les valeurs estimées et les valeurs observées

Les variantes de tri directionnel et de tri saisonnier sont à considérer lors de l'estimation des valeurs de l'indicateur  $ITL_i(j)$ . L'étape de la procédure dans laquelle ils doivent intervenir est donc mentionnée sur le schéma. Les autres variantes ont eu, comme le tri saisonnier et le tri directionnel, une influence sur les modèles produits. Cependant, ils ne nécessitent pas de réaliser de traitement particulier pour la production des estimations, d'où leur absence sur le schéma 4.1. Les indicateurs  $RMSE$  et  $\widehat{RMSE}$  sont utilisés pour juger des performances d'un EM.  $\widehat{RMSE}$  a été défini et illustré respectivement équation 2.26 et figure 2.36. Il a servi dans chapitre 2 à juger les performances des modèles  $corr_T$ , utilisés pour corriger les erreurs de mesure de température. L'équation 4.1 est ici adaptée à la variable à estimer ( $ITL$ ) :

$$\widehat{RMSE} = \sqrt{\frac{\frac{1}{nb_j} \frac{1}{nb_i} \sum_j^{nb_j} \sum_i^{nb_i} (ITL_{obs_i}(j) - ITL_{est_i}(j))^2}{\frac{1}{nb_j} \frac{1}{nb_i} \sum_j^{nb_j} \sum_i^{nb_i} (ITL_{obs_i}(j) - ITL_{obs_{ref}}(j))^2}} \quad (4.1)$$

avec  $ITL_{obs_i}(j)$  et  $ITL_{obs_{ref}}(j)$  respectivement les valeurs de l'indicateur  $ITL$  mesurées à la station  $i$  et à la station de référence pour le jour  $j$ ,  $ITL_{est_i}(j)$  les valeurs de l'indicateur  $ITL$  estimées à partir des modèles à la station  $i$  pour le jour  $j$ ,  $nb_i$  le nombre de stations du jeu de données utilisé pour la vérification,  $nb_j$  le nombre de journées utilisées pour la vérification et

$ITL_{est_i}(j)$  la température d'air estimée à la station  $i$  pour le jour  $j$ .

Si l'on ne dispose pas d'outils pour quantifier les microclimats urbains, on considère que la température de l'agglomération est homogène et égale à celle mesurée à la station de référence (Météo-France)  $T_{ref}$ . Dans ce cas, l'erreur, qui peut être considérée comme une erreur de référence, est calculé à partir du numérateur de l'équation 4.1. En revanche, si des modèles sont utilisés pour calculer l'écart de température entre la station de référence et plusieurs points  $i$  d'une agglomération, l'erreur du modèle, qui correspond grossièrement à la différence entre l'estimation et l'observation, est calculée à partir du numérateur de l'équation 4.1. Ces deux erreurs sont des erreurs quadratiques et le rapport des deux permet d'évaluer si l'utilisation des modèles proposés dans cette thèse sont plus pertinents que simplement considérer la température de l'agglomération comme homogène : si  $\widehat{RMSE}$  est inférieur à un, l'utilisation du modèle est justifiée. Si ce n'est pas le cas, il est préférable de considérer la température provenant de la station Météo-France comme homogène pour tout le territoire de l'agglomération.

## 4.2 Performances générales des ensembles de modèles

Les différences de performance des 32 EM proposés au chapitre 3 sont illustrées figure 4.2 par deux histogrammes. Ils représentent la distribution des valeurs de  $RMSE$  et de  $\widehat{RMSE}$  lorsque l'on considère tous les EM. Au préalable, les résultats ont été triés par jeu de données. Ceci permet d'observer l'évolution des performances des EM lorsque les données de vérification s'éloignent de celles utilisées pour la calibration. Les jeux de données  $JDD_{spat}$  et  $JDD_{spat\&temp}$  ont des données seulement pour certains mois de l'année. Afin de pouvoir comparer nos jeux de données pour des mêmes périodes, les résultats sont donc également répartis en deux catégories : les saisons chaudes (été, printemps) et les saisons froides (automne, hiver). Pour certains EM, la valeur des indicateurs est parfois trop élevée pour apparaître sur l'histogramme. Pour savoir quels jeux de données sont concernés par ce problème, le pourcentage de valeurs visibles sur les graphiques est indiqué dans la légende. Pour l'indicateur nocturne, trois observations sont faites.

1. La  $RMSE$  des saisons froides est inférieur à celui des saisons chaudes. Le contraire est observé pour l'indicateur  $\widehat{RMSE}$ .
2. Pour tous les jeux de données produits sur l'agglomération nantaise, la  $\widehat{RMSE}$  est toujours inférieur à 1. Pour les jeux de données d'Angers et de La Roche-sur-Yon, seulement



une moitié des EM correspond à cette situation.

3. Plus on s'éloigne des conditions spatiales et temporelles de calibration, plus les valeurs de  $\widehat{RMSE}$  sont élevées. Ce constat est moins vrai pour le  $RMSE$  dont les différences de valeurs entre jeux de données sont moins marquées.

L'indicateur  $RMSE$  est intéressant pour avoir un ordre de grandeur de l'erreur d'un EM. Trop sensible à la valeur des différences de température à modéliser, il n'est cependant pas adapté pour comparer les performances de deux EM entre eux. Si la différence de température est élevée (exemple : été ou jeu de données  $JDD_{spat\&temp}$ ), la  $RMSE$  est souvent élevée. Elle sera faible dans le cas contraire (hiver). L'indicateur  $\widehat{RMSE}$  est beaucoup moins sensible à ce phénomène puisqu'il est rapporté à la valeur de différence de température. C'est donc, en quelque sorte, une erreur relative aux conditions du problème tandis que l'indicateur  $RMSE$  représente une erreur absolue.

L'observation 2 permet de conclure que tous les EM appliqués au territoire utilisé pour la calibration sont pertinents. Il est donc préférable d'utiliser n'importe lequel de ces EM plutôt que de supposer l'agglomération à température homogène. Ce n'est pas le cas pour l'extrapolation à d'autres zones d'étude où plus de la moitié des EM ont une  $\widehat{RMSE} > 1$ .

Pour l'indicateur diurne, les observations 1 et 3 restent valables. Par contre, quel que soit le jeu de données, seuls quelques EM sont pertinents pour modéliser les différences de température. Pour certains jeux de données, il s'avère même parfois qu'aucun EM ne permet de respecter l'inégalité  $\widehat{RMSE} < 1$ .

Les différences entre les 32 EM proposés sont l'utilisation ou non des variantes. Pour rappel, celles-ci correspondent : au tri des données par direction de vent, au tri par saison, à l'utilisation de régressions robustes, à l'utilisation d'une méthode de centrage réduction robuste, à l'utilisation de seulement quelques stations pour établir les modèles temporels. Leur utilisation peut avoir des répercussions positives ou négatives sur la performance des EM. Pour chaque saison et chaque jeu de données, les cinq EM qui donnent les meilleurs résultats (selon l'indicateur  $\widehat{RMSE}$ ) sont analysés. Un tableau peut ainsi être dressé pour chaque combinaison {saison, jeu de données,  $ITL$  (diurne ou nocturne)}. Le tableau 4.1 correspond par exemple à la combinaison {automne, calibration, indicateur nocturne}. La variante *tri directionnel* n'est utilisée par aucun des cinq meilleurs EM de cette combinaison. Le tri des données par direction de vent n'est donc pas propice à l'élaboration d'EM performants pour cette combinaison {saison, jeu de données,  $ITL$ }.

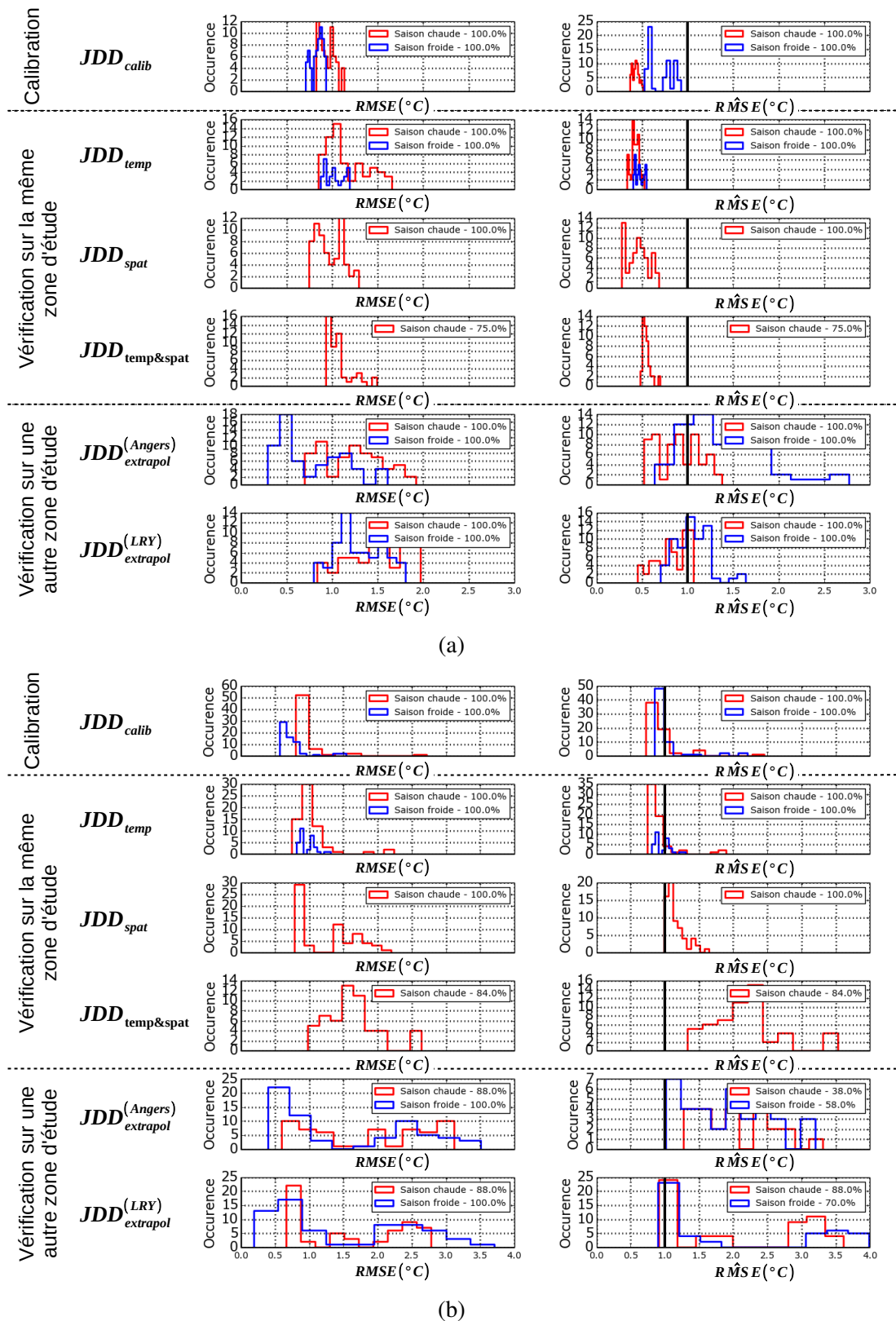


FIGURE 4.2 – Distribution saisonnière des valeurs de  $RMSE$  et de  $\widehat{RMSE}$  pour chacun des jeux de données. Cette figure permet de juger de la performance générale des ensembles de modèles (EM) utilisés pour estimer (a) l'indicateur nocturne  $ITL^{(nuit)}$  et (b) l'indicateur diurne  $ITL^{(jour)}$  selon le jeu de données considéré. La saison chaude comprend le printemps et l'été tandis que la saison froide comprend l'automne et l'hiver

classement de l'EM	Tri saisonnier	Tri directionnel	Utilisation de quelques stations seulement	Standardisation robuste	Régression robuste
1	Oui	Non	Oui	Non	Non
2	Oui	Non	Non	Non	Non
3	Oui	Non	Oui	Non	Oui
4	Oui	Non	Non	Oui	Oui
5	Oui	Non	Oui	Oui	Non

TABLE 4.1 – Variantes utilisées par les cinq meilleurs ensembles de modèles pour la combinaison {saison : automne, jeu de données : calibration, indicateur  $ITL$  : indicateur nocturne  $ITL^{(nuit)}$ }.

À partir de tous ces tableaux, une analyse des performances de chaque variante peut être conduite par saison et par indicateur  $ITL$ . Seules deux variantes ont une réelle influence (positive ou négative) sur les résultats : le *tri directionnel* et le *tri saisonnier*.

- Le tri directionnel a un impact sur la modélisation tant de l'indicateur nocturne que de l'indicateur diurne. La figure 4.3 permet en effet de constater qu'il est utilisé quelle que soit la saison par la plupart des EM obtenus pour les jeux de données utilisant les réseaux URBIO (c'est-à-dire les jeux  $JDD_{spat\&temp}$ ,  $JDD_{extrapol}^{(Angers)}$  et  $JDD_{extrapol}^{(LRY)}$ ). Les réseaux URBIO ont été installés en veillant à placer le capteur au sein de sites les plus isotropes possibles (dans un rayon de 100 m). Le fait que le tri directionnel améliore les résultats n'est donc pas intuitif pour ces jeux de données puisque la valeur des indicateurs géographiques aurait dû être presque égale quelle que soit la direction du vent. Cependant, il ne faut pas oublier que les modèles ont été élaborés à partir du jeu de données de calibration. Celui-ci utilise très peu de sites de mesure (10) souvent fortement anisotropes. Dans ce cas, le tri directionnel a donc un intérêt car il permet de multiplier le nombre de sites pour l'élaboration des modèles. Si ce tri a peu d'intérêt lorsque les jeux de données de vérification ont des stations de mesure ou des périodes de mesure en commun (respectivement  $JDD_{spat}$  et  $JDD_{temp}$ ), elle est tout-à-fait pertinente dans les autres cas ( $JDD_{spat\&temp}$ ,  $JDD_{extrapol}^{(Angers)}$  et  $JDD_{extrapol}^{(LRY)}$ ).
- Le tri saisonnier n'a un impact que lorsque l'indicateur nocturne est modélisé (figure 4.4). Il est clairement préférable de l'utiliser du point de vue du jeu de données de calibration.

Cependant, pour les autres jeux de données, les cinq meilleurs EM n'ont pas été construits à partir de modèles saisonniers. Ceci signifie que les EM saisonniers établis à partir du jeu de données de calibration ne sont pas robustes. Ils ne permettent pas une bonne estimation lorsqu'ils sont extrapolés à d'autres jeux de données, sauf pour le printemps. Trier les données saisonnièrement n'est donc pas très intéressant compte tenu des performances résultantes.

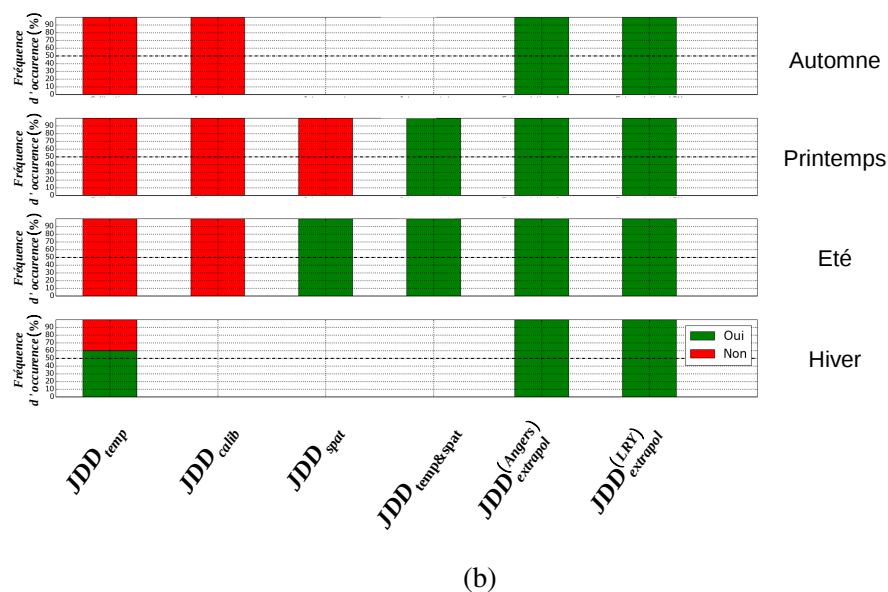
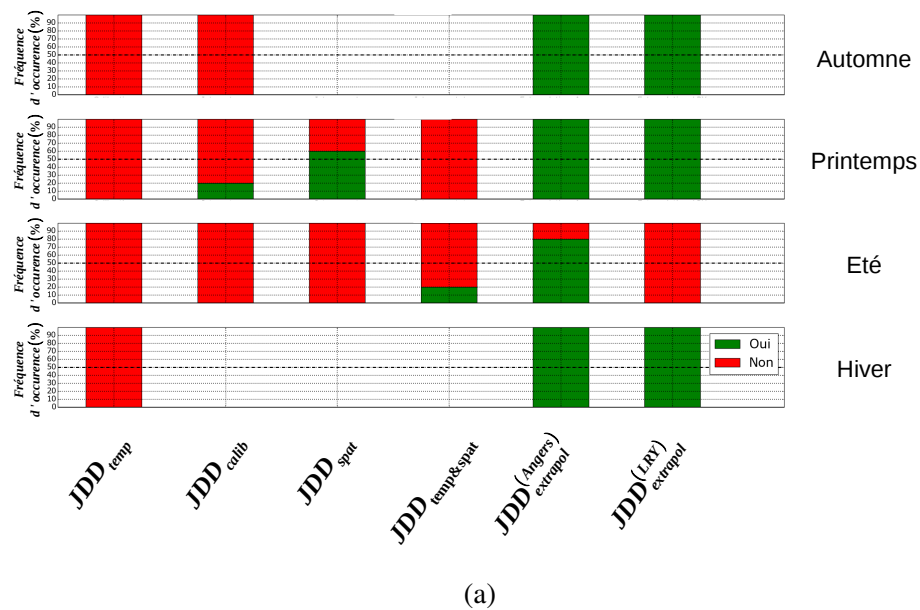


FIGURE 4.3 – Fréquence d'utilisation de la variante *tri directionnel* par les cinq meilleurs EM obtenus pour chaque combinaison {jeu de données, saison} (a) lorsque l'indicateur nocturne est modélisé, (b) lorsque l'indicateur diurne est modélisé

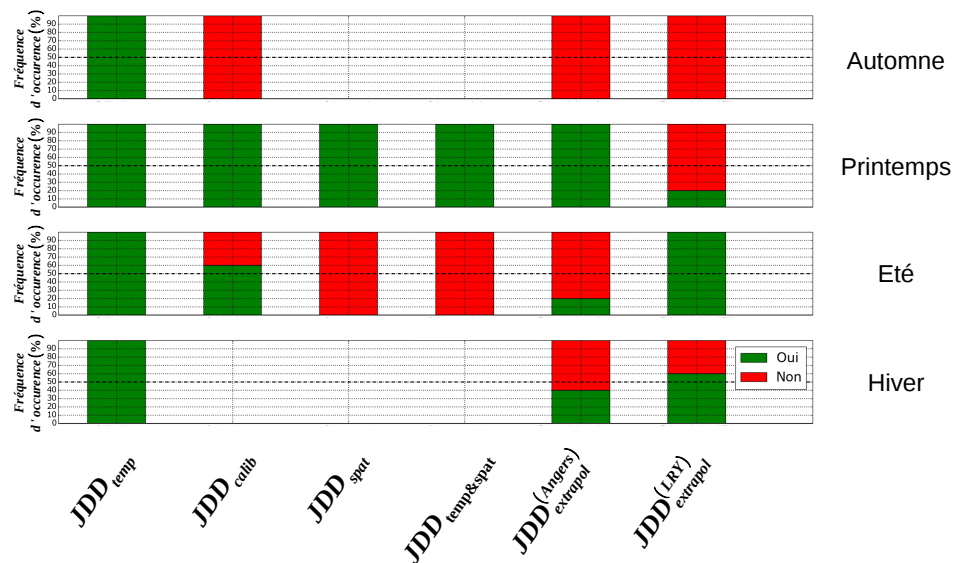


FIGURE 4.4 – Fréquence d'utilisation de la variante *tri saisonnier* par les cinq meilleurs EM obtenus pour chaque combinaison {jeu de données, saison} lorsque l'indicateur nocturne est modélisé

### 4.3 Choix du meilleur ensemble de modèles

Un EM performant pour un jeu de données ne l'est pas nécessairement pour tous les autres. Le but de cette thèse est de trouver un ensemble de modèles qui permette de modéliser le climat de n'importe quelle agglomération pour n'importe quel type de temps. Cet EM doit donc avoir des performances satisfaisantes quel que soit le jeu de données. Pour le choisir, la procédure présentée figure 4.5 est utilisée : pour chaque saison et chaque jeu de données, les EM sont classés par valeur de  $\widehat{RMSE}$  croissant. Si le meilleur EM obtenu par chaque jeu de données est identique, alors cet EM est retenu. Sinon, les deux meilleurs EM sont considérés. Si l'un d'entre eux est commun à tous les jeux de données, il est retenu. Sinon, les trois meilleurs EM sont considérés et ainsi de suite jusqu'à trouver un EM qui soit partagé par tous les jeux de données.

Un EM par saison est retenu pour chaque indicateur ( $ITL^{(nuit)}$  et  $ITL^{(jour)}$ ). Pour l'indicateur nocturne, le tableau 4.2 est établi afin d'observer pour chaque EM :

- la valeur de  $\widehat{RMSE}$  associée à chaque jeu de données : ceci permet d'identifier à quel(s) jeu(x) de données cet EM est mal adapté ;
- les variantes qui ont été utilisées : ceci permet d'affiner l'analyse des variantes réalisée auparavant.

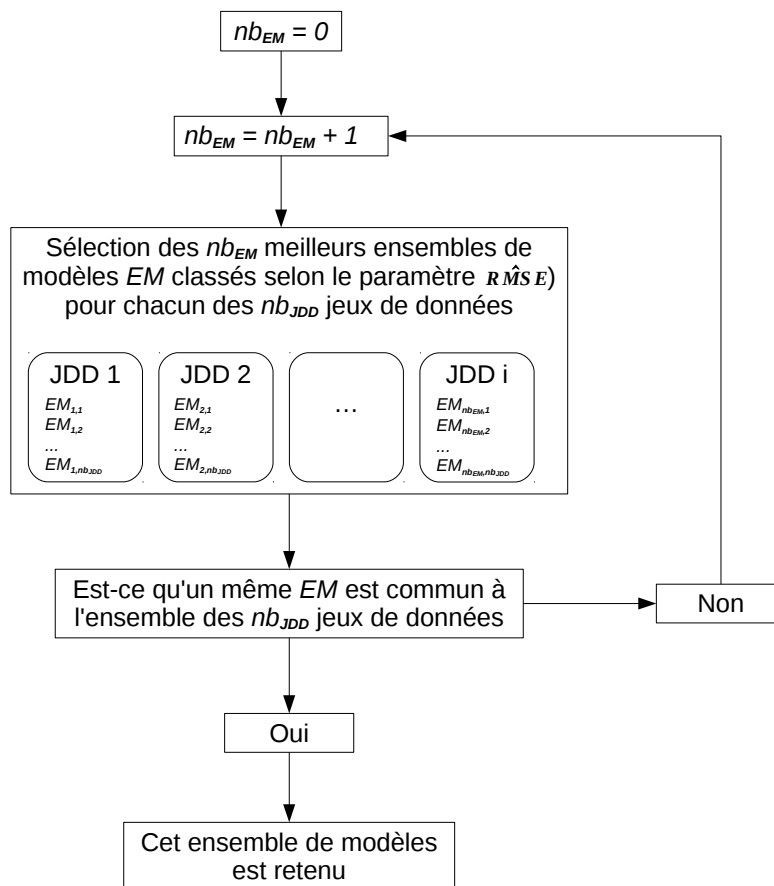


FIGURE 4.5 – Procédure utilisée pour choisir l’EM le plus performant en considérant tous nos jeux de données. À chaque itération, les  $nb_{EM}$  EM les plus performants au vu de chaque jeu de données sont retenus afin de vérifier si un EM est commun à l’ensemble des jeux

Saison modélisée	$\widehat{RMSE}$ pour chaque JDD	Variantes utilisées				
		Tri saison- nier	Tri direction- nel	Seulement certaines stations	Standardisation robuste	Régression robuste
Automne	$JDD_{calib} : 0,58$ $JDD_{temp} : 0,46$ $JDD_{extrapol}^{(LRY)} : 0,71$ $JDD_{extrapol}^{(Angers)} : 0,65$	<b>Non</b>	Oui	<b>Non</b>	<b>Non</b>	Oui
Hiver	$JDD_{calib} : 0,81$ $JDD_{extrapol}^{(LRY)} : 0,87$ $JDD_{extrapol}^{(Angers)} : 1,4$	Oui	Oui	Oui	Oui	Oui
Printemps	$JDD_{calib} : 0,42$ $JDD_{temp} : 0,42$ $JDD_{spat} : 0,29$ $JDD_{spattemp} : 0,53$ $JDD_{extrapol}^{(LRY)} : 0,87$ $JDD_{extrapol}^{(Angers)} : 1,4$	Oui	Oui	Oui	<b>Non</b>	Oui
Été	$JDD_{calib} : 0,46$ $JDD_{temp} : 0,45$ $JDD_{spat} : 0,54$ $JDD_{spattemp} : 0,54$ $JDD_{extrapol}^{(LRY)} : 0,76$ $JDD_{extrapol}^{(Angers)} : 1,2$	Oui	Oui	Oui	<b>Non</b>	Oui

TABLE 4.2 – Variantes utilisées par le modèle  $ITL^{(nuit)}$  (retenu pour chaque saison) et performances ( $\widehat{RMSE}$ ) obtenues pour chaque jeu de données.

#### 4.4. DISCUSSION DE LA PERTINENCE PHYSIQUE DES MODÈLES SPATIAUX ET DU MODÈLE TEM

Seul l'EM obtenu pour l'automne permet de maintenir le  $\widehat{RMSE}$  en-dessous de la barre unitaire pour tous les jeux de données. Pour toutes les autres saisons, le jeu de données angevin ( $JDD_{extrapol}^{(Angers)}$ ) a en effet une valeur de  $\widehat{RMSE}$  supérieure à un. Le tri directionnel est utilisé pour toutes les saisons, le tri saisonnier également sauf pour l'automne. Ces résultats sont cohérents avec l'analyse des variantes faite auparavant. Les modèles calibrés en utilisant seulement certaines stations donnent également les meilleurs résultats pour toutes les saisons sauf l'automne. Tous les EM sont également issus de régressions robustes tandis que seul l'EM hivernal est issu d'un centrage et d'une réduction robuste. L'utilisation de méthodes robustes de régression ou de centrage-réduction n'est donc pas toujours synonyme de robustesse des modèles obtenus.

Les valeurs de  $\widehat{RMSE}$  sont beaucoup moins bonnes pour l'indicateur diurne puisqu'aucune saison ne permet d'obtenir un EM qui soit intéressant pour plus de deux jeux de données. Ceci n'est pas étonnant au regard de la figure 4.3b où une minorité d'EM intéressants apparaissaient pour seulement trois des six jeux de données utilisés. Trois raisons peuvent expliquer ce résultat :

- soit la période de calcul de l'indicateur diurne a été mal choisie ;
- soit le jeu de données de calibration utilisé n'est pas approprié pour établir des modèles diurnes (dû aux erreurs de mesure des capteurs TRH) ;
- soit les différences de température diurne sont trop faibles pour être modélisées convenablement.

Finalement, seuls les EM permettant de calculer l'ICU nocturne seront analysés plus en détail par la suite.

## 4.4 Discussion de la pertinence physique des modèles spatiaux et du modèle temporel retenus

Les équations 4.2, 4.3, 4.4 et 4.5 présentent les modèles temporels choisis respectivement pour l'automne, l'hiver, le printemps et l'été. Les équations 4.6, 4.7, 4.8 et 4.9 présentent les modèles spatiaux de dispersion des écarts entre stations ( $ITLS_{std}$ ) respectivement pour l'automne, l'hiver, le printemps et l'été. Les équations 4.10, 4.11, 4.12 et 4.13 présentent les modèles spatiaux des écarts moyens ( $ITLS_{moy}$ ) respectivement pour l'automne, l'hiver, le printemps et l'été.



Pour rappel, les variables explicatives  $x_g$  sont calculées de sorte que  $x_g = IG_{g_i} - IG_{g_{ref}}$  pour chaque indicateur  $IG$  excepté pour  $d_{periph}$  pour lequel  $x_g = d_{periph_i}$ .

Les variables explicatives utilisées dans ces équations n'ont pas été centrées réduites. Il n'est donc pas possible de comparer leurs coefficients (l'influence d'une variable par rapport à une autre) mais leur analyse permet de mettre en évidence :

- quelles variables météorologiques ou géographiques sont influentes pour chaque saison (selon si elles sont présentes ou non dans l'équation) ;
- si ces variables augmentent ou au contraire diminuent l'ICU (coefficients  $\alpha$  positifs ou négatifs) ;

Et plus particulièrement pour les modèles temporels, on peut observer :

- le type d'évolution associé à chacune de ces variables (linéaire, exponentielle, etc.) ;
- la période pendant laquelle leur influence est la plus forte ( $P_A$ ,  $P_B$ ,  $P_C$  ou  $P_D$ ).

$$ITLT^{(aut)} = 0,0016K \downarrow (P_C) + 6,3 \frac{1}{U(P_A) + 3,0} + 0,01T(P_C) - 0,095 \exp\left(\frac{N(P_B) - 6}{6}\right) - 1,1 \quad (4.2)$$

$$ITLT^{(hiv)} = 0,081 \exp(-U(P_B) + 3) - 0,12N(P_A) + 0,64 \quad (4.3)$$

$$ITLT^{(pri)} = 0,0011K \downarrow (P_C) + 11,0 \frac{1}{U(P_A) + 3,0} + 0,033T(P_D) - 0,12N(P_A) - 2,3 \quad (4.4)$$

$$ITLT^{(ete)} = -0,0083RH(P_D) - 0,36U(P_A) - 0,13T(P_A) + 3,2 \quad (4.5)$$

$$ITLS_{std}^{(aut)} = 3,2x_{D_{fac}} + 0,06 \ln(d_{periph} + 1) - 0,027x_{NDVI} - 2,9x_{D_{bat}} + 0,38 \quad (4.6)$$

$$ITLS_{std}^{(hiv)} = -3,0x_{D_{fac}} + 1,4x_{D_{surf}} + 0,065 \ln(d_{periph} + 1) + 0,5 \quad (4.7)$$

$$ITLS_{std}^{(pri)} = 1,7x_{D_{surf}} + 0,12 \ln(d_{periph} + 1) - 0,029x_{NDVI} - 3,4x_{D_{bat}} + 0,41 \quad (4.8)$$

$$ITLS_{std}^{(ete)} = -6,4x_{D_{fac}} + 3,0x_{D_{surf}} + 4,8x_{D_{eau}} + 0,15 \ln(d_{periph} + 1) + 0,86 \quad (4.9)$$

$$ITLS_{moy}^{(aut)} = 5,3x_{D_{fac}} + 0,15 \ln(d_{periph} + 1) - 0,057x_{NDVI} - 5,5x_{D_{bat}} - 0,39 \quad (4.10)$$

$$ITLS_{moy}^{(hiv)} = 2,0x_{D_{surf}} + 0,12 \ln(d_{periph} + 1) - 0,027x_{NDVI} - 4,0x_{D_{bat}} - 0,2 \quad (4.11)$$

$$ITLS_{moy}^{(pri)} = 3,3x_{D_{eau}} - 0,051x_{NDVI} + 0,031 \quad (4.12)$$

$$ITLS_{moy}^{(ete)} = 12,0x_{D_{eau}} - 0,045x_{NDVI} + 0,23 \quad (4.13)$$

Plusieurs enseignements peuvent être tirés des modèles temporels.

- L'été est étrangement la seule saison où aucune des deux variables assimilées au rayonnement solaire (rayonnement solaire descendant et nébulosité) n'est influente. C'est également la seule saison où l'humidité relative a une influence : une journée où elle aura été élevée est souvent synonyme d'ICU nocturne faible. On peut faire l'hypothèse que les précipitations sont à l'origine de cette relation. L'humidité élevée serait engendrée par des précipitations importantes qui réduiraient les différences de température spatiales à néant. Ce terme peut ainsi être vu comme un terme de rayonnement solaire mais ayant un effet contraire.
- Le vent est la seule variable qui a une influence quelle que soit la saison. Plus le vent augmente, plus l'ICU diminue. Cette relation est linéaire seulement pour l'été. Cette linéarité peut poser problème dans le cas d'extrapolation des modèles temporels à des régions où les vents sont élevés. Pour les autres saisons, la fonction utilisée est conforme aux observations qui avaient pu être réalisées auparavant (ie l'ICU tend vers zéro lorsque la vitesse de vent tend vers l'infini). La période où le vent a le plus d'influence sur la valeur de l'ICU est toujours une période nocturne (la période  $P_B$  pour l'hiver et la période  $P_A$  pour toutes les autres saisons).
- Un terme de nébulosité est présent pour trois des quatre modèles saisonniers. Les variations d'ICU qu'il engendre sont, comme pour la vitesse du vent, simples à interpréter physiquement : plus la nébulosité est importante, plus l'ICU est faible. Sa période d'influence est également nocturne : la période  $P_A$  est utilisée dans deux cas,  $P_B$  dans le troisième cas.
- Le terme de température apparaît trois fois sous forme linéaire. Étrangement, il n'est pas présent pour l'hiver, où une température faible aurait pu correspondre à des charges anthropiques importantes entraînant un ICU élevé. En automne et au printemps, plus la température est importante en journée, plus l'ICU est important la nuit. Une température élevée en journée est souvent corrélée à un rayonnement solaire important (et à une nébulosité faible). Les surfaces urbaines stockent ainsi des quantités de chaleur importantes en journée et les déstockent pendant la nuit, phénomène propice au développement de l'ICU. Au printemps et en automne, des termes de nébulosité et de rayonnement solaire sont déjà présents. Le terme de température pourrait donc apparaître comme redondant.

En été, c'est la température mesurée pendant la nuit qui explique le mieux l'ICU. Plus la température mesurée à la station de référence est faible, plus l'ICU est important. Une température nocturne faible en dehors de la ville est souvent associée à une couverture nuageuse faible. Le terme de température est donc sans doute une alternative au terme de nébulosité qui n'est pas utilisé dans le modèle estival.

- La pression atmosphérique, qui semblait pourtant être étroitement liée à l'ICU à certaines saisons, n'apparaît dans aucun modèle temporel. Même si aucune explication physique ne pouvait expliquer une liaison directe entre la pression atmosphérique et l'ICU, des liaisons indirectes pouvaient être entrevues : une pression élevée est souvent synonyme de ciel dégagé et de vent faible, conditions propices au développement de l'ICU. L'utilisation de  $PA$  aurait donc été redondant dans le cas où la vitesse du vent et le rayonnement solaire étaient déjà utilisés.

Le même travail est réalisé à partir des équations de dispersion et de valeur moyenne des écarts de température. Ces deux indicateurs  $ITLS_{moy}$  et  $ITLS_{std}$  sont étroitement liés : lorsque la moyenne de l'écart augmente, l'écart-type augmente souvent également. On utilisera donc le terme ICU pour qualifier indifféremment l'ICU moyen (défini par  $ITLS_{moy}$ ) mais également les variations d'ICU (définies par  $ITLS_{std}$ ). Cinq conclusions sont tirées.

- Le  $NDVI$  est l'un des indicateurs les plus fréquemment utilisés dans les équations. Plus une station a un  $NDVI$  élevé (plus la végétation est dense), plus elle a un ICU faible.
- Plus la distance  $d_{periph}$  d'un site à une zone non bâtie est importante, plus son ICU sera élevé.
- Plus la densité de bâtiment  $D_{bat}$  d'un quartier est élevée, plus son ICU est faible. Son terme, dont le coefficient  $\alpha_{sD_{bat}}$  est toujours négatif, est parfois accompagné d'un terme de densité de surface  $D_{surf}$  dont le coefficient  $\alpha_{sD_{surf}}$  est toujours positif. Cette situation n'apparaît qu'en hiver et au printemps, ce qui correspond grossièrement aux saisons de chauffe. Ceci est cohérent avec le fait que la combinaison d'un terme  $\alpha_{sD_{surf}} x D_{surf}$  positif et d'un terme  $\alpha_{sD_{bat}} x D_{bat}$  négatif a été associée section 3.3.1.1 au flux de chaleur anthropique émis par les bâtiments. Lorsque  $\alpha_{sD_{bat}} x D_{bat}$  apparaît seul (en automne), son effet est difficilement interprétable physiquement. En revanche, lorsque  $\alpha_{sD_{surf}} x D_{surf}$  apparaît seul,  $\alpha_{sD_{surf}}$  est toujours positif et peut être associé au flux de stockage/destockage de chaleur. Plus ce terme est important, plus l'ICU est donc élevé.

- Le terme  $\alpha_{sD_{fac}} x_{D_{fac}}$  est négatif lorsque le terme  $\alpha_{sD_{surf}} x_{D_{surf}}$  est présent dans l'équation, et positif dans le cas contraire. La densité de façades  $D_{fac}$  a été dans notre cas (section 3.3.1.1) assimilée au rayonnement piégé dans la canopée. Le coefficient  $\alpha_{sD_{fac}}$  devrait donc normalement toujours être positif. Cependant,  $D_{surf}$  et  $D_{fac}$  sont corrélés : l'utilisation d'une seule variable permet donc d'expliquer les effets majeurs des deux variables sur l'ICU. Si la seconde variable est ajoutée, c'est pour expliquer son effet résiduel mineur sur l'ICU. Cet effet peut ne pas être intuitif. C'est par exemple le cas du signe négatif de  $\alpha_{sD_{fac}}$ .
- Plus la densité de surface en eau est importante, plus l'ICU nocturne est important. Cette observation est tirée uniquement des modèles pour les saisons d'été et de printemps. Les surfaces d'eau sont souvent assimilées à un dispositif de rafraîchissement, ce qui est généralement vrai lorsqu'il s'agit de rivières (Hathway et Sharples 2012). Au contraire, lorsqu'il s'agit de bassins de faible profondeur, leur présence peut contribuer à exacerber l'ICU localement. Ceci est particulièrement vrai en fin d'été, lorsque plusieurs journées consécutives très chaudes sont observées. Ainsi, l'eau peut être plus chaude que l'air pendant la nuit, et donc décharger sa chaleur dans l'atmosphère (Steenveld et al. 2014). Si cela permet de comprendre la valeur positive du coefficient  $\alpha_{sD_{eau}}$  en été, la valeur négative obtenue pour le printemps ne correspond a priori à aucune réalité physique. Il est nécessaire de rappeler que le  $NDVI$  des surfaces en eau est beaucoup plus faible que celui des surfaces de végétation. Le  $NDVI$  est pourtant utilisé ici comme un indicateur représentatif de la capacité d'une surface à évaporer. De la même manière qu'une surface couverte de végétation, une surface d'eau est source d'évaporation. La combinaison des indicateurs  $NDVI$  et  $D_{eau}$  nuit donc à l'interprétation des résultats. Il serait donc opportun de supprimer  $D_{eau}$  de la phase de calibration et d'assigner aux surfaces d'eau une valeur de  $NDVI$  cohérente avec leur comportement évaporatif. Une deuxième solution pourrait être de les retirer du calcul de  $NDVI$  moyen au sein de la zone d'influence tout en conservant l'indicateur  $D_{eau}$ .

Certains termes des équations temporelles et spatiales peuvent être prépondérants. Afin de comparer l'influence relative des différents termes constitutifs de ces équations, la procédure suivante est appliquée.

- Pour les modèles temporels, le terme de rayonnement  $0,0016K \downarrow (P_C)$  du modèle automnal sera utilisé pour illustrer la procédure.

1. L'intercentile  $\int_{q_1=0,05}^{q_2=0,95} (x_{VM_m}(j))_j$  de chaque variable météorologique  $VM_m$  est calculé pour chaque saison (ceci équivaut à prendre le minimum et le maximum en éliminant préalablement les cinq pourcents des valeurs les plus élevées et les cinq pourcents des valeurs les plus faibles). Pour cela, les données utilisées sont celles du jeu de calibration et l'intercentile calculé est défini comme l'amplitude  $amp_{x_{VM}}$  de cette variable.

$$amp_{x_{K\downarrow}} = \int_{q_1=0,05}^{q_2=0,95} (K \downarrow_{PC}(j))_j = 337,5 W.m^{-2} \quad (4.14)$$

2. La valeur de chaque terme  $\alpha_{t_{x_m}} x_m$  est calculée pour  $x_m = amp_{x_m}$ . Ceci permet de calculer les variations maximales de l'indicateur *ITLT* attribuables à chacun des termes.

$$\alpha_{t_{K\downarrow PC}} amp_{x_{K\downarrow}} = 0,54 \quad (4.15)$$

3. Les valeurs *normalisées* qui correspondent à chacun des termes  $\alpha_{t_{x_m}} x_m$  issus des modèles temporels saisonniers sont rassemblés dans le tableau 4.3 pour comparaison.
- Pour les modèles spatiaux (de dispersion (*ITLS<sub>std</sub>*) mais aussi de valeur moyenne (*ITLS<sub>moy</sub>*) des écarts), la même procédure est utilisée. Cependant, en complément des données des 10 points de mesure du jeu de calibration, les données géographiques issues de 200 points choisis aléatoirement sur le territoire de l'agglomération nantaise sont également utilisées. Pour rappel, un point est considéré comme invalide lorsqu'il est situé dans l'eau ou sur un bâtiment. Si c'est le cas, il est déplacé au point valide le plus proche. Les indicateurs géographiques sont calculés au sein de cercles de 500 m de rayon centrés sur chacun de ces 200 points. Cet échantillon est supposé représentatif de la variabilité géographique du territoire étudié. Les résultats de l'application de la procédure décrite précédemment sont présentés tableau 4.4 et tableau 4.5, respectivement pour les modèles spatiaux de dispersion et de valeurs moyennes des écarts.

En hiver, la vitesse du vent ayant un impact négligeable, seul le terme de nébulosité dicte les variations de l'ICU (tableau 4.5). En été, le terme  $\alpha_{t_{RH}} amp_{x_{RH}}$  est également négligeable devant les termes faisant intervenir la vitesse du vent et la température. En revanche, en automne et au printemps, tous les termes utilisés ont une influence comparable, avec un petit avantage néanmoins pour la vitesse du vent au printemps.

Saison	Jeu de données	$\alpha_{t_{K\downarrow}} amp_{x_{K\downarrow}}$	$\alpha_{t_{RH}} amp_{x_{RH}}$	$\alpha_{t_U} amp_{x_U}$	$\alpha_{t_T} amp_{x_T}$	$\alpha_{t_N} amp_{x_N}$
Automne	Calibration	0,54	-	0,73	0,22	-0,93
Hiver	Calibration	-	-	0,01	-	-0,93
Printemps	Calibration	0,34	-	1,43	0,42	-0,95
Été	Calibration	-	-0,33	-1,29	-1,23	-

TABLE 4.3 – Valeurs *normalisées* des termes  $\alpha_{t_{VMm}} amp_{x_{VMm}}$  issus des équations régissant le calcul de l'indicateur  $ITLT^{(nuit)}$

Saison	Jeu de données	$\alpha_{ssD_{fac}} \cdot amp_{x_{D_{fac}}} (\text{°C})$	$\alpha_{ssD_{surf}} \cdot amp_{x_{D_{surf}}} (\text{°C})$	$\alpha_{ssD_{eau}} \cdot amp_{x_{D_{eau}}} (\text{°C})$	$\alpha_{ssd_{periph}} \cdot amp_{x_{d_{periph}}} (\text{°C})$	$\alpha_{ssNDVI} \cdot amp_{x_{NDVI}} (\text{°C})$	$\alpha_{ssD_{bat}} \cdot amp_{x_{D_{bat}}} (\text{°C})$
Automne	Calibration	1,32	-	-	0,37	-1,16	-1,71
	Échantillon 200 points	0,91	-	-	0,35	-1,07	-1,30
Hiver	Calibration	-1,25	1,05	-	0,40	-	-
	Échantillon 200 points	-0,86	0,55	-	0,38	-	-
Printemps	Calibration	-	1,27	-	0,73	-1,23	-1,97
	Échantillon 200 points	-	0,66	-	0,70	-1,13	-1,50
Été	Calibration	-2,64	2,33	<b>0,23</b>	0,94	-	-
	Échantillon 200 points	-1,83	1,21	<b>1,72</b>	0,90	-	-

TABLE 4.4 – Valeurs *normalisées* des termes  $\alpha_{ssIG_g} amp_{x_{IG_g}}$  issus des équations régissant le calcul de l'indicateur  $ITLS_{std}^{(nuit)}$ . Les termes sont calculés pour chacun des jeux de données et également pour 200 points tirés aléatoirement sur le territoire. La comparaison des valeurs obtenues permet de vérifier que les jeux de données utilisés sont bien représentatifs du territoire en regard des indicateurs géographiques utilisés dans chacun des termes du modèle.

Concernant les modèles spatiaux, en plus de la comparaison des termes entre eux, une comparaison d'un même terme entre les deux jeux de données doit être effectuée. Afin de considérer les sites utilisés dans le jeu de données de calibration comme suffisamment représentatifs de la variabilité de notre agglomération, il faut que la valeur du terme  $\alpha_{sIG_g} amp_{x_{IG_g}}$  calculé à partir des données de calibration soit supérieure ou égale à celle du terme calculé à partir des données de l'échantillon de 200 points. Si ce n'est pas le cas pour un indicateur  $IG_g$ , cela si-

Saison	Jeu de données	$\alpha_{smD_{fac}} \cdot amp_{x_{D_{fac}}} \text{ (}^\circ\text{C)}$	$\alpha_{smD_{surf}} \cdot amp_{x_{D_{surf}}} \text{ (}^\circ\text{C)}$	$\alpha_{smD_{eau}} \cdot amp_{x_{D_{eau}}} \text{ (}^\circ\text{C)}$	$\alpha_{sm_{d_{periph}}} \cdot amp_{x_{d_{periph}}} \text{ (}^\circ\text{C)}$	$\alpha_{smNDVI} \cdot amp_{x_{NDVI}} \text{ (}^\circ\text{C)}$	$\alpha_{smD_{bat}} \cdot amp_{x_{D_{bat}}} \text{ (}^\circ\text{C)}$
Automne	Calibration	2,19	-	-	0,90	-2,42	-3,18
	Échantillon 200 points	1,51	-	-	0,86	-2,23	-2,43
Hiver	Calibration	-	1,52	-	0,76	-1,15	-2,30
	Échantillon 200 points	-	0,79	-	0,72	-1,06	-1,75
Printemps	Calibration	-	-	<b>0,16</b>	-	-2,16	-
	Échantillon 200 points	-	-	<b>1,20</b>	-	-2,0	-
Été	Calibration	-	-	<b>0,56</b>	-	-1,92	-
	Échantillon 200 points	-	-	<b>4,17</b>	-	-1,77	-

TABLE 4.5 – Valeurs *normalisées* des termes  $\alpha_{smIG_g} amp_{x_{IG_g}}$  issus des équations régissant le calcul de l'indicateur  $ITLS_{moy}^{(nuit)}$ . Les termes sont calculés pour chacun des jeux de données et également pour 200 points tirés aléatoirement sur le territoire. La comparaison des valeurs obtenues permet de vérifier que les jeux de données utilisés sont bien représentatifs du territoire en regard des indicateurs géographiques utilisés dans chacun des termes du modèle.

gnifie qu'appliqué à certaines zones de l'agglomération, le modèle spatial risque de donner des résultats peu satisfaisants. C'est le cas pour la densité de surfaces en eau, qui est toujours largement supérieure pour l'échantillon de 200 points. La température d'un site situé au bord d'une grande étendue d'eau sera donc certainement mal estimée. Au vu des équations spatiales, plus la densité d'eau est importante, plus l'ICU est important. La température d'un site situé près de l'eau sera donc très probablement sur-estimée. Cela peut poser problème si l'ensemble de modèles devait être appliqué à des zones d'études où les surfaces en eau recouvrent une grande partie du territoire.

D'une manière générale, le terme de distance à la périphérie n'est un terme prédominant pour aucun des modèles spatiaux. En automne, le  $NDVI$ ,  $D_{bat}$  et  $D_{fac}$  ont environ la même influence potentielle. L'absence de terme prédominant pour cette saison est sans doute associée au fait que le modèle utilisé est le seul à ne pas être issu d'un tri saisonnier. Aucune spécificité saisonnière ne ressort. En hiver, ce sont les indicateurs liés à la chaleur d'origine anthropique ( $D_{surf}$  et  $D_{bat}$ ), au flux de stockage/destockage ( $D_{surf}$ ) et au rayonnement solaire ( $D_{fac}$ ) qui dominent. Le  $NDVI$  ne fait pas partie des indicateurs ayant une forte influence



pendant cette saison. Au printemps, le  $NDVI$  domine largement le modèle de  $ITLS_{moy}$  mais aucune terme ne se dégage du modèle de  $ITLS_{std}$ . Par conséquent, les indicateurs associés au flux de chaleur latente ( $NDVI$ ) et à la chaleur d'origine anthropique ( $D_{surf}$  et  $Q_B$ ) ont presque le même effet  $\alpha_{ssIG_g} amp_{xIG_g}$ . En été, le terme contenant le  $NDVI$  est, comme au printemps, le terme prépondérant pour le modèle de valeur moyenne. Ce sont les termes liés au rayonnement solaire (indicateur  $D_{fac}$ ) et au stockage/déstockage de chaleur (indicateur  $D_{surf}$ ) qui sont prépondérants pour le modèle de dispersion (cf section 3.3.1.1). Toutes ces observations sont physiquement pertinentes, confortant le choix de cet ensemble de modèles. Cependant, la présence des indicateurs  $NDVI$  et  $D_{eau}$  dans la plupart des modèles spatiaux peut poser problème : nous avons vu section 3.3.3 que les sites du jeu de calibration n'étaient pas représentatifs de la zone d'étude au regard de ces deux indicateurs. Nous verrons dans la section suivante entre autres quelles implications auront ces éléments sur les résultats.

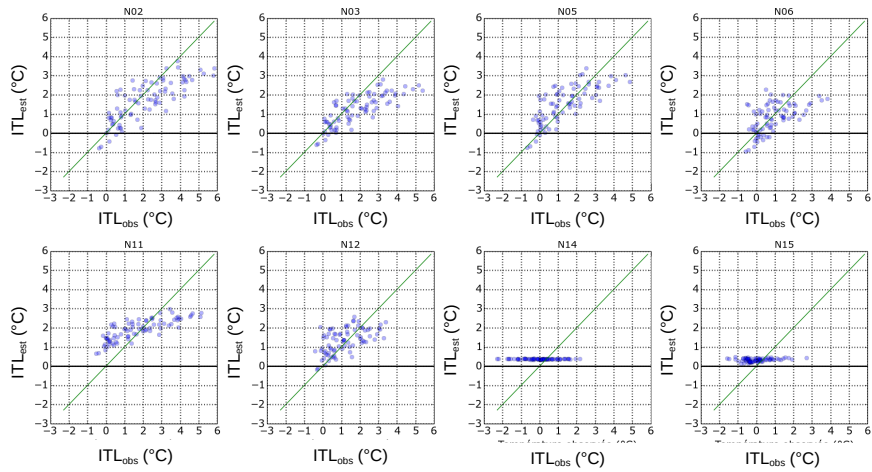
## 4.5 Analyse spatiale et temporelle des résultats

Il s'agit maintenant de comprendre pourquoi les performances de l'EM finalement adopté sont limitées. Ceci est particulièrement le cas des jeux de données produits à partir des réseaux URBIO (les deux jeux de données d'extrapolation -  $JDD_{extrapol}^{(Angers)}$  et  $JDD_{extrapol}^{(LRY)}$  - et le jeu  $JDD_{spat\&temp}$ ). Les valeurs  $ITL_{est}$  estimées pour chacun des sites qui constituent ces jeux de données seront comparées aux valeurs des données observées  $ITL_{obs}$ . Les nuages de points de  $ITL_{est} = f(ITL_{obs})$  sont tracés figure 4.6 pour chaque site de chacun de nos jeux de données. Ceux-ci correspondent seulement aux données printanières.

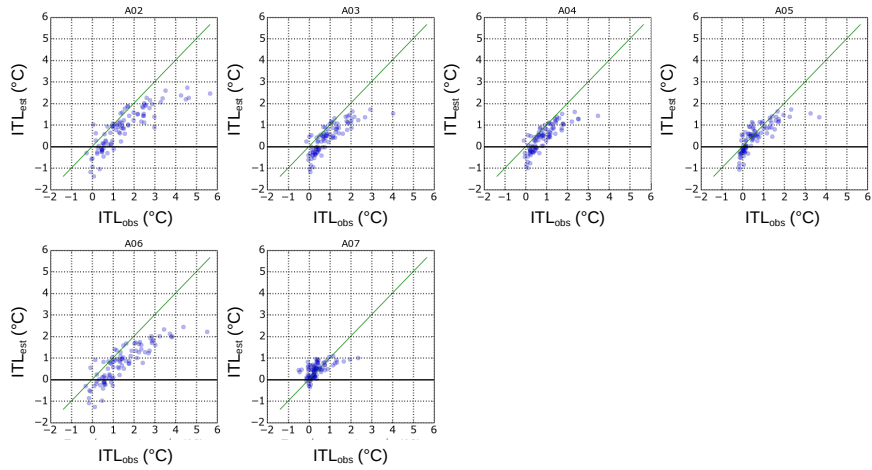
### 4.5.1 Principes d'analyse des résultats

À partir du nuage de points d'un site donné, quatre types de dysfonctionnement peuvent être observés. Chacun d'eux est représenté figure 4.7 et peut être relié à une cause à partir de l'analyse de l'équation 3.9 ( $ITL_i(j) = \frac{ITLT_i(j) \cdot ITLS_{std_i} + ITLS_{moy_i}}{\sigma_i}$ ) :

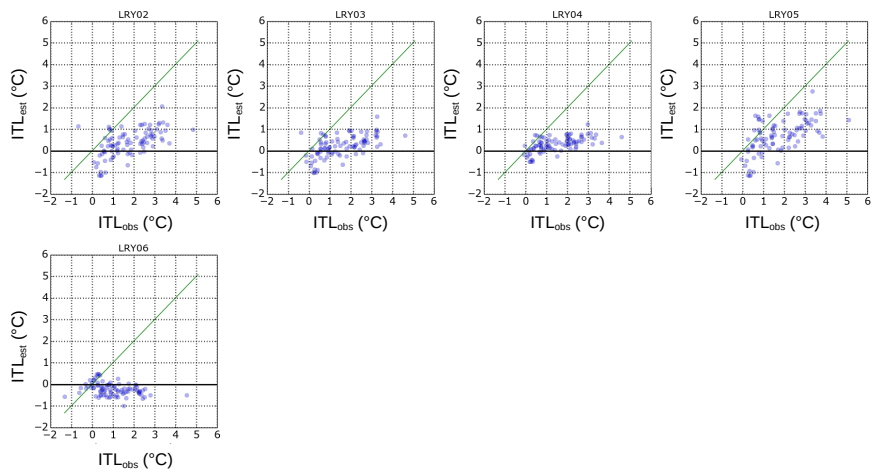
- cas où le nuage de points n'est pas linéaire (figure 4.7a) : ceci correspond à une modélisation temporelle défailante ;
- cas où le nuage de points est linéaire mais n'est pas parallèle à la droite  $ITL_{est} = ITL_{obs}$  (figure 4.7b) : ceci correspond à une modélisation spatiale de la dispersion de l'écart défailante (pente mal calculée) ;



(a)



(b)



(c)

FIGURE 4.6 – Valeurs estimées de l'indicateur  $ITL^{(pri)}$  en fonction des valeurs observées pour chaque station de mesure des jeux de données (a)  $JDD_{spat\&temp}$ , (b)  $JDD_{extrapol}^{(Angers)}$  et (c)  $JDD_{extrapol}^{(LRY)}$

- cas où le nuage de points est colinéaire à la droite  $ITL_{est} = ITL_{obs}$  mais est situé au-dessus ou en-dessous d'elle (figure 4.7c) : ceci correspond cette fois à un problème de modélisation de la valeur moyenne de l'écart à la station de référence ;
- cas où certaines parties du nuage de points sont clairement décalées du reste (figure 4.7d) : modèles spatiaux non-adaptés pour une ou plusieurs directions de vent.

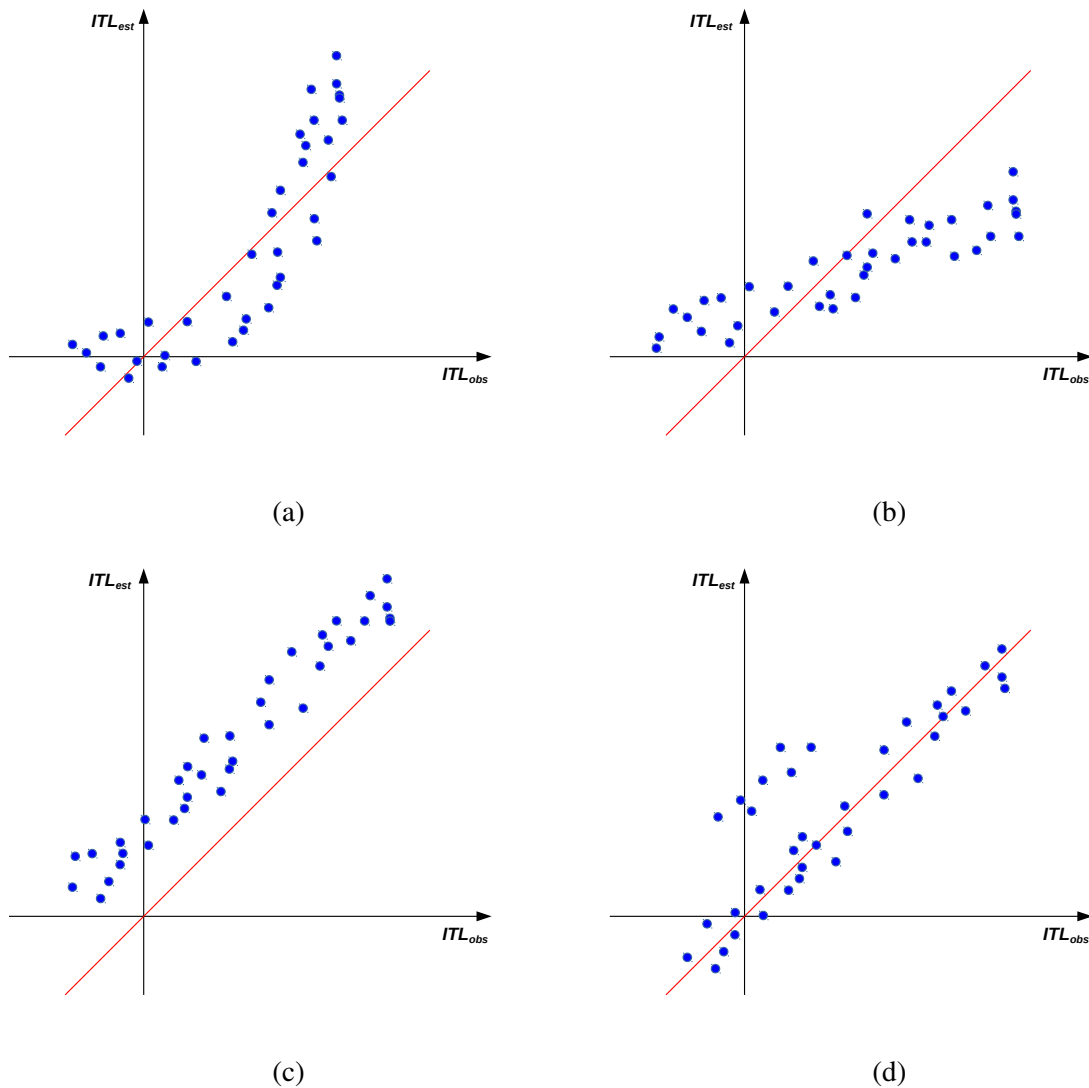


FIGURE 4.7 – Types de dysfonctionnement observables lorsque l'on trace le nuage de point  $ITL_{est} = f(ITL_{obs})$  pour une station de mesure donnée : (a) modèle temporel erroné (indicateur  $ITL_{LT}$ ), (b) modèle spatial de la dispersion erroné (indicateur  $ITL_{S_{std}}$ ), (c) modèle spatial de la moyenne erroné (indicateur  $ITL_{S_{moy}}$ ) et (d) modèles spatiaux erronés pour une ou plusieurs directions de vent

Pour les modèles spatiaux, si les mêmes défaillances apparaissent seulement pour une minorité de sites d'un jeu de données, la cause peut être double.

- Le problème peut venir du modèle : il n'est peut-être pas adapté pour modéliser certains

types de sites (les sites à vérifier ne ressemblent géographiquement à aucun autre site utilisé pour calibrer les modèles).

- Le problème peut venir des sites : les indicateurs géographiques utilisés dans les modèles sont peut-être mal calculés pour certains sites. La taille de la zone d'influence peut ne pas être adaptée ou l'environnement réel des capteurs est trop fortement différent des informations géographiques (issues de bases de données) utilisées pour caractériser les sites.

Au contraire, si les mêmes défaillances sont observées pour une majorité des sites, il est fort probable que le problème provienne du site de référence utilisé. La taille de la zone d'influence peut ne pas être adaptée ou l'environnement réel du capteur de référence est trop fortement différent des informations géographiques utilisées pour caractériser le site.

#### 4.5.2 Analyse spatiale des résultats

Selon cette méthode d'analyse des résultats, le fonctionnement des modèles spatiaux appliqués au jeu de données nantais (figure 4.6a) paraît globalement bon : une majorité des sites ont leur nuage de points presque parallèle et centré sur la droite  $ITL_{est} = ITL_{obs}$ . Cependant, les estimations de certaines stations (N14, N15 et dans une moindre mesure N11) sont beaucoup moins bonnes. Bien que l'indicateur  $ITLS_{moy}$  soit relativement bien modélisé (la moyenne des températures observées est proche de celle des températures estimées), les modèles de dispersion  $ITLS_{std}$  ne fonctionnent pas du tout. Ceci peut s'expliquer par le fait que le jeu de calibration est peu représentatif des variations du NDVI au sein de l'agglomération nantaise, contrairement au jeu de vérification  $JDD_{spat\&temp}$  qui l'est grâce aux stations N14 et N15. Ceci est mis en évidence par la figure 4.8 où, si l'on ne tient pas compte de la station de référence, il apparaît que :

- deux stations du jeu de données  $JDD_{spat\&temp}$  (les stations N14 et N15) ont un NDVI supérieur à 150 ;
- aucune station du jeu de calibration  $JDD_{calib}$  n'a un NDVI supérieur à 150.

La calibration du modèle de dispersion est donc peut-être erronée faute d'avoir été réalisée à partir d'un jeu de données contenant au moins un site à fort NDVI.

Le problème de la station N11 provient probablement du fait qu'elle est située au sein d'un cimetière très minéral (figure 4.9). Son NDVI est faible, mais la quantité de bâtiments

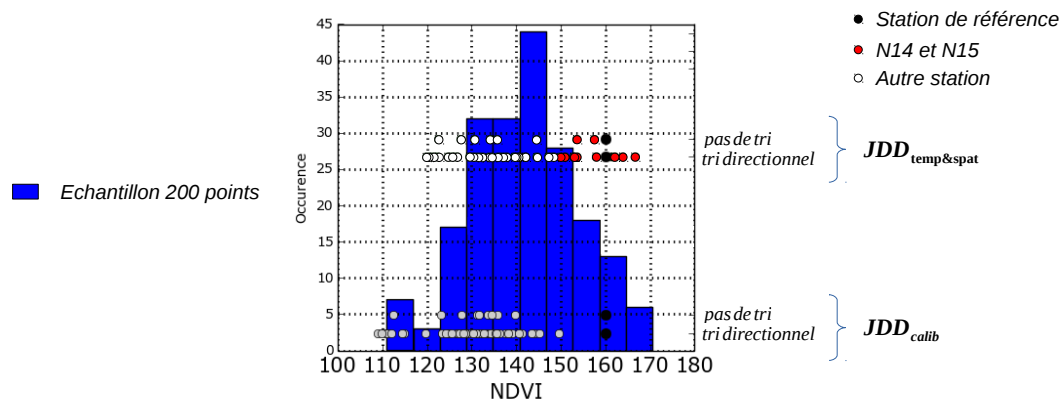


FIGURE 4.8 – Distribution des valeurs de l'indicateur  $NDVI$  calculées pour 200 points aléatoirement choisis sur l'agglomération de Nantes et pour les stations des jeux de données  $JDD_{calib}$  et  $JDD_{spat\&temp}$ . Cette figure permet de vérifier que les deux jeux de données sont bien géographiquement représentatifs (au regard de l'indicateur  $NDVI$ ) de la zone d'étude de Nantes.

présente alentour est également faible. Cette particularité géographique n'est observée pour aucun site du jeu de calibration. De ce fait, les modèles spatiaux n'intègrent pas les spécificités microclimatiques liées à ce type d'environnement géographique.

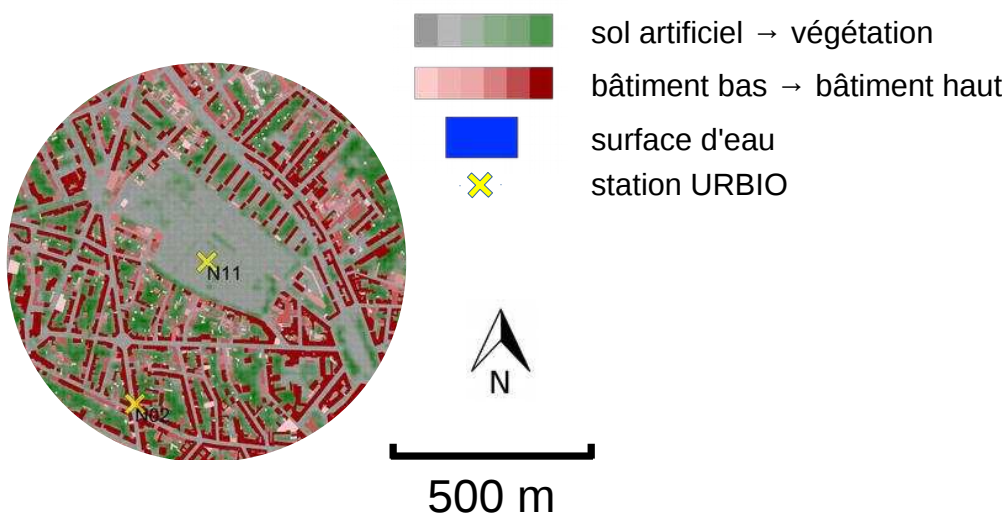


FIGURE 4.9 – Contexte géographique du site N11

Aucun des deux modèles spatiaux n'est globalement bon pour le jeu de données d'Angers. Sur la figure figure 4.6, la droite de tendance de tous les sites semble avoir une pente et une valeur moyenne sous-estimées. Le problème ne vient donc pas des modèles. Si tel était le cas, seuls quelques sites n'auraient pas été bien modélisés. Le problème vient donc de la station de référence. Un (ou plusieurs) des indicateurs géographiques utilisés dans les modèles spatiaux a (ont) dû être mal calculé(s) pour la station MF\_Angers. Au printemps, le  $NDVI$  est la seule

variable présente dans les deux modèles  $ITLS_{moy}$  et  $ITLS_{std}$ . On notera  $e_{ITLS_{std}}$  et  $e_{ITLS_{moy}}$  respectivement les erreurs de modélisation de  $ITLS_{std}$  et  $ITLS_{moy}$ , attribuables à une erreur de calcul du  $NDVI$  de la station de référence. La correction des modèles spatiaux suit donc les relations :

$$\begin{aligned} ITLS_{std_{cor}} &= ITLS_{std_{ini}} - e_{ITLS_{std}} \\ ITLS_{moy_{cor}} &= ITLS_{moy_{ini}} - e_{ITLS_{moy}} \end{aligned} \quad (4.16)$$

avec  $ITLS_{std_{ini}}$  et  $ITLS_{std_{cor}}$  le modèle de dispersion respectivement avant et après correction,  $ITLS_{moy_{ini}}$  et  $ITLS_{moy_{cor}}$  le modèle de valeur moyenne respectivement avant et après correction.

En considérant tous les sites, il apparaît que le modèle  $ITLS_{std}$  (la pente) est en moyenne sous-évalué d'environ  $\frac{2}{5}^{\circ}\text{C}$  (soit  $e_{ITLS_{std}} = -\frac{2}{5}^{\circ}\text{C}$ ), tandis que  $ITLS_{moy}$  est sous-évalué de  $1^{\circ}\text{C}$  (soit  $e_{ITLS_{moy}} = -1^{\circ}\text{C}$ ). À l'origine de cette erreur, les termes  $\alpha_{ss_{NDVI}}x_{NDVI}$  et  $\alpha_{sm_{NDVI}}x_{NDVI}$  peuvent être assimilés respectivement aux termes  $ITLS_{std}$  et  $ITLS_{moy}$  des équations 4.16 :

$$\begin{aligned} -0,029(NDVI_{i_{cor}} - NDVI_{ref_{cor}}) &= -0,029(NDVI_{i_{ini}} - NDVI_{ref_{ini}}) + \frac{2}{5} \\ -0,051(NDVI_{i_{cor}} - NDVI_{ref_{cor}}) &= -0,051(NDVI_{i_{ini}} - NDVI_{ref_{ini}}) + 1 \end{aligned} \quad (4.17)$$

avec  $NDVI_{i_{ini}}$  et  $NDVI_{i_{cor}}$  les valeurs de  $NDVI$  de la station  $i$  respectivement avant et après correction et  $NDVI_{ref_{ini}}$  et  $NDVI_{ref_{cor}}$  les valeurs de  $NDVI$  de la station  $ref$  respectivement avant et après correction.

D'où, en considérant que seule la valeur calculée à la station de référence est erronée (et donc que  $NDVI_{i_{ini}} = NDVI_{i_{cor}}$ ), le système d'équations 4.17 devient :

$$\begin{aligned} NDVI_{ref_{cor}} - NDVI_{ref_{ini}} &= 13,8 \\ NDVI_{ref_{cor}} - NDVI_{ref_{ini}} &= 20 \end{aligned}$$

Après correction, la valeur de  $NDVI$  de la station de référence devra donc être supérieure à la valeur initialement calculée. On considérera la nouvelle valeur de  $NDVI$  comme bonne si son augmentation est proche de l'intervalle  $[13,8 - 20]$ . L'erreur de calcul du  $NDVI$  de la station de référence est associée à un mauvais choix de la taille du rayon d'influence, fixé

arbitrairement à 500 m. Afin de voir si un autre rayon se prête mieux à l'exercice, le  $NDVI$  est calculé pour sept autres cercles de taille différente. Au final, les rayons de cercle testés sont 50, 100, 200, 300, 400, 500, 750 et 1000 m. Le site de référence des jeux de données nantais (MF\_Nort\_s\_e) ne semble pas engendrer d'erreurs géographiques. De plus, il est situé en-dehors de la zone d'étude. La valeur de ses indicateurs a seulement été estimées visuellement à partir de photos satellitaires. Sa valeur de  $NDVI$  est supposée équivalente aux valeurs couramment rencontrées pour des zones rurales enherbées. Il est difficile d'estimer précisément quelles pourraient être les variations de  $NDVI$  entre deux cercles de taille différente centrés sur le même site. Les calculs sont donc réalisés seulement pour les sites MF\_Angers et MF\_LRY, respectivement références des jeux de données d'extrapolation d'Angers et de LRY. Les résultats sont affichés sur la figure 4.10.

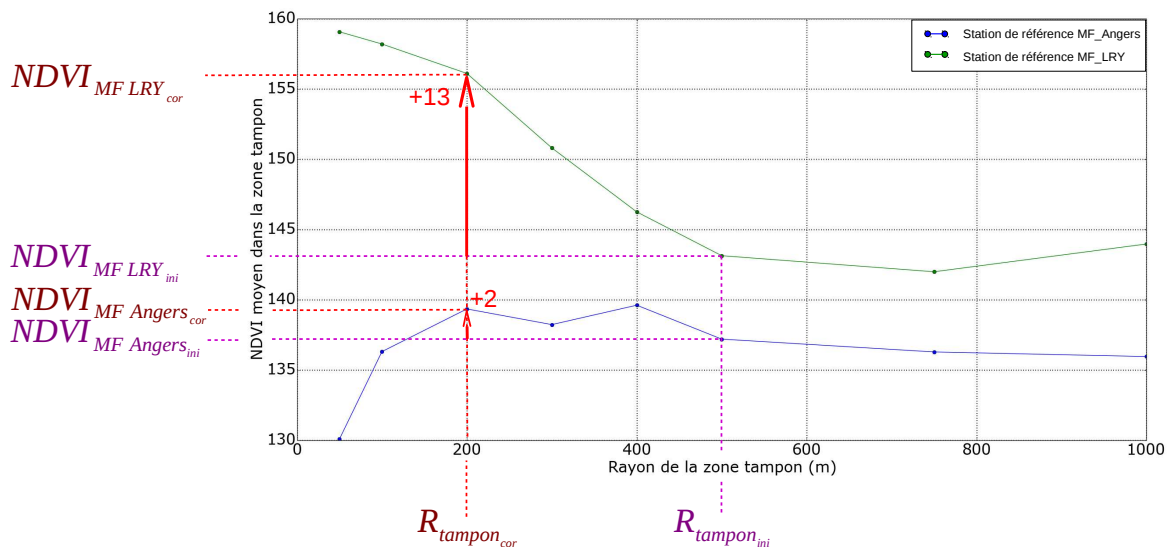


FIGURE 4.10 –  $NDVI$  des stations de référence  $MF\_Angers$  et  $MF\_LRY$  en fonction du rayon de la zone tampon utilisée pour le calcul des indicateurs géographiques thématiques.

Pour la station MF\_LRY, la valeur de  $NDVI$  diminue drastiquement pour des rayons compris entre 50 et 500 m. Elle se stabilise ensuite jusqu'à 1000 m. Ceci se comprend aisément lorsqu'on analyse visuellement le contexte géographique de cette station (figure 4.11a) : aucun bâtiment n'est présent dans les cercles de 50 et 100 m de rayon, et les sols ont un  $NDVI$  élevé. Dès lors que le cercle dépasse les 100 m de rayon, des bâtiments apparaissent et les sols sont composés de moins de végétation. La valeur  $NDVI_{ref\_cor}$  (après correction) doit être plus élevée que celle calculée initialement dans le cercle de 500 m. Au vue de l'allure de la courbe associée au site MF\_LRY (figure 4.10), le rayon d'influence qui doit être utilisé est nécessairement plus faible que 500 m. Dès lors que le rayon est inférieur ou égal à 200m, le  $NDVI_{ref\_cor}$  prend une

valeur proche de la valeur  $NDVI_{ref,cor}$  attendue (augmentation de 13 par rapport au  $NDVI$  initialement calculé pour un cercle de 500 m de rayon). Le choix d'un rayon inférieur ou égal à 200 m pour tous nos sites pourrait donc a priori s'avérer pertinent.

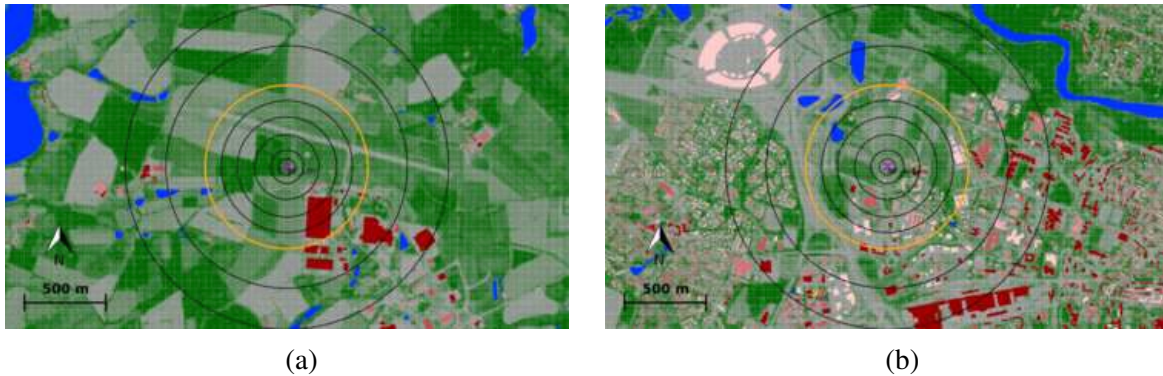


FIGURE 4.11 – Contexte géographique des stations de référence (a)  $MF\_LRY$  et (b)  $MF\_Angers$  au sein d'unités géographiques de référence concentriques de rayon 50, 100, 200, 300, 400, 500, 750 et 1000 m.

Le site LRY06 ne suit pas du tout la tendance de la droite  $ITL_{est} = ITL_{obs}$  (figure 4.6c). Il se différencie en ce sens des autres sites. De la même manière que pour les sites N14 et N15, l'explication vient probablement du fait que ce site a un  $NDVI$  très élevé (dont plusieurs directions correspondent à un  $NDVI > 150$  - cf. figure 4.12).

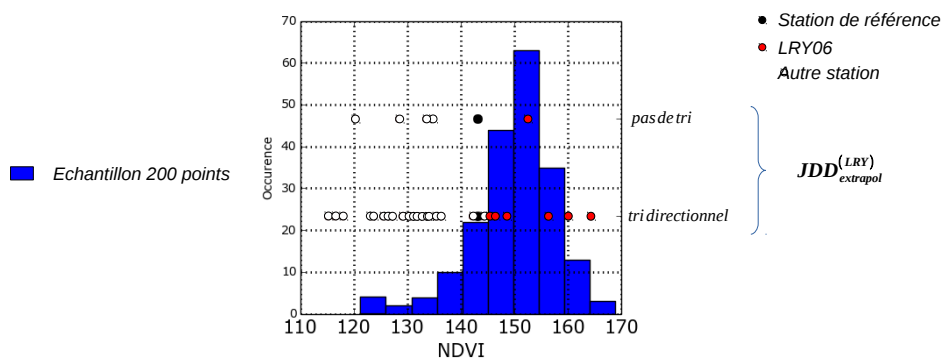


FIGURE 4.12 – Distribution des valeurs de l'indicateur  $NDVI$  calculées pour 200 points aléatoirement choisis sur l'agglomération de La Roche-sur-Yon et pour les stations du jeu de données  $JDD_{extrapol}^{(LRY)}$ . Cette figure permet de vérifier que les deux jeux de données sont bien géographiquement représentatifs (au vu de l'indicateur  $NDVI$ ) de la zone d'étude de LRY.

Concernant l'agglomération d'Angers, la pente est globalement bien calculée pour la majorité des sites (figure 4.6b) mais la valeur moyenne est légèrement sous-évaluée. La même procédure de diagnostic que celle effectuée pour le jeu de données de LRY est effectuée. Au vu des équations 4.8 et 4.12, seule la variable  $D_{eau}$  permettrait de modifier le modèle de  $ITLS_{moy}^{(pri)}$  sans altérer le modèle de  $ITLS_{std}^{(pri)}$ . Toujours d'après l'équation 4.12, il faudrait que la valeur



prise par cet indicateur pour le site de référence soit plus élevée que celle calculée actuellement. En considérant une sous-estimation de la valeur de  $ITLS_{moy}^{(pri)}$  de  $0,5^{\circ}\text{C}$  et d'après la valeur du coefficient  $\alpha_{ssD_{eau}} = 3,3$  de l'équation 4.12, il faudrait que la nouvelle valeur de densité de surface d'eau de la station MF\_Angers ( $D_{eau_{ref_{cor}}}$ ) soit supérieure de 0,15 à celle de la valeur initialement calculée pour le rayon de 500 m ( $D_{eau_{ref_{ini}}}$ ). Comme le montre la figure 4.11b, le rayon qui permettrait d'augmenter la valeur de  $D_{eau}$  est nécessairement supérieur à 500 m (il n'y a en effet pas d'eau dans les cercles de rayon inférieur ou égale à 400 m). Pour le cercle de 750 m de rayon, cette valeur est de 0,042 contre 0,021 pour le cercle de 500 m de rayon. L'augmentation de 0,021 correspond à une augmentation de la température d'environ  $0,1^{\circ}\text{C}$ , encore loin des  $0,5^{\circ}\text{C}$  nécessaires pour augmenter la valeur moyenne de tous les sites de  $0,5^{\circ}\text{C}$ . Le cercle de 1000 m ne permettant pas d'augmenter au-dessus de la valeur de 0,042, une autre explication doit être trouvée. À nouveau, le  $NDVI$  pourrait expliquer la sous-estimation du modèle de valeur moyenne en supposant qu'une augmentation de la valeur du  $NDVI$  de la station MF\_Angers permettrait de corriger ce biais. Au vue de la figure 4.10, seulement trois valeurs de rayon permettent d'augmenter la valeur du  $NDVI$  : 200 m, 300 m et 400 m. L'augmentation étant approximativement identique entre les rayons 200 et 400 m, on préférera le rayon de 200 m qui permettrait d'être cohérent avec le rayon de correction du site MF\_LRY. La différence de  $NDVI$  avant et après correction est de 2, ce qui correspond selon l'équation 4.12 à une augmentation de seulement  $0,1^{\circ}\text{C}$  de la valeur moyenne. Cette correction est loin d'être suffisante mais permet :

- de s'approcher de la valeur observée ;
- de rester cohérent avec la proposition de réduire la taille du cercle d'influence pour la station MF\_LRY.

Le modèle de dispersion dépend lui aussi du  $NDVI$ . Cette correction du site de référence ferait augmenter la pente du nuage de points  $ITL_{est} = f(ITL_{obs})$  d'une valeur de 0,05, augmentation considérée comme négligeable devant la valeur théorique idéale de 1. Aucun des six sites angevins observés figure 4.6b n'a de comportement particulier, contrairement aux jeux de données nantais ou yonnais où certains sites se démarquaient des autres. Ceci s'explique très probablement par le fait qu'aucun de ses sites n'a de  $NDVI$  particulièrement élevé. Lorsque le tri directionnel est effectué, seul un site dépasse très légèrement la valeur de 150. Les valeurs de  $NDVI$  de ce jeu de données sont donc d'après la figure 3.22 dans le même intervalle que les valeurs du jeu de calibration.

### 4.5.3 Analyse temporelle des résultats

Pour le jeu de données nantais, les variations temporelles sont assez mal représentées lorsque l'ICU est très important. Ceci est particulièrement flagrant pour le site N05 (figure 4.13) : le nuage de points suit relativement bien la droite d'équation  $ITL_{est} = ITL_{obs}$  jusqu'à une valeur  $ITL_{obs} = 3^{\circ}\text{C}$  (notée zone 2). Au-delà, la température estimée stagne à  $3^{\circ}\text{C}$  (zone 1).

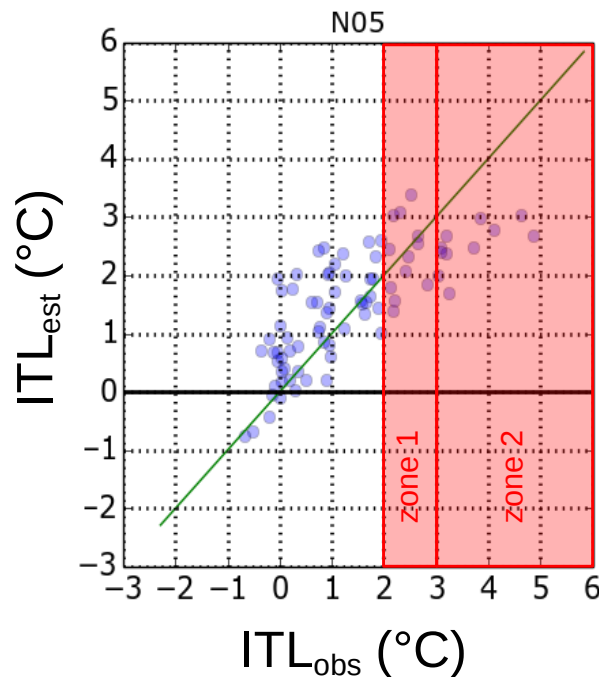


FIGURE 4.13 – Valeurs estimées de l'indicateur  $ITL$  en fonction des valeurs observées pour le site N05. La zone 1 est représentative de valeurs hautes bien modélisées, en opposition à la zone 2 contenant des valeurs hautes moins bien modélisées

Il s'agit donc d'identifier si une variable météorologique a été mal prise en compte dans les modèles temporels pour les fortes valeurs d'ICU. Pour cela, les conditions météorologiques qui ont engendré les valeurs pour  $ITL_{obs} > 3^{\circ}\text{C}$  (zone 1) sont comparées à celles qui ont engendré les  $ITL_{est}$  tels que  $2^{\circ}\text{C} < ITL_{obs} < 3^{\circ}\text{C}$  (zone 2). Cette comparaison est réalisée figure 4.14. Si, pour une variable météorologique donnée, les valeurs qui correspondent à la zone 1 diffèrent nettement de celles qui correspondent à la zone 2, cette variable est potentiellement à l'origine de la défaillance du modèle. La seule variable météorologique qui corresponde à cette description est la nébulosité calculée pendant la période  $P_B$ . Cette combinaison {variable, période} est la seule pour laquelle les points situés hors tendance ont une abscisse différente des points situés dans la tendance. Un terme en nébulosité existe déjà dans le modèle temporel printanier, mais pour la période  $P_A$ . Pour rappel, nous avons décidé au chapitre 3 de n'utiliser qu'une seule période par variable pour limiter les collinéarités. Cependant, autoriser l'ajout

d'une même variable pour une période différente permettrait dans certains cas d'améliorer les modèles temporels pour les valeurs hautes de l'ICU.

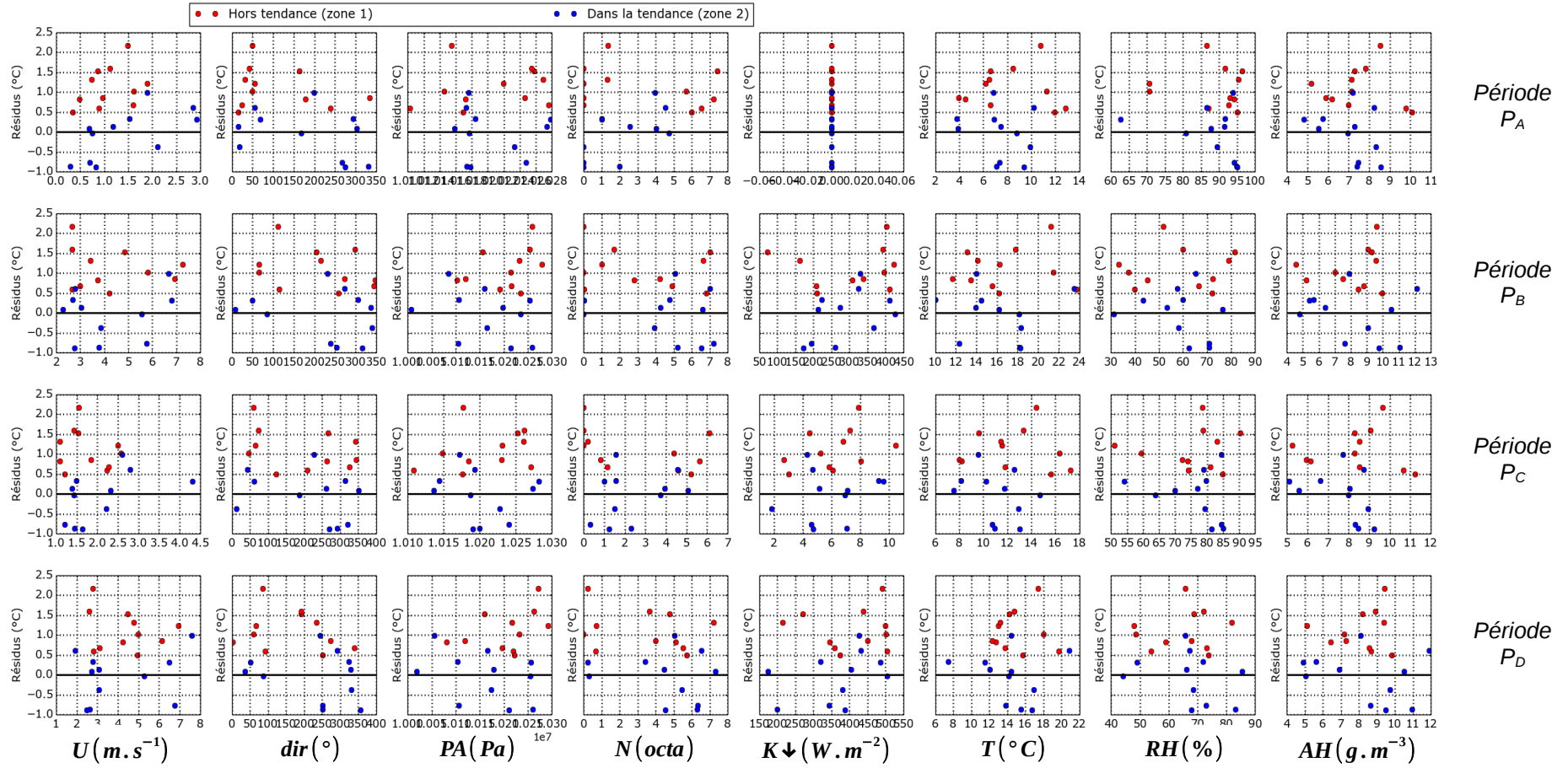


FIGURE 4.14 – Résidus du modèle  $ITL^{(pri)}$  (correspondant aux zones 1 et 2 de la figure 4.13) en fonction de chaque indicateur météorologique calculé pour chaque période  $P_m$ .

Il est difficile d'analyser les performances du modèle temporel pour le jeu de données LRY. En effet, ces performances s'évaluent en observant la linéarité de la courbe de tendance formée par le nuage de point  $ITL_{est} = f(ITL_{obs})$ . Cette pente étant très faible (due aux défaillances du modèle spatial de dispersion), il n'est pas possible de détecter une non-linéarité à l'oeil nu.

La non-linéarité du nuage de points est plus facile à détecter pour le jeu de données angevin (figure 4.6b). Elle apparaît assez clairement pour les sites A02, A03, A04 et A05. Le site A02 est choisi pour détecter d'où viennent les défaillances des modèles temporels. Comme cela a été fait pour Nantes, des zones hors tendances et dans la tendance sont identifiées figure 4.15. Comme précédemment, celles-ci correspondent à des valeurs élevées d'ICU, mais également à des valeurs faibles.

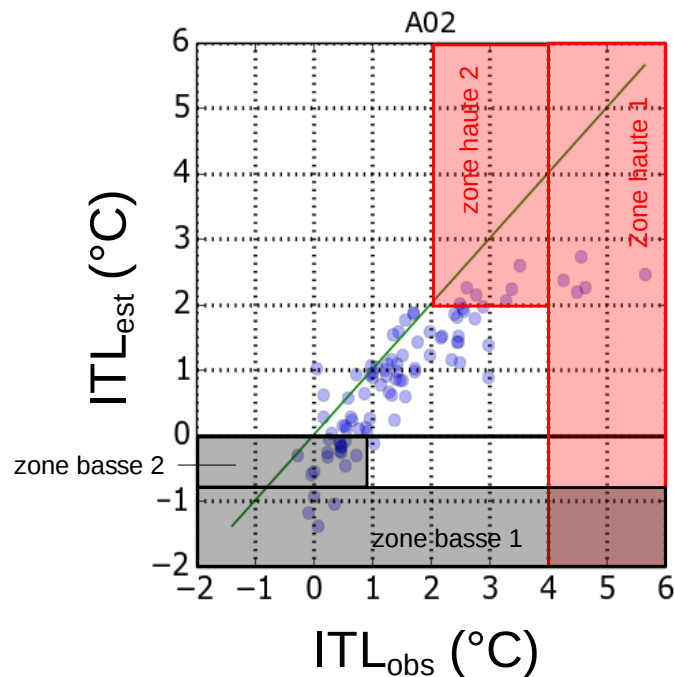


FIGURE 4.15 – Valeurs estimées de l'indicateur  $ITL$  en fonction des valeurs observées pour le site A02. La *zone haute 1* est représentative de valeurs hautes bien modélisées, en opposition à la *zone hautes 2* contenant des valeurs hautes moins bien modélisées. La *zone basse 1* est représentative de valeurs basses bien modélisées, en opposition à la *zone basse 2* contenant des valeurs basses moins bien modélisées.

L'objectif est de voir si des différences notables des conditions météorologiques sont observées entre chacune de ces deux zones. Si c'est le cas, les variables météorologiques concernées pourraient être à l'origine de la faille du modèle temporel. Les figures 4.16 et 4.17 présentent les résultats de cette analyse respectivement pour des zones hors tendance correspondant à un ICU élevé (zone haute) et à un ICU faible (zone basse).

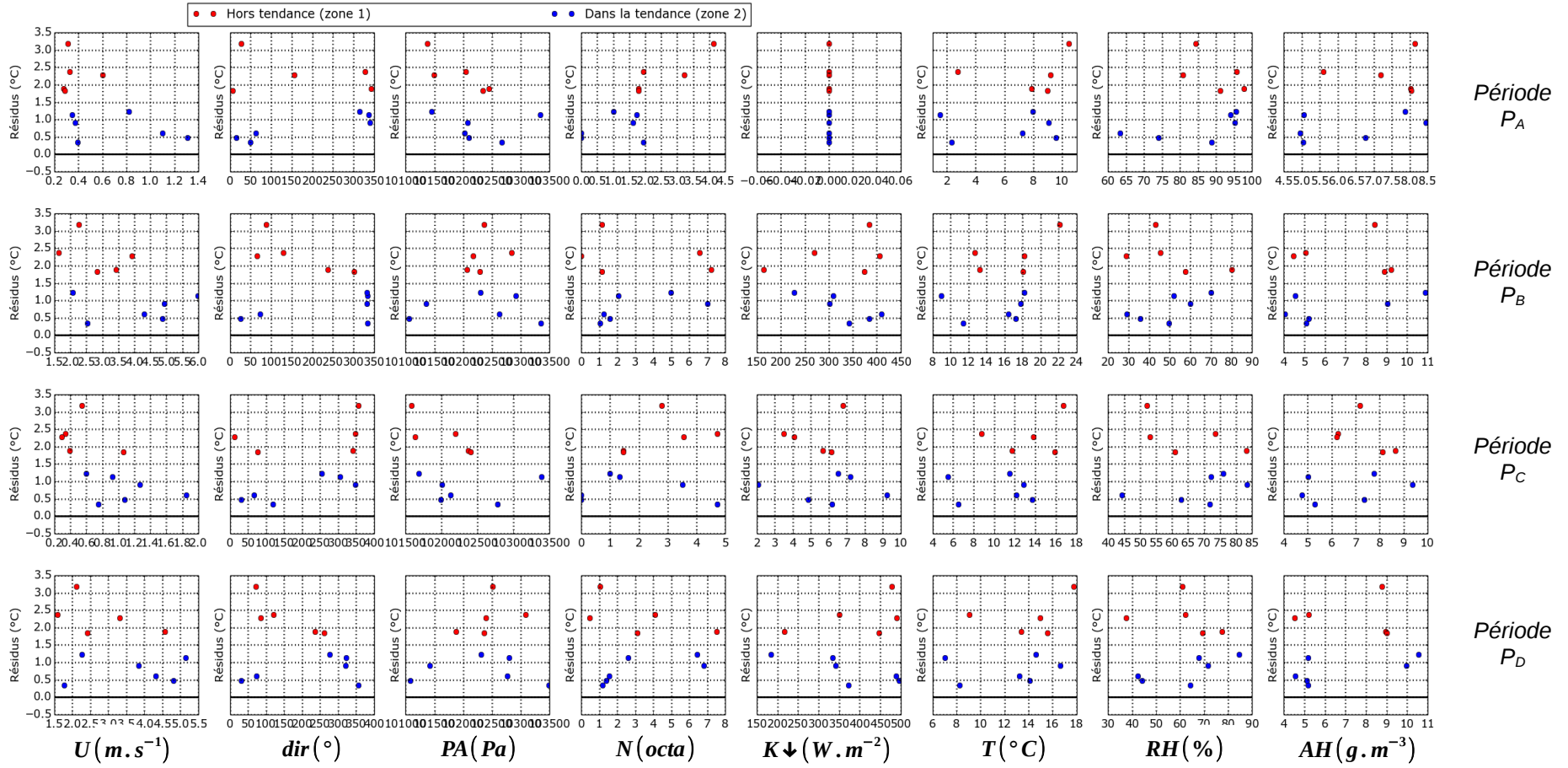


FIGURE 4.16 – Résidus du modèle  $ITL^{(pri)}$  (correspondant aux zones hautes 1 et 2 de la figure 4.15) en fonction de chaque indicateur météorologique calculé pour chaque période  $P_m$ .

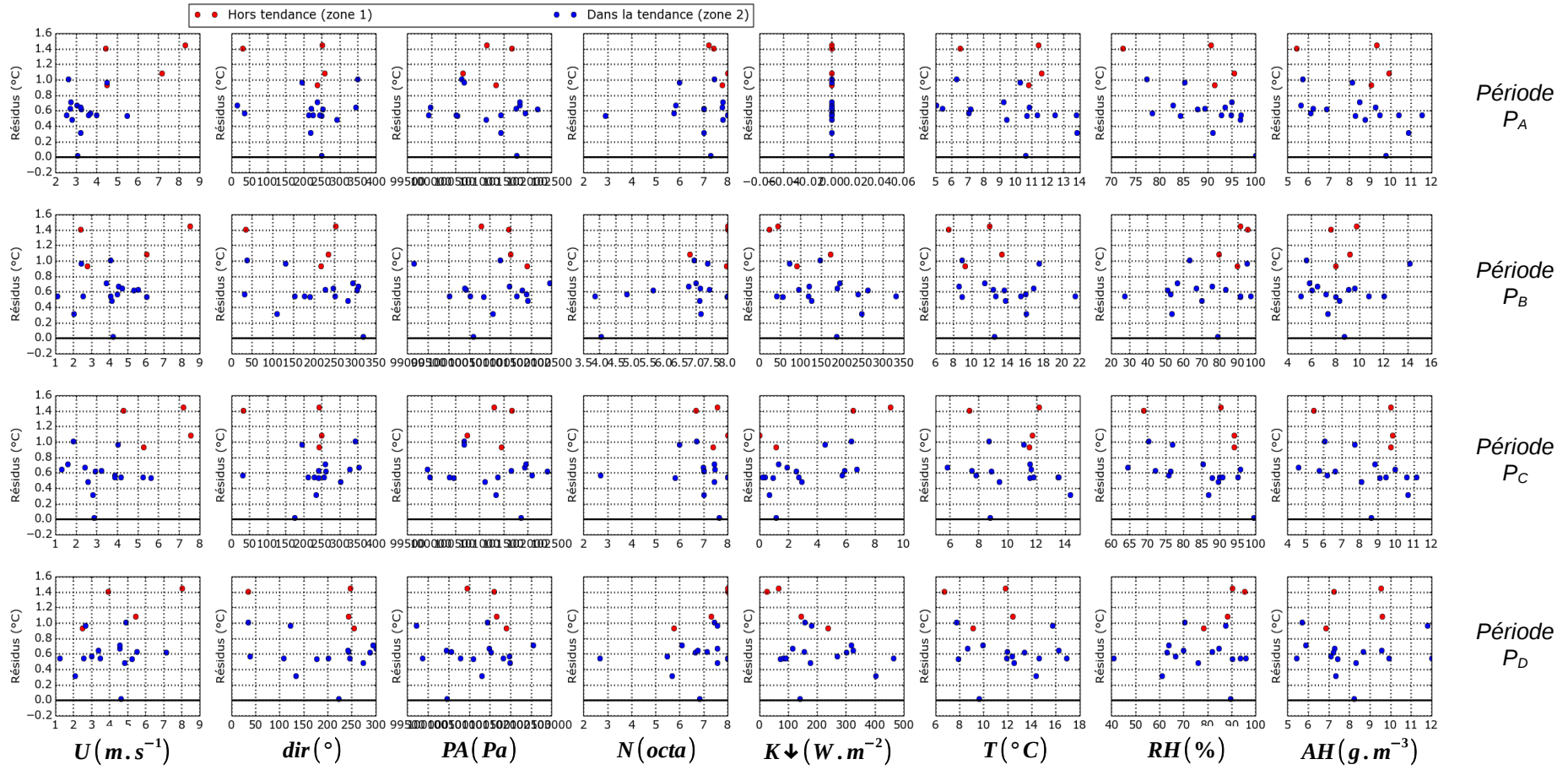


FIGURE 4.17 – Résidus du modèle  $ITL^{(pri)}$  (correspondant aux zones basses 1 et 2 de la figure 4.15) en fonction de chaque indicateur météorologique calculé pour chaque période  $P_m$ .

Pour les points hors tendance correspondant à un fort ICU, deux variables pourraient expliquer l'erreur du modèle temporel : la nébulosité pendant la période  $P_A$ , et la vitesse du vent pendant la période  $P_C$ . Au vu de la figure 4.16, l'ICU augmente linéairement avec la nébulosité. Or, la nébulosité pour le période  $P_A$  est une variable déjà présente dans l'équation 4.4 du modèle printanier. Elle est présente sous forme linéaire mais a une pente négative, c'est-à-dire contraire à celle observée figure 4.16. Il est donc possible que ce modèle ne soit pas approprié aux valeurs hautes de l'ICU, souvent associées dans la littérature à une nébulosité faible. Certains auteurs ont remarqué que l'ICU n'évoluait pas tout-à-fait linéairement avec la nébulosité. Svensson et al. (2002) utilisent quatre modèles linéaires définis pour quatre plages de valeurs de nébulosité : [0 - 2] octa, [2 - 4] octa, [4 - 6] octa, [6 - 8] octa. La pente est la plus forte pour la portion [2 - 4] octa. Dans notre cas, nous avons observé au chapitre 3 que l'ICU était presque constant pour des valeurs comprises entre 0 et 6 octa puis diminuait notablement au-dessus de 6 octa. Un modèle exponentiel avait donc été proposé. Ce modèle est d'ailleurs adopté pour l'automne (équation 4.2). Un tel modèle permettrait peut-être de mieux représenter les valeurs hautes de l'ICU puisque la pente négative de l'exponentielle serait faible dans l'intervalle [0 - 6]. Cependant, il n'a pas été retenu dans la procédure décrite figure 3.16. Cela signifie que la fonction linéaire maximisait le coefficient de corrélation. Il faut donc comprendre pourquoi. Tous les jours ont initialement le même poids dans l'élaboration du modèle temporel : les écarts de température observés, qu'ils soient élevés ou faibles, pèsent de la même façon sur la régression. Ils pèsent de façon proportionnelle au carré de leur distance à la droite de régression. Or, les valeurs extrêmes de l'ICU sont dans notre cas celles dont on souhaiterait une modélisation la plus fidèle possible. Il en résulte qu'un système de pondération doit être intégré à la méthode de régression, afin de mieux modéliser les journées où des valeurs hautes de l'ICU sont observées. À noter que dans ce cas, la qualité des estimations pour les autres journées sera sans doute revue à la baisse. De très faibles valeurs de vent pendant la période  $P_C$  peuvent également être à l'origine des faiblesses du modèle temporel pour modéliser les valeurs hautes de l'ICU (figure 4.16). La vitesse du vent est prise en compte dans le modèle mais pour la période  $P_A$ . Or, cette période ne semble pas poser de problèmes puisque des valeurs faibles de vent sont enregistrées pour la zone hors de la tendance mais également pour celle dans la tendance. L'ajout d'un deuxième terme de vitesse de vent correspondant cette fois-ci à la période  $P_C$  permettrait peut-être de corriger le problème.

Concernant les valeurs basses de l'ICU, la figure 4.17 met en évidence que la vitesse du vent est probablement la variable à l'origine de la mauvaise qualité des estimations du mo-



dèle, puisque les valeurs de vent les plus fortes correspondent aux valeurs hors de la tendance (surtout pour les périodes  $P_A$  et  $P_C$ ). Ces valeurs sont largement sous-estimées. Cela signifie que la pente du modèle inverse utilisé (équation 4.4) pour modéliser l'influence du vent sur l'ICU est trop inclinée. La constante 3 présente dans le terme de l'inverse de la vitesse du vent permet de déplacer la courbe vers des valeurs plus faibles de vent (voir figure 4.18). Cette constante a été déterminée par maximisation du coefficient de corrélation lorsque l'ICU était modélisé par la fonction inverse de la vitesse du vent en utilisant tout le jeu de données (toutes saisons confondues). Il s'agirait donc de vérifier que la valeur de cette constante est bien valable pour les données printanières. Il se peut en effet qu'elle soit plus importante et que la courbe inverse soit d'autant plus décalée vers la gauche, réduisant potentiellement la pente pour les valeurs élevées de vent. Néanmoins, cette modification réduirait la différence de pente entre les valeurs faibles et fortes de vitesse de vent (voir figure 4.18). Ceci pourrait contribuer d'autant plus à la dégradation des estimations pour les valeurs hautes de l'ICU.

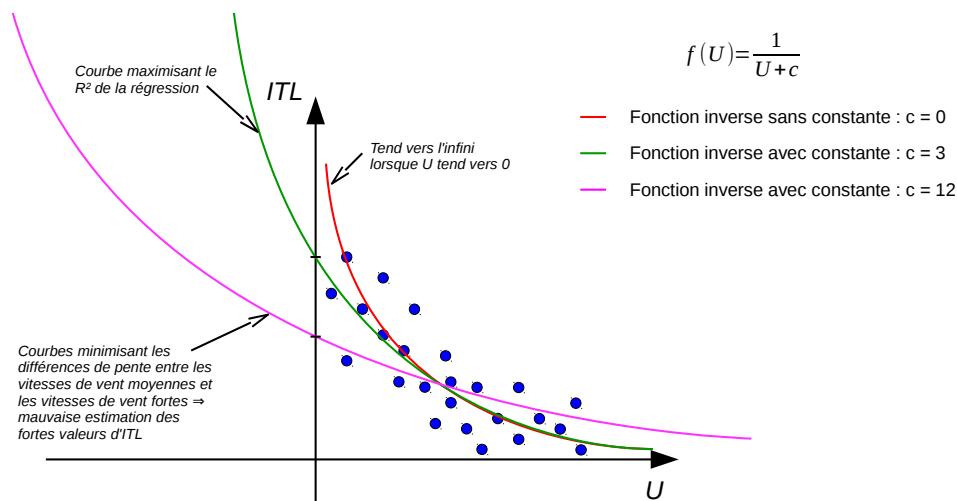


FIGURE 4.18 – Évolution de la fonction inverse  $f(U) = \frac{1}{U+c}$  pour plusieurs valeurs de constante  $c$ .

## 4.6 Améliorations à apporter

Afin d'améliorer les performances des modèles, plusieurs propositions d'améliorations sont émises, tant pour la modélisation des phénomènes spatiaux que temporels.

- Pour les modèles spatiaux, les modifications pourraient être relatives à la taille de la zone tampon ou aux variables explicatives utilisées.

- Un travail d’identification de la taille optimale de la zone tampon utilisée pour calculer les indicateurs géographiques autour de chaque site devrait être réalisée ;
  - Autoriser les termes croisés dans les équations 3.20 permettrait de donner un sens physique à chaque terme directement lié au soleil. Par exemple, l’indicateur  $NDVI$  (dont le terme est  $\alpha_{s_{NDVI}} x_{NDVI}$ , assimilé au flux de chaleur sensible), pourrait être pondéré par  $D_{fac}$  (dont le terme est  $\alpha_{s_{D_{fac}}} x_{D_{fac}}$ , terme qui définit le rayonnement solaire piégé dans la canopée).
- Pour les modèles temporels, les valeurs extrêmes sont souvent mal estimées. Pour remédier à ce problème, deux solutions sont proposées, qui devront être testées l’une après l’autre :
    1. laisser l’opportunité au modèle de contenir deux fois la même variable météorologique pour deux périodes différentes : au printemps, c’est le cas pour la nébulosité (période  $P_B$  en plus de la période  $P_A$ ) et pour la vitesse de vent (période  $P_C$  en plus de la période  $P_A$ ) ;
    2. si des méthodes de régression sont utilisées, accorder plus d’importance aux valeurs d’ICU les plus élevées en les pondérant plus fortement permettra de mieux les modéliser.

## 4.7 Conclusion du chapitre

Un ensemble de modèles spatiaux et temporels a été retenu dans ce chapitre. Il permet de calculer les différences de température nocturnes pouvant se développer au sein d’une agglomération. Les ensembles de modèles calibrés au chapitre précédent pour estimer les différences de température diurnes ont montré des performances qui n’ont pas été jugées satisfaisantes. Ceci s’explique probablement par le fait que l’intensité de l’ICU soit moins marquée le jour que la nuit. Finalement, l’ensemble de modèles nocturnes retenu est celui qui permet d’obtenir les valeurs de température d’air les plus proches des températures réellement observées pour l’ensemble des sites des six jeux de données. Les modèles spatiaux, composés d’indicateurs géographiques caractérisant le contexte géographique de chaque site de mesure, ainsi que les modèles temporels, composés d’indicateurs météorologiques caractérisant le type de temps, sont différents pour chacune des quatre saisons. La pertinence des modèles, au-delà de leurs performances, est justifiée physiquement lorsque l’on analyse les termes de chaque équation. Nous allons voir

dans le chapitre suivant comment cet ensemble de modèles peut maintenant être utilisé dans le cadre d'applications plus opérationnelles.



# 5

## Modélisation du climat urbain : exemples d'application

Les chapitres précédents ont permis de décrire la méthodologie proposée pour modéliser l'îlot de chaleur urbain, et de présenter les performances de cette modélisation. Dans ce chapitre, nous illustrons l'application du travail à partir de deux approches : la reconstitution du signal de température ou l'utilisation directe des indicateurs de température modélisés précédemment. Ces deux approches sont appliquées pour trois thématiques climatiques différentes : la modélisation microclimatique d'un quartier, la thermique du bâtiment et la biodiversité.

### 5.1 Reconstitution du signal de température de n'importe quel point du territoire

Les modèles développés dans cette thèse permettent d'estimer la température de n'importe quel point de l'agglomération mais pas à n'importe quel instant  $t$ . Les différences de température sont seulement modélisées pour les périodes  $P_{ICU}^{(jour)} = [0,5 - 0,7]^{(jour)}$  et  $P_{ICU}^{(nuit)} = [0,4 - 0,8]^{(nuit)}$  (qui correspondent respectivement aux indicateurs  $ITL^{(jour)}$  et  $ITL^{(nuit)}$ ). Afin d'avoir une estimation de la température au point  $i$  pour un instant  $t$ , il est donc nécessaire de reconstituer le signal de température sur la base de celui de la station de référence. Nous proposons dans cette

section une méthode de reconstitution ainsi qu'un exemple applicatif.

### 5.1.1 Procédure de reconstitution

Les EM proposés dans cette thèse donnent des résultats pertinents seulement pour la période nocturne. Ceci peut se comprendre en étudiant comment évolue la dispersion spatiale de la température et du taux de réchauffement au cours d'une journée moyenne. Pour cela, l'indicateur de dispersion  $std(x_{i,k})_i$  (défini équation 3.3) est utilisé. Pour rappel, l'indice  $i$  représente une station et l'indice  $k$  un pas de temps. La figure 5.1, élaborée à partir de cet indicateur et du jeu de données  $JDD_{spat\&temp}$  (le plus fiable des jeux de données nantais au regard des résultats du chapitre 2), permet de mettre en évidence deux phénomènes.

- La dispersion spatiale de la température est plus faible le jour que la nuit. La variance des écarts de température entre les stations (variable à modéliser) est donc plus élevée la nuit que le jour.
- La dispersion spatiale du taux de réchauffement est plus forte le jour que la nuit. Les écarts de température entre les stations (variable à modéliser) sont donc en moyenne plus stables la nuit que le jour.

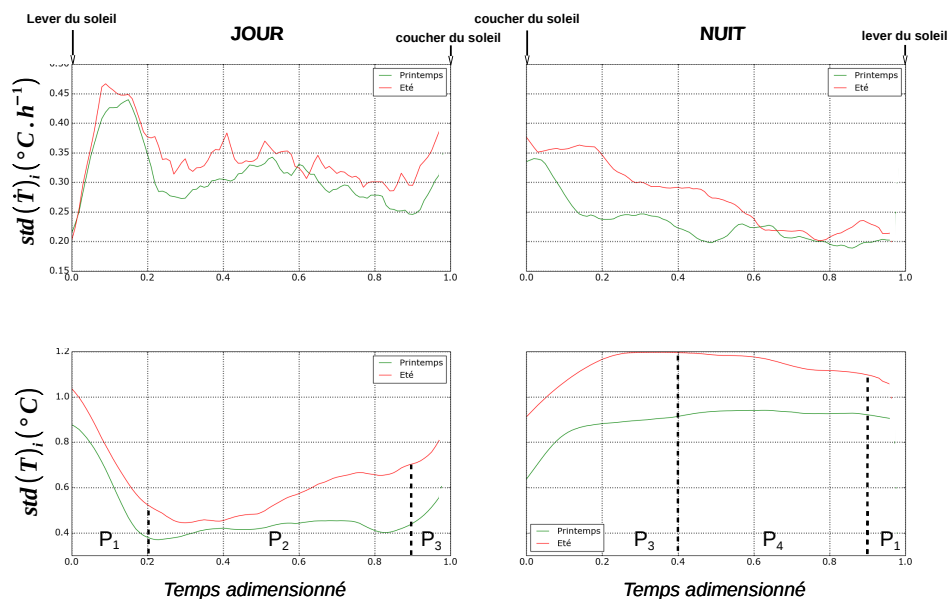


FIGURE 5.1 – Évolution (en temps adimensionné) de la dispersion interstation de la température et de celle du taux de réchauffement au cours d'une journée moyenne pour le jeu de données de vérification  $JDD_{spat\&temp}$ . L'indicateur de dispersion utilisé est l'écart-type. La figure permet d'identifier au sein d'un cycle diurne quatre périodes distinctes :  $P_1$ ,  $P_2$ ,  $P_3$  et  $P_4$ .

Quatre périodes peuvent être distinguées au sein d'un cycle diurne à partir de la figure 5.1. Pendant chacune d'elles, l'écart de température entre deux points de la ville évolue de façon particulière. La figure 5.2 schématise l'évolution temporelle de la température d'un point  $i$  situé dans la zone urbaine ainsi que de la station de référence  $ref$  située en périphérie de la ville.

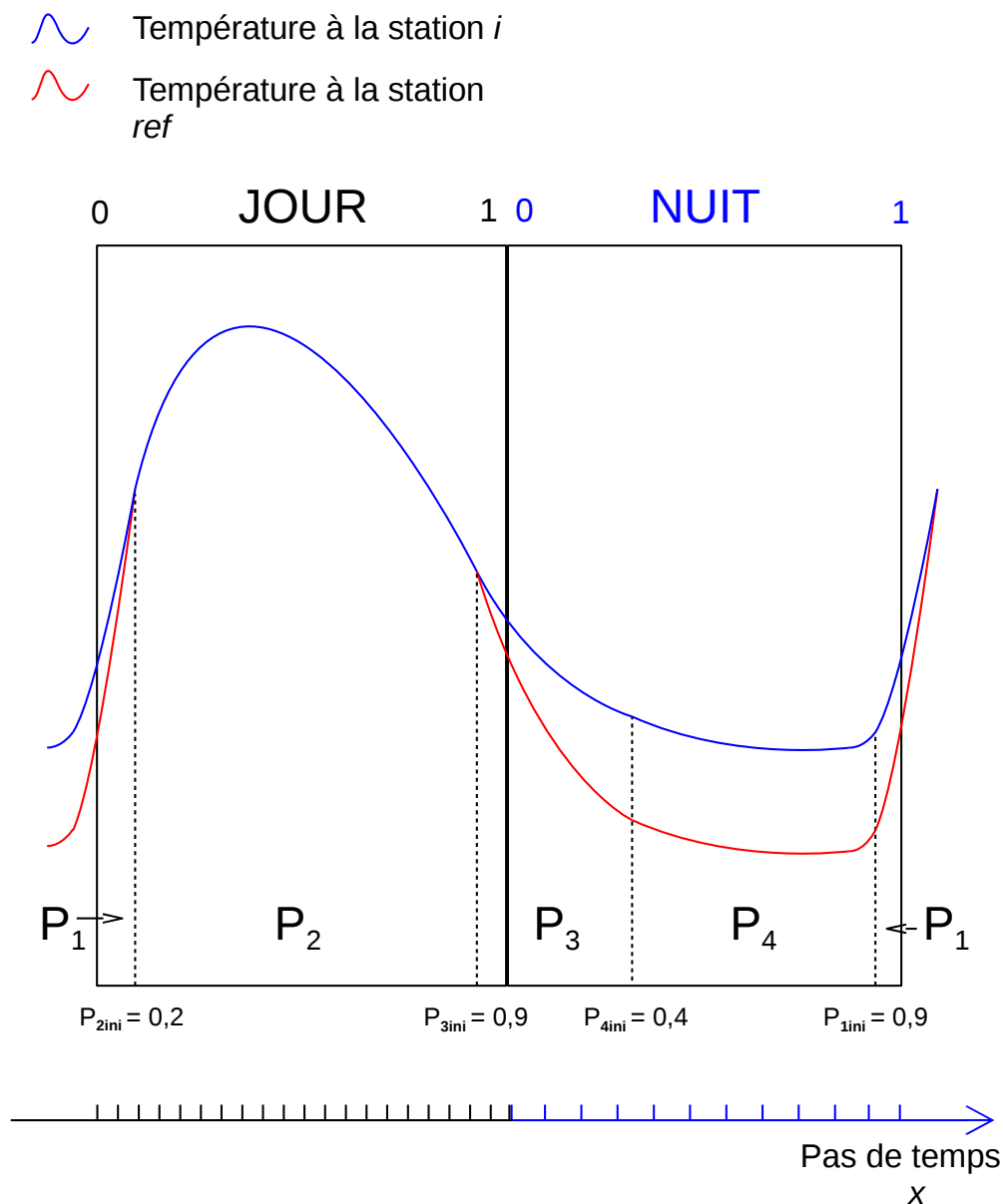


FIGURE 5.2 – Représentation schématique de l'évolution au cours d'un cycle diurne de la température de l'air pour une station  $i$  et d'une station de référence  $ref$ .

Pendant  $P_2 = [0,2 - 0,9]^{(jour)}$ , la dispersion spatiale des températures est faible et évolue peu. De plus, les résultats de cette thèse n'ont pas permis de modéliser de manière satisfaisante les différences de température entre un point  $i$  du territoire et une station de référence puisque l'indicateur  $RMSE_{norm}$  (représentatif des performances d'un EM) était pour l'indicateur diurne  $ITL^{(jour)}$  toujours supérieur à 1. Ceci signifie qu'il est préférable de considérer

l'agglomération comme homogène en température pendant cette période plutôt que d'utiliser les modèles. On considère donc que  $T_{i,k \in P_2}(j) = T_{ref,k \in P_2}(j)$  (avec  $j$  un jour de l'année). La période  $P_4 = [0,4 - 0,95]^{(nuit)}$  est l'union de la période  $P_{ICU}^{(nuit)}$  ( $= [0,4 - 0,8]^{(nuit)}$ ), utilisée pour le calcul de l'indicateur diurne  $ITL^{(nuit)}$ , et de la période  $[0,8 - 0,95]^{(nuit)}$ . La dispersion des températures évolue peu entre ces deux périodes. Les estimations de  $ITL^{(nuit)}$  seront donc allouées à la période  $P_4$ . La période  $P_1$ , constituée des intervalles  $[0,95 - 1]^{(nuit)}$  et  $[0 - 0,2]^{(jour)}$ , est la période pendant laquelle la température de l'air s'homogénéise à l'échelle de l'agglomération. À la fin de cette période, n'importe quel point  $i$  de l'agglomération est donc à la température de référence. La période  $P_3$ , constituée des intervalles  $[0,9 - 1]^{(jour)}$  et  $[0 - 0,4]^{(nuit)}$ , est l'intervalle de temps pendant lequel les différences de température entre un site  $i$  et le site de référence vont se creuser. Les écarts de température  $\Delta T_{i,k \in P_p} = T_{i,P_p} - T_{ref,P_p}$  pour  $P_p = P_1, P_2, P_3, P_4$  sont régis respectivement par les équations 5.1, 5.2, 5.3 et 5.4 :

$$T_{i,k \in P_1}(j) = T_{i,k \in P_1}(j-1) + \dot{T}_{i,k \in P_1}(j) \Delta t \quad (5.1)$$

$$T_{i,k \in P_2}(j) = T_{ref,k \in P_2}(j) \quad (5.2)$$

$$T_{i,k \in P_3}(j) = T_{i,k \in P_3}(j-1) + \dot{T}_{i,k \in P_3}(j) \Delta t \quad (5.3)$$

$$T_{i,k \in P_4}(j) = T_{ref,k \in P_4}(j) + \Delta T_{i,k \in P_4} \quad (5.4)$$

avec  $\dot{T}_{i,k \in P_p}(j)$  le taux de réchauffement (en °C/s) de la station  $i$  le jour  $j$  au pas de temps  $k$  lorsque  $k$  est compris dans la période  $P_p$  et  $\Delta t$  la valeur d'un pas de temps (en s).

Le taux de réchauffement  $\dot{T}_{i,k}$  pour chaque pas de temps  $k$  des périodes  $P_1$  et  $P_3$  est supposé proportionnel au taux de réchauffement  $\dot{T}_{ref,k}$  de la station de référence. Il est calculé respectivement à partir des équations 5.5 (période  $P_1$ ) et 5.6 (période  $P_3$ ). À noter que la variable  $j$  n'est pas conservée pour alléger l'écriture des équations, qui restent toutefois valable pour n'importe quel cycle diurne  $j$  de l'année.

$$\dot{T}_{i,k \in P_1} = \dot{T}_{ref,k \in P_1} \cdot \frac{T_{ref,P_{2ini}} - (T_{ref,P_{1ini}} + \Delta T_{i,k \in P_4})}{T_{ref,P_{2ini}} - T_{ref,P_{1ini}}} \quad (5.5)$$

$$\dot{T}_{i,k \in P_3} = \dot{T}_{ref,k \in P_3} \cdot \frac{T_{ref,P_{4ini}} - (T_{ref,P_{3ini}} + \Delta T_{i,k \in P_4})}{T_{ref,P_{4ini}} - T_{ref,P_{3ini}}} \quad (5.6)$$

avec  $\dot{T}_{ref,k \in P_p}$  le taux de réchauffement (en °C/s) de la station  $ref$  au pas de temps  $k$  lorsque  $k$  est compris dans la période  $P_p$ ,  $T_{ref,P_{1ini}}$ ,  $T_{ref,P_{2ini}}$ ,  $T_{ref,P_{3ini}}$  et  $T_{ref,P_{4ini}}$  la température (en °C) de la station de référence respectivement au premier pas de temps de chacune des périodes  $P_1$ ,

$P_2$ ,  $P_3$  et  $P_4$  et  $\Delta T_{i,k \in P_4}$  la différence de température entre une station  $i$  et la station de référence (correspondant à l'indicateur  $ITL^{(nuit)}$ , calculé à partir de l'EM finalement adopté au chapitre 4 - équations 4.2 à 4.13 et équation 3.9).

Afin de simplifier l'écriture de ces équations, on pose  $\theta_{i,k \in P_1}$  et  $\theta_{i,k \in P_3}$  tels que :

- $\dot{T}_{i,k \in P_1} = \dot{T}_{ref,k \in P_1} \cdot \theta_{i,k \in P_1}$  ;
- $\dot{T}_{i,k \in P_3} = \dot{T}_{ref,k \in P_3} \cdot \theta_{i,k \in P_3}$ .

La procédure suivie pour reconstituer les signaux de température est illustrée figure 5.3. Deux variables temporelles sont différenciées :  $j$  indique la journée utilisée tandis que  $k$  indique le pas de temps auquel on s'intéresse dans cette journée.



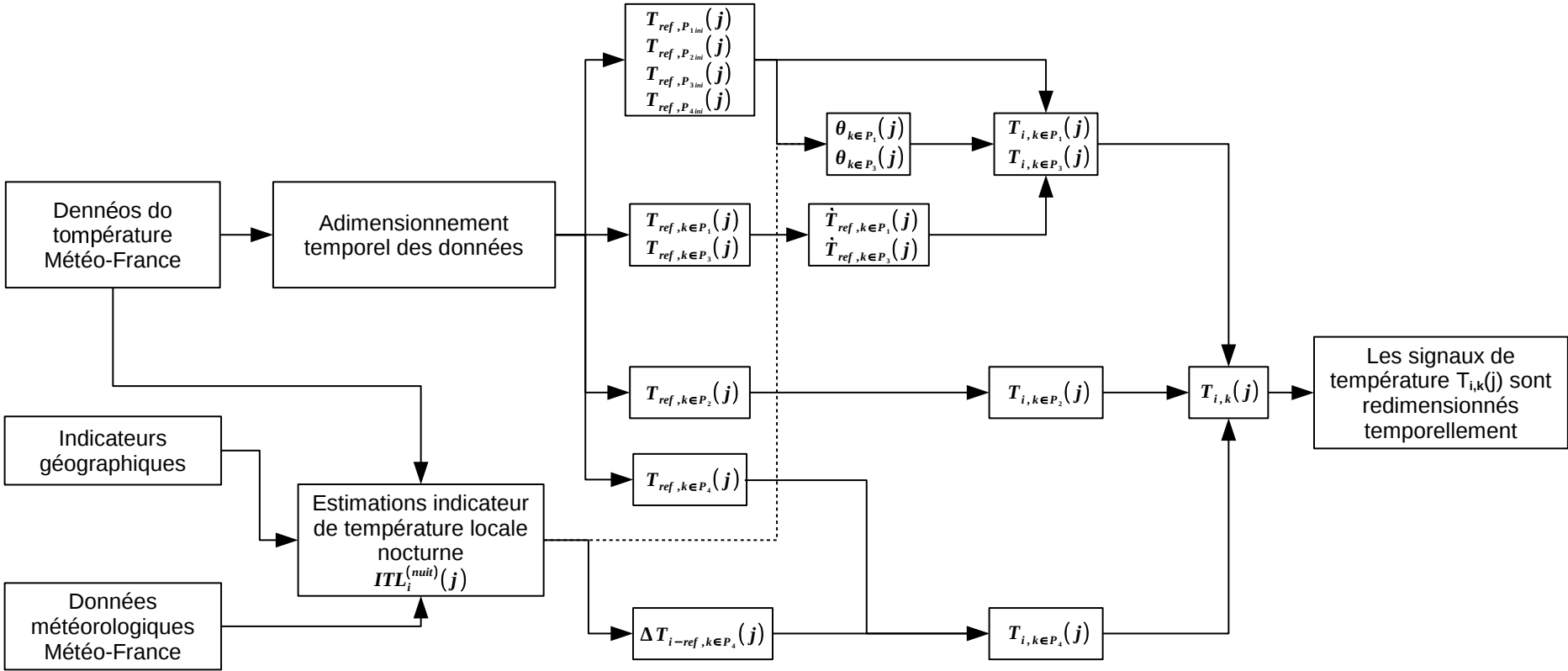


FIGURE 5.3 – Procédure utilisée pour reconstituer à un pas de temps  $k$  et pour un ensemble de journées  $j$  donnés le signal de température de n'importe quel point  $i$  du territoire.

### 5.1.2 Application : simulation numérique du microclimat d'un quartier

Certains logiciels permettent de simuler numériquement l'évolution temporelle des conditions météorologiques à l'échelle d'un quartier. Ils permettent d'évaluer la pertinence d'un projet urbain d'un quartier en termes de confort ou de besoins énergétiques. Pour fonctionner, ils ont besoin de données d'entrée telles que l'évolution temporelle de la température de l'air pendant la période simulée. Trois solutions existent pour obtenir cette donnée, chacune dotée de ses avantages et de ses inconvénients.

La solution 1 consiste à utiliser un fichier météorologique issu de la station Météo-France la plus proche. Ces données présentent l'avantage d'être simples à récupérer et sont souvent disponibles et continues pour de nombreuses années. Cependant, les stations Météo-France sont généralement situées à l'extérieur de la ville et ne sont donc pas concernées par l'îlot de chaleur urbain. Si le quartier à étudier est situé au sein d'une agglomération densément bâtie, la température réelle du quartier a donc de fortes probabilités d'être sous-estimée.

La solution 2 consiste à mesurer localement la température pendant la période à simuler. Contrairement à la première méthode, ceci permet de prendre en compte l'effet d'ICU mais implique que le dispositif de mesure soit placé à un emplacement représentatif du microclimat du site à caractériser. Il faut également que ce dispositif de mesure soit constitué d'un abri ventilé de qualité afin de ne pas sur-estimer les températures en période diurne. Par rapport à la première solution (station Météo-France), la température sera donc probablement plus juste la nuit mais potentiellement fausse le jour. Il est également souvent complexe d'installer des dispositifs de mesure sur la voie publique. Le temps passé à choisir le bon matériel de mesure, la position idéale au sein du site et les délais liés à l'obtention d'une autorisation pour installer ce matériel en zone urbaine peuvent constituer des freins à l'utilisation de cette solution. Par ailleurs, cette approche ne sera pas utilisable pour simuler des évolutions climatiques issues des scénarios climatiques produits par le GIEC, ni pour comparer l'influence microclimatique de plusieurs scénarios d'aménagement au stade du projet.

La solution 3 consiste à utiliser des données de température issues de sorties de modèles climatiques d'échelles supérieures. Plusieurs outils permettent de modéliser le climat urbain à l'échelle de l'agglomération. La température peut ainsi être estimée à n'importe quel instant  $t$  pour le quartier qui nous intéresse à partir de la température de la maille qui lui est associée (résolution variable souvent supérieure à 100 m selon l'échelle horizontale). L'avantage de cette méthode est qu'elle considère l'influence du microclimat qui se développe localement. L'inconvénient majeur est le temps de mise en œuvre d'un tel outil : il est nécessaire de préparer

les données d'entrée des modèles (collecte et traitements) et de réaliser les calculs. Ces deux étapes peuvent prendre un temps plus ou moins long selon les outils utilisés.

La méthode de calcul de l'ICU développée dans cette thèse est un bon compromis. L'EM retenu au chapitre 4 et la procédure de reconstitution du signal décrite précédemment permettent en effet de calculer la température à n'importe quel instant  $t$  et pour n'importe quel point de l'agglomération.

Pour illustrer les performances de chacune des trois solutions, elles sont chacune appliquée au quartier où se trouve la station de mesure TRH06. Le contexte géographique de ce quartier est illustré figure 5.4. Celui-ci est globalement peu dense et très végétalisé comparé à certains quartiers du centre-ville de Nantes.

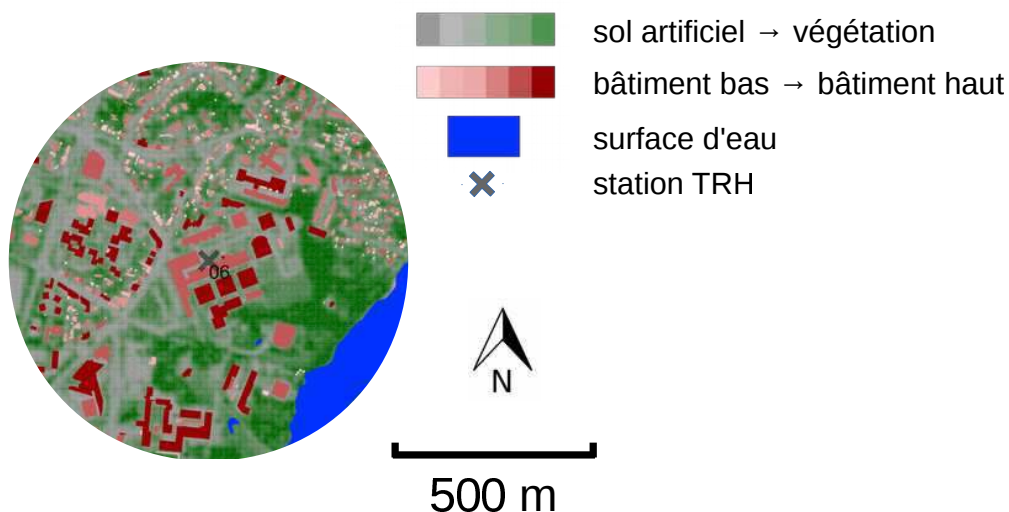


FIGURE 5.4 – Contexte géographique du site TRH06

Pendant le mois de juillet 2015, la température de l'air a été mesurée au sein de ce site sous deux abris : l'abri TRH qui a montré de mauvaises performances (chapitre 2), et l'abri URBIO, qui a montré de bonnes performances même lorsqu'il était soumis à un rayonnement solaire important. La température mesurée sous l'abri TRH illustrera les limites d'une mesure locale utilisant un bon emplacement mais un abri peu fiable (solution 2). La température mesurée à Météo-France Nort-sur-Erdre est utilisée comme solution 1 et la méthode développée dans cette thèse comme solution 3. Les données de l'abri URBIO (qualifié de fiable), constitueront les données de référence, qu'il faudrait idéalement considérer comme la température d'entrée d'une simulation numérique. Trois jours particulièrement chauds de la campagne de mesure (présentée en section 2.4.2.2) sont utilisés pour illustrer les différences de performance entre les trois solutions. Le signal de température obtenu pour chaque solution est comparé à

une température de référence figure 5.5. Cette température de référence est celle mesurée sous l'abri URBIO, très peu sensible au rayonnement solaire (chapitre 2).

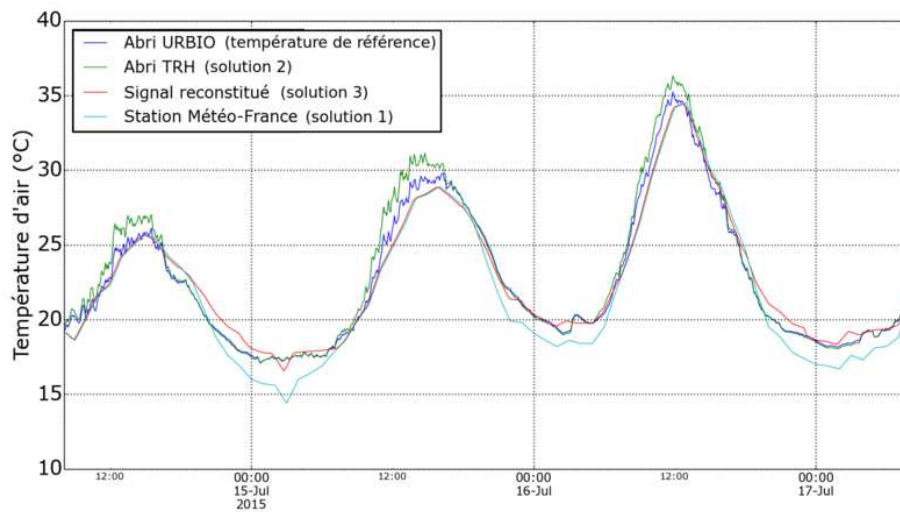


FIGURE 5.5 – Évolution de la température d'air d'un quartier estimée à partir des solutions 1, 2 et 3.

La température sous l'abri TRH suit très bien la dynamique de la température de référence, malgré une légère sur-estimation en première partie de journée. Le signal reconstitué est également proche du signal de référence, notamment entre 8h et 9h du matin. La température Météo-France est au plus proche entre 8h et 9h également mais est relativement éloignée le reste du temps. Les plus grosses différences de température apparaissent comme prévu la nuit. La performance de chaque solution est quantifiée par la racine carrée de l'écart quadratique moyen (RMSE) entre la température estimée par une solution donnée et la température de référence. Selon cet indicateur, la méthodologie développée pendant cette thèse est aussi performante qu'utiliser l'abri TRH (la valeur de RMSE est respectivement égale à 0,69 °C et 0,67 °C). La solution qui consiste simplement à prendre les données Météo-France arrive largement en dernière position (RMSE = 1,14 °C). Le fait d'utiliser une période particulière de l'année (mois de juillet) ainsi qu'un quartier particulier (peu densément bâti) ne permet pas de conclure qu'une solution est meilleure qu'une autre. Ces différences de performances sont donc propres à notre exemple. Dans le cas où le quartier choisi serait plus densément bâti :

- l'erreur engendrée par l'utilisation des données Météo-France pourrait s'avérer encore plus élevée ;
- les ombres portées des bâtiments masqueraient d'autant plus l'abri TRH du rayonnement solaire direct, ce qui lui permettrait d'approcher d'autant mieux la température réelle du

quartier.

## 5.2 Spatialisation des résultats : cartes climatiques

Les informations géographiques sont utilisées au quotidien pour gérer les territoires. Représenter une information sur un fond de carte du territoire peut permettre d'identifier des différences entre les quartiers (différences d'ambiance climatique, d'accès aux espaces verts, de densité de population, etc.). Si ces différences ne sont pas souhaitables, un plan d'action doit être élaboré pour y remédier. Cette section propose de réaliser deux types de cartes climatiques permettant de diagnostiquer le territoire à partir d'indicateurs résultant de la modélisation du climat urbain. Chacun de ces indicateurs présente un intérêt particulier dans la gestion de la ville. Une première partie traitera des différences spatiales des consommations énergétiques entre bâtiments situés sur le territoire de la ville de Nantes. La seconde portera sur la prévention des risques allergiques liés aux pollens.

### 5.2.1 Îlot de chaleur urbain et modification des besoins de chaleur et de froid

L'indicateur Degrés-Jours-Unifiés ( $DJU$ ) est utilisé pour caractériser le climat d'une zone en vue de calculer les consommations de chauffage ou de climatisation d'un bâtiment. En effet, les besoins de chaleur  $B_{chaleur}$  d'un bâtiment peuvent être estimés en multipliant les  $DJU$  par une conductance thermique (notée  $UA$ ), représentative du flux thermique qui peut passer à travers l'enveloppe d'un bâtiment lorsqu'il est soumis à une différence de température de  $1^\circ\text{C}$  entre l'air intérieur et l'air extérieur. Ces besoins sont exprimés par la relation 5.7 (à noter que la même formulation peut être réalisée pour les besoins de rafraîchissement en remplaçant les variables  $B_{chaleur}$  et  $DJU_{chaleur}$  respectivement par  $B_{froid}$  et  $DJU_{froid}$ ).

$$B_{chaleur} = UA DJU_{chaleur} \quad (5.7)$$

avec  $B_{chaleur}$  en  $\text{W}\cdot\text{jr}$ ,  $UA$  en  $\text{W}/^\circ\text{C}$  et  $DJU_{chaleur}$  en  $^\circ\text{C}\cdot\text{jr}$

Selon ce modèle simplifié, les besoins de chaleur ou de froid d'un bâtiment sont donc directement proportionnels aux  $DJU$  d'une zone climatique. Il faut cependant noter que cette méthode ne prend pas en compte les apports solaires. Pour cette raison, entre autres, elle n'est souvent utilisée qu'en première approche. Pour calculer les  $DJU$ , les températures d'air mini-

minimales et maximales journalières de la zone étudiée (respectivement  $T_{min}(j) = \min(T_k(j))_k$  et  $T_{max}(j) = \max(T_k(j))_k$ ) doivent être connues pour chaque jour  $j$  de la période d'intérêt (période de chauffage ou de climatisation). La température moyenne  $T_{moy}(j)$  est également définie par l'équation  $T_{moy}(j) = \frac{T_{max}(j)+T_{min}(j)}{2}$ . Pour chaque jour  $j$ , on calcule le nombre de degré jours  $DJ(j)$  par rapport à une température de référence  $T_{seuil}$  en utilisant les équations suivantes (Météo-France 2005) :

- pour calculer des besoins de chauffage ;
  - si  $T_{seuil} > T_{max}(j) : DJ(j) = T_{seuil} - T_{moy}(j)$ ,
  - si  $T_{seuil} \leq T_{min}(j) : DJ(j) = 0$ ,
  - si  $T_{min}(j) < T_{seuil} \leq T_{max}(j) : DJ(j) = (T_{seuil} - T_{min}(j))(0,08 + 0,42 \frac{T_{seuil} - T_{min}(j)}{T_{max}(j) - T_{min}(j)})$ ,
- pour calculer des besoins de climatisation ;
  - si  $T_{seuil} > T_{max}(j) : DJ(j) = 0$ ,
  - si  $T_{seuil} \leq T_{min}(j) : DJ(j) = T_{moy}(j) - S$ ,
  - si  $T_{min}(j) < T_{seuil} \leq T_{max}(j) : DJ(j) = (T_{max}(j) - T_{seuil})(0,08 + 0,42 \frac{T_{max}(j) - T_{seuil}}{T_{max}(j) - T_{min}(j)})$ .

Les  $DJ(j)$  sont ensuite cumulés sur toute la période de chauffage ou de climatisation pour calculer l'indicateur  $DJU$ . L'équation 5.8 est ainsi utilisée :

$$DJU = \sum_j DJ(j) \quad (5.8)$$

La période de chauffage sera pour notre exemple définie entre le 1er octobre et le 20 mai (bornes traditionnellement utilisées en France métropolitaine). La température de référence  $T_{seuil_{chauffeur}}$  est fixée à 18°C. La période de climatisation sera définie entre le 21 mai et le 30 septembre. La température de référence  $T_{seuil_{froid}}$  sera fixée à 26°C.

Les données de température minimale et maximale traditionnellement utilisées pour calculer les  $DJU$  sont celles issues des stations Météo-France. Ceci revient à considérer que tout le territoire de l'agglomération a les mêmes minima et maxima de température. En réalité, les  $DJU$  des zones urbanisées sont sans doute sur-estimés en hiver et sous-estimés en été. Afin de quantifier ces erreurs d'estimations, les modèles décrits au chapitre 4 seront utilisés. Nous avons en effet vu section 3.1.1.4 que :

- $ITL^{(jour)}$  pouvait être assimilé à l'écart de température entre une station  $i$  et la station de référence  $ref$  lorsque la température atteignait son maximum journalier  $T_{max}(j)$  ;

- $ITL^{(nuit)}$  pouvait être assimilé à l'écart de température entre une station  $i$  et la station de référence  $ref$  lorsque la température atteignait son minimum journalier  $T_{min}(j)$ .

L'ensemble de modèles qui permet de calculer les écarts diurnes de température  $ITL^{(jour)}$ , contrairement à celui qui permet le calcul des écarts nocturnes  $ITL^{(nuit)}$ , donne des résultats peu fiables. Nous considérerons donc que la température maximale observée à un point  $i$  de l'agglomération est égale à celle observée à la station de référence ( $T_{max_i}(j) = T_{max_{ref}}(j)$ ). La température minimale  $T_{min_i}(j)$  observée à un point  $i$  sera quant à elle calculée à partir de l'EM qui permet de calculer l'indicateur nocturne  $ITL^{(nuit)}$  ainsi que de la température minimale  $T_{min_{ref}}(j)$  observée à la station Météo-France selon l'équation 5.9.

$$T_{min_i}(j) = T_{min_{ref}}(j) + ITL_{nuit_i}(t) \quad (5.9)$$

Il est donc possible de calculer les besoins relatifs de chaleur  $B_{rel_{chaleur_i}}$  entre un bâtiment situé à n'importe quel point  $i$  du territoire et un bâtiment situé sur le site de référence Météo-France :

$$B_{rel_{chaleur_i}} = 100 \cdot \frac{DJU_{chaleur_i}}{DJU_{chaleur_{ref}}} \quad (5.10)$$

avec  $B_{rel_{chaleur}}$  les besoins de chaleur pour le quartier  $i$  relativement à ceux du site de référence  $ref$  (en %),  $DJU_{chaleur_{ref}}$  les degrés-jours-unifiés de chauffage du site Météo-France et  $DJU_{chaleur_i}$  ceux d'un site  $i$ .

De la même manière, les besoins relatifs de climatisation  $B_{rel_{froid_i}}$  peuvent être calculés :

$$B_{rel_{froid_i}} = 100 \cdot \frac{DJU_{froid_i}}{DJU_{froid_{ref}}} \quad (5.11)$$

avec  $B_{rel_{froid}}$  les besoins de froid pour le quartier  $i$  relativement à ceux du site de référence  $ref$  (en %),  $DJU_{froid_{ref}}$  les degrés-jours-unifiés de climatisation du site Météo-France et  $DJU_{froid_i}$  ceux d'un site  $i$ .

Les équations 5.10 et 5.11 sont utilisées pour calculer respectivement les besoins relatifs de chaleur et de fraîcheur de 855 points du territoire de la ville de Nantes. Ces 855 points sont initialement régulièrement répartis. Un point est ensuite considéré non-valide s'il est placé sur l'eau ou sur un bâtiment. Lorsque c'est le cas, il est déplacé au point valide le plus proche. La figure 5.6 représente la localisation géographique des 855 points finalement utilisés.

Avant le calcul des besoins de chaud et de froid,  $ITL^{(nuit)}$  doit être calculé pour cha-

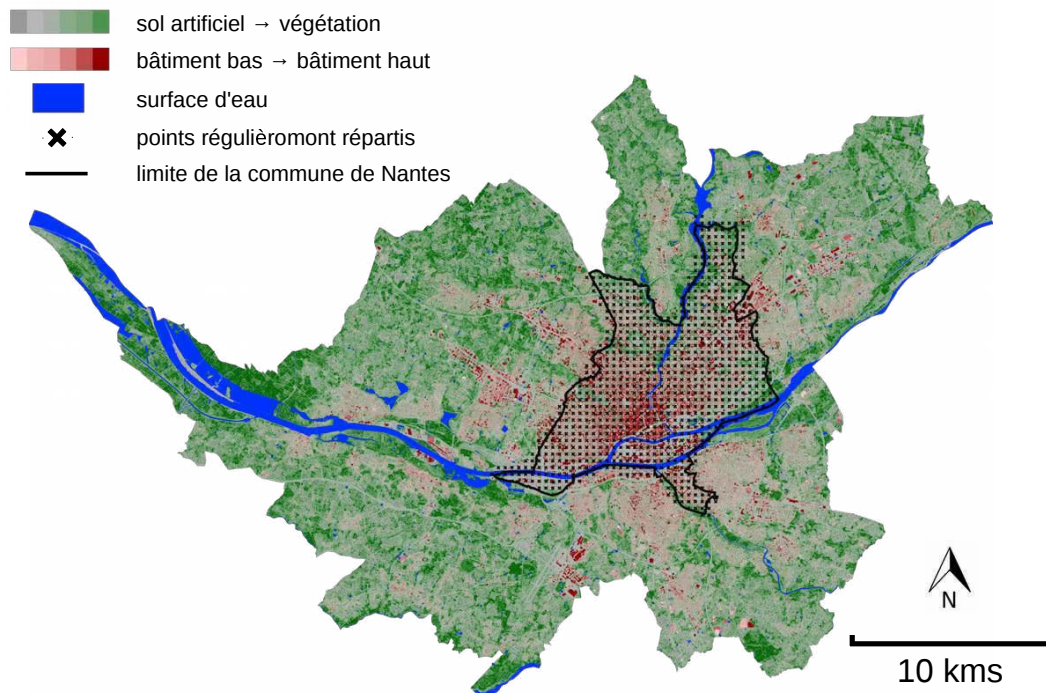


FIGURE 5.6 – Localisation des 855 points répartis sur le territoire de la ville de Nantes

chacun des 855 points à partir des modèles spatiaux et temporels retenus (chapitre 4). La période de chauffe de l'hiver 2014-2015 et celle de climatisation de l'été 2015 sont considérées pour les calculs. Une interpolation spatiale des résultats de consommations énergétiques permet de ne pas se restreindre aux 855 points de calcul et donc d'avoir une représentation continue des résultats pour l'ensemble du territoire de la ville de Nantes. Celle-ci est réalisée à partir du plugin *Interpolation* du logiciel QGIS. La méthode utilisée est l'interpolation par triangulation irrégulière. Celle-ci repose sur la triangulation de Delaunay et une méthode d'interpolation linéaire qui sont illustrées figure 5.7. La triangulation de Delaunay consiste à construire un réseau de triangles qui relie des points  $i$  de mesure entre eux. Chaque sommet de triangle correspond à un point  $i$ . Le réseau créé a pour contrainte de maximiser le plus petit angle de l'ensemble des triangles qui le constituent. Les triangles les plus *allongés* sont donc proscrits. En deux dimensions, un réseau de triangles est appelé *triangulation de Delaunay* seulement si aucun sommet de triangle n'est placé à l'intérieur d'un cercle circonscrit (figure 5.7a). Soit  $X_{mes_i}$  la grandeur physique mesurée à chacun des points  $i$ . La valeur  $X_{int_m}$  d'un point  $m$  situé dans un triangle ( $Tri_m$ ) du réseau est ensuite calculée par interpolation linéaire. Cela consiste, dans un espace à trois dimensions (où la grandeur physique à interpoler  $X$  correspond à la troisième dimension), à déterminer la coordonnée  $X_m$  du point  $m$  à partir de l'équation du plan ( $PL_m$ ) correspondant à la facette triangulaire  $Tri_m$  (figure 5.7b).



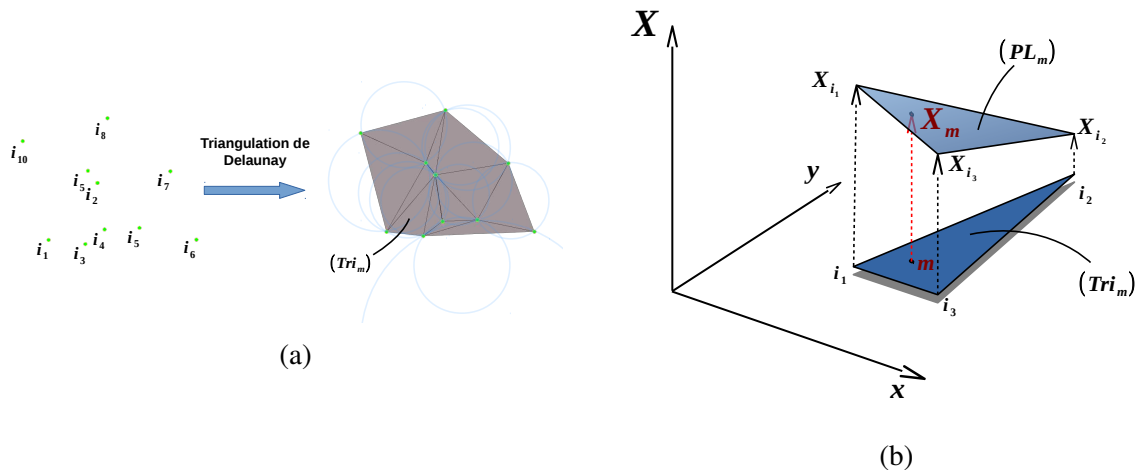
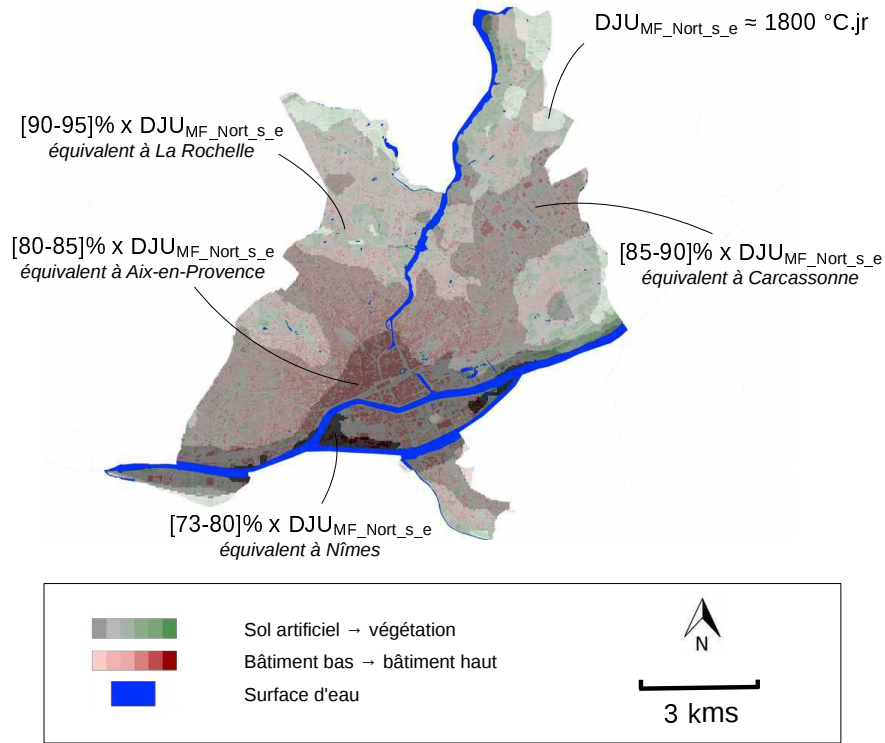


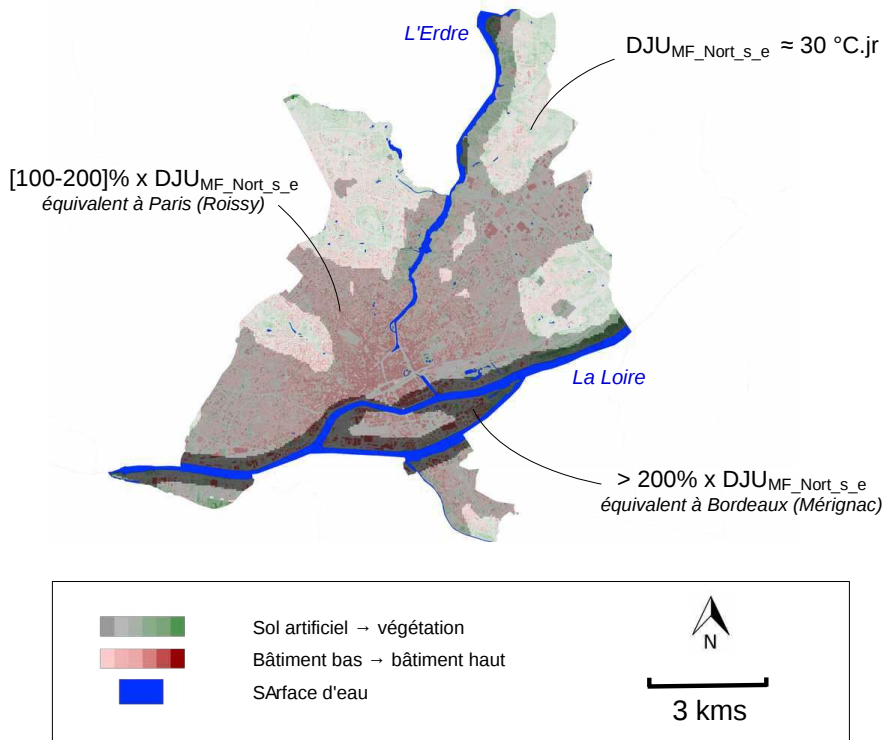
FIGURE 5.7 – (a) Triangulation de Delaunay et (b) interpolation linéaire appliquée à une série de points de mesure d'une grandeur physique  $X$

Les  $DJU$  de chaleur et de climatisation calculés pour les 855 points du territoire nantais sont interpolés linéairement pour toute la zone d'étude en appliquant cette méthode pour  $X = B_{rel}$ . Un maillage de cellules carrées de taille  $40 \times 40m^2$  est utilisé. Le centre de chaque cellule correspond à un point  $m$  pour lequel les besoins sont interpolés (figure 5.7b). La figure 5.8a représente les besoins relatifs de chaleur et la figure 5.8b les besoins relatifs de climatisation. Pour mieux situer les ordres de grandeur obtenus, une ville dont les températures sont plus élevées que celles de Nantes est associée à chaque zone. Cette association est réalisée après calcul des  $DJU$  de plusieurs villes françaises à partir de l'outil en ligne de CEGIBAT (Cegibat 2016) pour les données Météo-France de l'année 2015 (hiver 2014-2015 pour les  $DJU_{chaleur}$  et été 2015 pour les  $DJU_{froid}$ ). Seule une zone n'est pas recouverte d'un voile grisé. Elle correspond aux parties du territoire dont les  $DJU$  sont équivalents à ceux du site utilisé comme référence pour qualifier le climat de l'agglomération (ici la station Météo-France de Nort-sur-Erdre). Cette façon de présenter les résultats permet de rendre compte des modifications microclimatiques engendrées par la ville : *placer un bâtiment en zone urbaine dense à Nantes correspond en termes de  $DJU$  à le placer dans une autre région climatique française.*

L'analyse des besoins relatifs d'énergie est plus intéressante en hiver qu'en été. En hiver, certains quartiers nantais sont associés à des villes méditerranéennes comme Nîmes. Ces quartiers sont situés le long de la Loire ou dans l'hyper-centre de Nantes. Le fait que ces zones soient localisées le long de la Loire rejoint les critiques émises à l'égard de l'EM décrit au chapitre 4. Nous avons vu que la façon de prendre en compte les surfaces d'eau dans les modèles spatiaux était sans doute inadaptée au problème. De plus, trop peu de stations proches de l'eau ont été utilisées pour permettre l'extrapolation des modèles dans des zones proches de



(a)



(b)

FIGURE 5.8 – Cartographie des besoins relatifs (a) de chauffage et (b) de climatisation pour la ville de Nantes

l'eau. Ces résultats sont donc peu pertinents, contrairement à ceux obtenus pour les autres *zones microclimatiques*.

- La zone de centre-ville, la plus chaude, est constituée de bâtiments hauts et de très peu de surfaces végétales.
- En périphérie s'inscrivent ensuite des zones assimilables aux valeurs de *DJU* de Carcassonne : ces zones sont souvent densément bâties mais la hauteur des bâtiments est nettement plus basse qu'en centre-ville. La végétation est également souvent plus dense.
- Des trouées apparaissent ensuite dans cette zone périphérique, propices à la génération de zones plus fraîches. Celles-ci sont toujours constituées d'espaces verts de taille importante ou de zones pavillonnaires caractérisées par des tailles de bâtiments faibles et par une végétation dense.
- Enfin, quelques tâches correspondent à la valeur de *DJU* de la station de référence de la région nantaise. Elles sont identifiables dans des zones majoritairement constituées d'espaces vert.

La figure 5.8a permet également d'évaluer la réduction des consommations énergétiques théoriquement engendrées par l'îlot de chaleur urbain. Ainsi, tout l'hyper-centre a une valeur de *DIU* près de 20% plus faible que celle calculée pour la station Météo-France normalement utilisée. Les consommations de chaleur de cette zone pourront ainsi être 20% plus faibles que celles qui auront pu être estimées à partir d'un outil de calcul utilisant comme fichier météorologique les données de Météo-France (Nort-sur-Erdre).

Les besoins estivaux de climatisation pour la région de Nantes sont naturellement faibles (moins de 30 °C.jr/an calculés pour à la station Météo-France de Nort-sur-Erdre), même si la présence de la ville augmente la température de l'air et de ce fait les besoins de fraîcheur (plus de 60°C.jr/an pour certaines zones de l'agglomération). Les bords de Loire et quelques îlots situés le long de l'Erdre sont clairement identifiés comme les zones les plus chaudes, ce qui peut sembler étrange. Comme pour les besoins de chaleur, les estimations de température pour les quartiers proches de grandes étendues d'eau sont peu fiables, et donc les besoins de climatisation également. La deuxième zone la plus chaude est constituée de l'hyper-centre et de la zone périphérique identifiée auparavant pour les besoins de chaleur. En termes de besoins de fraîcheur, elle peut être assimilée à la région parisienne (site Météo-France de Roissy). Les espaces verts et zones pavillonnaires, qui étaient distingués pour les besoins de chaleur, sont ici

rassemblés dans une même zone dont les besoins de fraîcheur sont identiques à ceux du site de référence (Météo-France Nort-sur-Erdre).

En termes de consommations de chaleur et de froid, l'exemple nantais ne met pas forcément en avant les effets négatifs de l'ICU : les consommations de chauffage sont plus faibles en ville qu'en périphérie tandis que les besoins supplémentaires de froid en ville sont négligeables devant les besoins de chaleur. Néanmoins, la température moyenne ainsi que la fréquence des événements caniculaires devraient augmenter ces prochaines décennies. Une étude pilotée par le Secrétariat général aux affaires régionales (SGAR) a analysé l'évolution pour le siècle en cours de plusieurs paramètres climatiques en retenant trois des quatre scénarios du GIEC. Le résultat des simulations pour la région nantaise sont donnés sous forme d'indicateurs calculés à trois horizons : 2030, 2050 et 2080. À noter que les indicateurs sont soit moyennés, soit sommés sur une période de 30 ans centrée sur chaque horizon. Les résultats suivants sont tirés du rapport remis au SGAR (Poupard 2013).

- La température moyenne est donnée sous la forme d'un écart à la période de référence, qui correspond à l'intervalle 1971-2000. Les écarts de température sont moyennés pendant les 30 années. Dans le cas du scénario le plus optimiste, cet écart sera de 1,7°C, 1,7°C et 2,7°C respectivement pour les horizons 2030, 2050 et 2080. Dans le cas du scénario le plus pessimiste, il sera de 1,7°C, 2,7°C et 5,2°C pour les horizons 2030, 2050 et 2080.
- Le nombre de jours de canicule est sommé sur la période de 30 ans. On compte un jour de canicule dès lors que pendant trois jours consécutifs, les températures minimales et maximales journalières dépassent respectivement 18,5°C et 33,5°C. Dans le cas du scénario le plus optimiste, ce nombre de jours sera de 50, 90 et 170 respectivement pour les horizons 2030, 2050 et 2080. Dans le cas du scénario le plus pessimiste, il sera de 50, 120 et 550 respectivement pour les horizons 2030, 2050 et 2080. Pour bien prendre conscience de ce dernier résultat, enregistrer 550 jours de canicules en 30 ans est équivalent à avoir en moyenne pendant l'été un épisode caniculaire tous les cinq jours (sachant qu'un épisode caniculaire dure trois jours).

La commune de Nantes pourrait par exemple se fixer l'objectif de ne pas augmenter les consommations de climatisation d'ici 2030. Un diagnostic et des préconisations peuvent être émis à partir de la méthodologie développée dans le cadre de cette thèse. La carte de la figure 5.8b pourrait ainsi être croisée avec d'autres données géographiques dont dispose la collectivité. Les bâtiments publics pourraient par exemple être positionnés sur le fond de carte afin d'identi-

fier les bâtiments à rénover thermiquement en priorité. Des données de population de l'INSEE pourraient également être utilisées : les quartiers les plus densément peuplés situés au sein de zones où les besoins de climatisation sont les plus forts peuvent être mis en évidence. D'après les équations 4.13 et 4.9 qui permettent de calculer l'ICU pour un quartier de l'agglomération, le *NDVI* est l'une des variables géographiques les plus influentes sur la valeur de l'ICU. Les zones où les besoins de climatisation sont les plus fortes pourraient donc faire l'objet d'un projet de végétalisation. Ceci montre l'intérêt que peut avoir ce type de résultats sur les préconisations qu'une collectivité peut mettre en oeuvre dans l'aménagement urbain.

### 5.2.2 Îlot de chaleur urbain et projets d'alerte aux allergènes d'origine végétale

Des applications pour téléphone mobile existent aujourd'hui afin d'être prévenu lorsqu'une espèce allergène démarre ou termine une période d'émission de pollens<sup>1</sup>. Ces informations permettent aux personnes allergiques de démarrer le traitement avant les premiers symptômes et de le stopper dès la fin de la période d'émission. Ces applications reposent sur l'utilisation de pollinariums sentinelles, qui sont des espaces d'observation dans lesquels sont réunies les principales espèces de plantes dont le pollen est allergisant. Une observation quotidienne du pollinarium permet de déterminer le début et la fin de la période d'émission de chaque espèce. Les pollinariums sentinelles sont souvent placés dans le jardin des plantes des villes, ce qui peut être critiquable puisque la date à laquelle chaque phase phénologique (dont l'émission de pollens) démarre est très fortement corrélée à la température de l'air environnant (qui varie à l'échelle de l'agglomération). En agronomie, il est courant d'associer à chaque espèce et à chaque stade phénologique une somme de degrés-jours. La méthode de calcul de cette somme est relativement proche de celle utilisée pour calculer les *DJU*. Les températures  $T_{max}(j)$  et  $T_{min}(j)$ , respectivement les températures maximales et minimales de la journée  $j$ , sont également utilisées. La température seuil (précédemment notée  $T_{seuil}$ ) est appelée en agronomie température de base et est notée  $T_{base}$ . Elle est propre à chaque espèce et représente la température minimale nécessaire pour assurer la croissance de la plante. Plus la température de l'air est forte au-dessus de cette température, plus la croissance de la plante est importante et donc plus les phases phénologiques vont apparaître tôt dans l'année. Pour une journée  $j$ , on calcule les degrés-jours de croissance  $DJ(j)$  par la procédure suivante :

<sup>1</sup> disponible sur la page [http : //www.airpl.org/Pollens/pollinariums – sentinelles](http://www.airpl.org/Pollens/pollinariums-sentinelles) [Accédé le 28/09/2016]

- si  $\frac{T_{max}(j)+T_{min}(j)}{2} \leq T_{base} : DJ(j) = 0 ;$
- si  $\frac{T_{max}(j)+T_{min}(j)}{2} > T_{base} : DJ(j) = \frac{T_{max}(j)+T_{min}(j)}{2} - T_{base}.$

Les degrés-jours de croissance sont ensuite sommés sur une période donnée afin de calculer la somme des températures. La difficulté consiste à identifier le début de cette période de décompte. Pour des espèces vivaces, il correspond à la fin de la période de dormance du spécimen. Pour les espèces dont on sème les graines chaque année, le décompte peut démarrer dès le premier jour du semis.

Pour chaque espèce, l'étape qui correspond à la dispersion des pollens par la plante est associée à une somme de température donnée. Le fait qu'une plante commence à émettre des pollens signifie donc qu'elle a atteint cette valeur seuil. Les parcs sont souvent qualifiés d'îlots de fraîcheur car ils représentent des espaces tempérés au sein de zones urbaines chaudes. Les espèces allergènes, qui sont sous surveillance dans les parcs, peuvent donc émettre des pollens plus précocément si elles se trouvent dans des espaces densément bâtis. Les alertes peuvent dans ce cas arriver trop tardivement pour un grand nombre de zones de la ville. L'objet de ce travail est d'identifier l'espace vert géré par la ville de Nantes qui serait le mieux adapté climatiquement pour accueillir les espèces allergènes étudiées.

Pour notre exemple d'application, nous étudierons la phase de floraison du lierre. Cette espèce n'est pas spécifiquement une espèce allergène mais a été étudiée dans le cadre du projet URBIO. Elle se trouve facilement en ville dans n'importe quel quartier. Plusieurs spécimens de lierre ont été identifiés sur chacun des sites de mesure de température du réseau angevin. Le suivi de leur période d'inflorescence par les écologues du projet a permis d'étudier les liens existants entre la période de floraison et la somme des degré-jour de croissance. Pour une température de base de 16°C, la floraison a eu lieu en moyenne pour une somme de température correspondant à 385 °C.jr de croissance, avec un écart-type de l'ordre de 29 °C.jr.

Ce résultat est utilisé pour identifier les périodes de floraison théoriques du lierre selon l'espace vert dans lequel il se trouverait pour la ville de Nantes. Dix parcs gérés par le service des espaces verts de Nantes sont étudiés. Les degrés-jours de croissance de chacun de ces sites sont calculés jour après jour à partir de trois informations : des données Météo-France, de l'EM développé dans cette thèse pour calculer l'indicateur  $ITL^{(nuit)}$  et de la définition des degrés-jours de croissance donnée précédemment. Le début de la période de calcul est fixé arbitrairement au 1er janvier et la température de base du lierre est fixée à 16°C. L'évolution de la somme des degrés-jour au cours de l'année 2014 est présentée figure 5.9.

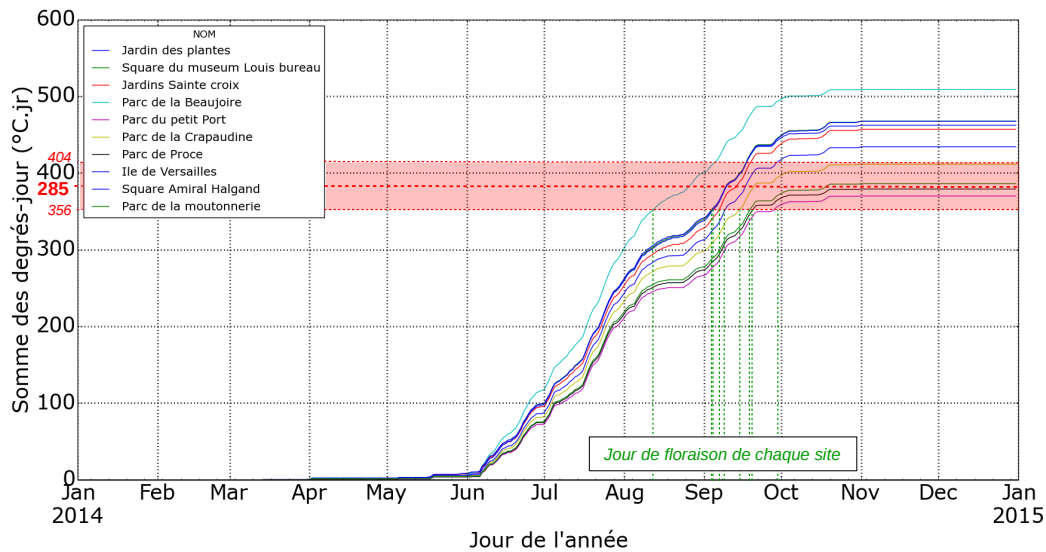


FIGURE 5.9 – Évolution temporelle de la somme des degrés-jour de croissance pour différents parcs de la ville de Nantes. Calcul réalisé à partir des modèles développés dans cette thèse et de la méthode de calcul des degrés jour de croissance. Le jour de floraison du lierre est établi pour chaque site dès lors que la somme de ses degrés-jour atteint l'intervalle de tolérance de la somme des degrés-jour du lierre déterminé par l'observation.

Certaines stations n'atteignent jamais la valeur moyenne de  $385^{\circ}\text{C.jr}$ . On considérera donc la valeur de  $356^{\circ}\text{C.jr}$  (la valeur moyenne moins l'écart-type) comme valeur seuil pour l'identification du jour de l'année que l'on assimile à l'émission des pollens pour un site donné. La figure 5.10 permet d'identifier chacun des parcs dans l'espace et de mettre en évidence les différences spatiales de début de floraison du lierre.

Le jour de l'année où les premières fleurs devraient théoriquement apparaître est rapporté pour chaque site au site phénologique de référence (Jardin des Plantes). Il est ainsi plus évident d'identifier les stations où les fleurs apparaîtront plus précocement ou plus tardivement que pour ce site. Le tableau 5.1 regroupe ces valeurs.

Seul le parc de la Beaujoire se distingue nettement des autres pour sa précocité phénologique. Ceci s'explique principalement du fait qu'il se trouve proche de l'eau. Or, nous avons vu au chapitre 4 que les estimations pour les sites caractérisés par une forte densité d'eau sont peu fiables. Les résultats pour ce parc ne sont donc pas considérés. Le lierre de tous les parcs de centre-ville fleurit ensuite presque au même moment que celui du Jardin des Plantes. Seuls les parcs très végétalisés ou plus périphériques (parc de Procé, parc du Petit Port, parc de la moutonnerie, parc de la Crapaudine), plus frais, ont un réel retard de floraison (souvent une dizaine de jours). Pour la ville de Nantes, le Jardin des Plantes semble donc être un parc



FIGURE 5.10 – Cartographie de la date de floraison de chaque parc étudié par rapport à celle du jardin des plantes.

particulièrement adapté au suivi d'émission des pollens. Cependant, une étude plus précise devra être menée à partir de la température de base de vraies espèces allergènes pour vérifier ce constat, et également d'un nombre de parcs plus important. De la même manière que pour les cartographies des besoins énergétiques, il serait intéressant de produire des cartes de précocité des stades phénologiques par rapport au site de référence. Ceci permettrait d'identifier les quartiers qui correspondent à des zones où les émissions d'allergènes ont lieu le plus tôt. Il faudra ensuite vérifier que des spécimens de l'espèce allergène en question sont bien présents dans le



Nom de l'espace vert	Nombre de jours
Ile de Versailles	-4
Jardin des plantes	0
Jardins Sainte croix	-2
Parc de Proce	11
Parc de la Beaujoire	-27
Parc de la Crapaudine	6
Parc de la moutonnerie	10
Parc du petit Port	20
Square Amiral Halgand	-4
Square du museum Louis bureau	-4

TABLE 5.1 – Nombre de jours séparant la floraison du lierre dans chacun des parcs à celle dans le Jardin des Plantes

secteur. Des actions pourraient enfin être mises en place, soit pour informer les populations de ce quartier, soit pour lutter contre l'ICU localement afin de repousser la date d'émission des premiers pollens.

## Conclusions et discussions

### 6.1 Synthèse des résultats obtenus

L'objectif de cette thèse était de développer une méthode permettant d'estimer l'intensité de l'îlot de chaleur urbain de n'importe quel quartier d'une agglomération et pour n'importe quel instant de l'année. Deux contraintes avaient été fixées :

- la méthode devait pouvoir être reproduite pour n'importe quelle ville du territoire français métropolitain, ce qui a motivé le choix des bases de données géographiques et météorologiques utilisées en entrée des modèles (celles-ci devaient être disponibles et homogènes sur l'ensemble du territoire français) ;
- les modèles proposés devaient relier directement les différences de température à des variables géographiques caractérisant le territoire ; c'est donc une approche empirique qui a été choisie plutôt qu'une approche formelle.

Le premier chapitre a été consacré à l'analyse de la littérature existante dans le domaine de la microclimatologie urbaine. Il a permis d'identifier les enjeux liés à la modélisation du microclimat urbain, et plus particulièrement aux méthodes de modélisation dites de *régression géoclimatique*. Nous avons vu qu'aucun des travaux utilisant ces modèles empiriques ne permet d'estimer la variation des températures dans l'espace et dans le temps ; soit les diffé-

rences spatiales de température sont estimées mais moyennées saisonnièrement, soit les variations temporelles de l'ICU sont estimées mais seulement pour un point donné du territoire. L'originalité de ce travail est donc de proposer une méthodologie permettant de prévoir spatialement (pour n'importe quel point  $i$  du territoire) et temporellement (à n'importe quel instant  $t$ ) les différences de température observables au sein d'une agglomération.

Six jeux de données météorologiques ont été utilisés pour calibrer et vérifier les modèles. Ils ont été présentés dans le deuxième chapitre. Quatre sont issus de mesures sur l'agglomération de Nantes, un sur l'agglomération d'Angers et le dernier sur l'agglomération de La Roche-sur-Yon. Le matériel et les méthodes d'acquisition ont été décrits pour chacun d'eux. À l'échelle de l'agglomération, une analyse des données météorologiques a permis de mettre en évidence que certaines directions de vent étaient plus propices à certains types de temps. L'importance de la précision des données a également été mise en évidence. En effet, la mesure de température de l'air est très sensible au rayonnement solaire : pour certains abris de mauvaise qualité, un capteur positionné au soleil peut mesurer une température supérieure de plus de 5°C à celle de l'air. Les deux types d'abris ventilés naturellement ont été positionnés à côté d'un abri ventilé mécaniquement afin de comparer leurs performances. La température a été mesurée sous chacun d'eux pendant plusieurs mois. Pendant le jour, la température mesurée sous l'abri TRH a été supérieure de plus de 3°C à celle mesurée sous l'abri ventilé mécaniquement tandis que les différences de température observées entre l'abri URBIO et l'abri ventilé mécaniquement étaient négligeables. Pendant la nuit, aucune différence significative n'a été observée entre les abris. Afin de corriger les données diurnes enregistrées sous les abris du réseau TRH, des modèles d'erreur ont été élaborés à partir d'une méthode de régression linéaire multiple. Les variables explicatives utilisées, construites à partir de la vitesse du vent, du rayonnement CLO descendant et de la température de l'air mesurés à la station Météo-France la plus proche, ont été déterminées à partir d'une mise en équation simplifiée des phénomènes physiques s'exerçant sur le système {capteur - abri}. Le modèle a été vérifié à partir d'une seconde campagne de comparaison des abris TRH et URBIO. Les résultats se sont avérés satisfaisants : les modèles de correction permettent de réduire la  $RMSE$  de plus de 50% (pour plusieurs journées ensoleillées estivales, l'erreur initiale  $RMSE_{ini}$  était de 0,87°C contre 0,41°C après correction). L'influence de la saleté sur les performances de l'abri a également été investiguée à cette occasion, en comparant la température observée sous des abris sales avec celle observée sous des abris propres. Au final, il s'avère que l'erreur engendrée par la saleté de l'abri est négligeable en comparaison de celle liée au type d'abri utilisé. Enfin, une méthode de correction automa-

tique des données de température mesurée sous les abris TRH a été développée. Pour cela, une reconstitution en trois dimensions de l'environnement bâti alentour à chaque station de mesure est réalisée à partir des données de la BD Topo, de la position spatiale des abris dans la scène urbaine et du logiciel SketchUp. Un *plug-in* connecté à SketchUp, développé par l'équipe de recherche du Centre de Recherche Nantais Architectures Urbanités (CRENAU), permet ensuite d'identifier les périodes pendant lesquelles l'abri est frappé par le soleil. C'est pendant celles-ci que les modèles d'erreur décrits auparavant sont utilisés pour corriger les données obtenues sous les abris TRH. Pour certains mois de l'année, certaines stations ont ainsi subi des modifications de leur température moyenne de l'ordre de 1°C.

La procédure de modélisation des variations spatiales et temporelles de l'ICU a été décrite dans le troisième chapitre. Le jeu de données utilisé pour calibrer les modèles est issu de mesures mises en place sur le territoire de l'agglomération nantaise et qui ont eu lieu pendant plusieurs années. Deux périodes de la journée ont été jugées pertinentes pour l'étude de l'ICU : une période nocturne définie entre le milieu et la fin de la nuit, et une période diurne définie entre le milieu de la journée et le milieu de l'après-midi solaire. La différence de température *ITL*, définie entre chacune des stations du jeu de données et une station Météo-France située à l'extérieur de la ville, a été calculée pour chaque journée du jeu de données. Les variations spatiales de l'ICU correspondent aux différences :

- de valeur moyenne de l'indicateur *ITL* (notée  $ITLS_{moy}$ ) observées entre les sites de mesure ;
- d'amplitude moyenne de l'indicateur *ITL* (notée  $ITLS_{std}$ ) observées entre les sites de mesure.

Les indicateurs  $ITLS_{moy}$  et  $ITLS_{std}$  ont ensuite été modélisés à partir de régressions linéaires multiples, dont les variables explicatives sont construites à partir d'indicateurs caractérisant le contexte géographique de chaque site. À chaque flux du bilan énergétique d'un volume urbain a été associé un ou plusieurs indicateurs géographiques. Ces indicateurs décrivent la morphologie du bâti ou la nature des sols présents dans un rayon de 500 m autour de chaque site de mesure. Ils sont calculés soit à partir de données issues de la BD Topo, soit à partir de traitements spécifiques d'images satellitaires prises par le satellite *RapidEye*. Les variations temporelles correspondent aux variations d'intensité de l'ICU entre deux journées consécutives (*ITLT*). Elles ont été modélisées à partir de régressions linéaires multiples dont les variables explicatives sont construites à partir de données météorologiques (vitesse du vent, nébulosité, etc.). Ces

données sont mesurées à la station Météo-France la plus proche. La période où l'influence de chaque variable météorologique sur l'ICU est la plus forte est déterminée statistiquement. Les deux modèles spatiaux et le modèle temporel sont rassemblés sous une même dénomination : *ensemble de modèles*. Cinq *variantes* sont proposées afin de déterminer plusieurs ensembles de modèles plutôt qu'un seul :

- trois d'entre elles correspondent à l'utilisation de méthodes particulières de conditionnement des données avant d'appliquer les régressions : tri saisonnier, tri selon la direction du vent, utilisation d'un nombre limité de stations pour calibrer les modèles ;
- les deux autres sont des modifications de la procédure de calibration des modèles : centrage et réduction robuste des valeurs  $ITL$ , utilisation de méthodes robustes pour les régressions linéaires.

Au final, l'utilisation de ces variantes a permis de proposer 32 ensembles de modèles.

Le chapitre quatre consistait à évaluer la performance de chaque variante et de chaque ensemble de modèles. Les variables météorologiques et géographiques utilisées par l'ensemble de modèles le plus performant ont été analysées plus en détail pour chaque saison. Les estimations qu'il a permis de produire ont également été confrontées aux observations. La performance de chaque ensemble de modèles a été évaluée en comparant les valeurs  $ITL$  observées à celles estimées pour chaque jeu de données. Le critère  $\widehat{RMSE}$ , calculé à partir de ces valeurs, a été utilisé pour classer les ensembles de modèles entre eux. Les six jeux de données ont été utilisés pour l'évaluation. Au final, la modélisation de l'indicateur de température diurne  $ITL^{(jour)}$  n'est pas convaincante. Il est préférable de considérer la température comme homogène pendant la période diurne  $P_{ICU}^{(jour)}$  plutôt que d'utiliser un des modèles proposés. La modélisation de l'indicateur nocturne donne des résultats plus satisfaisants. Si l'on considère tous les modèles, seules les variantes de tri directionnel et de tri saisonnier utilisés pour modéliser  $ITL^{(nuit)}$  ont une influence notable sur la performance des modèles proposés :

- les modèles établis à partir de données triées selon la direction du vent semblent plus performants que ceux établis sans trier les données ;
- d'une manière générale, les modèles calibrés à partir de données météorologiques triées par saison permettent très rarement d'obtenir des meilleures estimations que ceux calibrés à partir de toutes les données.

Cependant, pour chaque saison sauf pour l'automne, l'ensemble de modèles qui permet d'obtenir les meilleures estimations de l'indicateur  $ITL^{(nuit)}$  quel que soit le jeu de données est issu d'un tri saisonnier. Pour cet EM, les variables météorologiques qui influencent le plus les variations temporelles de l'ICU nocturne sont la nébulosité, le rayonnement et la vitesse du vent. La température et l'humidité relative de l'air sont également deux variables influentes généralement de moindre importance. L'ICU de chaque saison n'est pas expliqué par les mêmes variables météorologiques :

- la vitesse du vent et la température de l'air, lorsqu'elles sont faibles, ont pour effet de maximiser l'ICU en été ;
- en hiver, c'est lorsque la nébulosité est faible que l'ICU est maximisé.

En général, les valeurs prises par l'indicateur  $ITL$  à une période  $P_p$  sont attribuables à la valeur prise par une variable météorologique à cette même période  $P_p$ . Ce n'est pas le cas pour le rayonnement solaire : sa valeur nocturne étant nulle, c'est sa valeur moyenne diurne qui explique le mieux la valeur de l'indicateur  $ITL$  nocturne. Les variations spatiales sont principalement expliquées par le  $NDVI$  moyen mesuré autour de chaque site, un indicateur de densité de végétation. Selon la saison, d'autres indicateurs viennent compléter le modèle : en hiver, ce sont par exemple les indicateurs  $D_{fac}$  et  $D_{surf}$ , représentatifs du flux de chaleur émis par les bâtiments, qui prédominent.

Les valeurs de  $ITL^{(nuit)}$  estimées à l'aide du meilleur ensemble de modèles sont comparées aux valeurs observées. Les stations de mesure dont les estimations sont de mauvaise qualité ont des particularités géographiques.

- Elles sont souvent situées à l'extérieur de la zone urbaine ou dans des espaces fortement végétalisés (valeur de  $NDVI$  élevée). Ces défauts sont associés au manque de stations rurales dans le jeu de données utilisé pour la calibration des modèles.
- D'autres correspondent à des sites fortement minéralisés et dépourvus de tout bâtiment. Ce défaut pourrait également être corrigé en utilisant ce type de site dans le jeu de données de calibration.
- La surface de calcul des indicateurs géographiques paraît pour certaines stations clairement trop large. C'est par exemple le cas pour les stations de référence des jeux de données angevins et yonnais, pour lesquelles un rayon de 200 m (au lieu de 500 m) permettrait d'améliorer nettement la performance des modèles spatiaux.

Le chapitre cinq a permis d'illustrer les applications potentielles des résultats de la thèse. Il est possible d'estimer la température de n'importe quel point  $i$  de l'agglomération pour n'importe quel instant  $t$  :

- à partir de deux types de données :
  - les données météorologiques correspondant aux données de vitesse et de direction du vent, le rayonnement solaire descendant, la nébulosité, la température et l'humidité relative de l'air. Elles doivent être mesurées de préférence au pas horaire à la station Météo-France la plus proche. Les journées de mesure utilisées correspondent à celle qui contient l'instant  $t$  et la journée précédente.
  - les données géographiques correspondant à la BD Topographique de l'IGN et à une image satellitaire (de préférence *RapidEye*). Ces informations géographiques doivent décrire le territoire de l'agglomération dans lequel est situé le point  $i$  à modéliser.
- à partir de trois étapes décrites dans la thèse :
  - calcul des indicateurs géographiques  $IG_g$ , des indicateurs de température locale  $ITL$  et des indicateurs de météorologie régionale  $IMR_m$  ;
  - application des huit modèles spatiaux (deux par saison) et des quatre modèles temporels (un par saison) déterminés au chapitre 4 ;
  - mise en oeuvre de l'algorithme de calcul permettant de reconstruire le signal de température d'air.

Finalement, le jeu de données créé à partir de ces estimations peut être utilisé dans le but de répondre à plusieurs objectifs.

- Lors de la définition d'une nouvelle zone d'aménagement, des études de confort thermique extérieur ou de consommations des bâtiments sont parfois menées. Les outils de simulation thermique et aéraulique qui permettent de les réaliser nécessitent de connaître la température du quartier. Or c'est le plus souvent la température mesurée à MF qui est utilisée, ce qui ne permet pas de prendre en considération l'effet d'îlot de chaleur urbain. La méthodologie proposée permettrait donc d'améliorer la qualité des estimations. Nous avons en effet montré que l'erreur d'estimation de la température d'air d'un quartier était plus élevée lorsque la température était assimilée à celle mesurée à la station

Météo-France (RMSE = 1,14 °C) que lorsqu'elle était estimée en utilisant les résultats de la thèse (RMSE = 0.69 °C) ou mesurée directement sur site avec un équipement de faible qualité (RMSE = 0.67 °C).

- L'ICU est responsable de l'augmentation des besoins de froid des bâtiments en été et de la diminution des besoins de chaleur en hiver. Identifier les zones de l'agglomération où les modifications microclimatiques sont les plus fortes pourrait donc permettre de mieux gérer les consommations d'énergie du territoire. Un diagnostic est ainsi réalisé sur le territoire de la ville de Nantes, permettant d'identifier des zones où les consommations d'énergie sont, du fait de l'ICU, réduites de près de 20 % en hiver et presque doublées en été.
- Enfin, l'ICU a une influence sur la végétation : certaines phases phénologiques des végétaux (par exemple l'apparition de pollens) apparaissent plus précocement en ville qu'en périphérie. Il est possible d'être averti par téléphone lors du début de la période d'émission dans l'air des pollens de certaines espèces allergènes. Les services des espaces verts des villes sont à l'origine de cette information : une observation quotidienne d'échantillons témoins d'espèces allergènes placés dans le Jardin des Plantes de certaines villes françaises leur permet de déclencher l'alerte lorsque les premiers pollens apparaissent. Les résultats de cette thèse ont permis de calculer pour plusieurs parcs nantais la somme de degrés-jour du lierre, permettant ainsi d'estimer pour chacun d'eux la date d'apparition des premiers pollens. Au final, aucun autre espace vert ne devrait avoir une apparition nettement plus précoce des premiers pollens. Le Jardin des Plantes semble donc être adapté pour ce type d'observation. D'autres quartiers, fortement minéralisés et densément bâtis, devraient être plus chauds et seraient peut-être plus adaptés pour la surveillance des périodes d'émission d'allergènes. Les résultats de cette thèse pourraient permettre d'identifier les emplacements les plus appropriés pour placer le pollinarium sentinelle.

## 6.2 Limites techniques du travail effectué et futures travaux

Certaines des données ou des méthodes employées dans cette thèse sont criticables. Les critiques qui nous ont semblé les plus pertinentes sont discutées dans cette section.



## 6.2.1 Modification des données

### Données météorologiques

Les données des stations TRH étaient erronées et ont donc été corrigées. Cependant, le modèle d'erreur proposé au chapitre 2 a été vérifié seulement pour un site de mesure situé proche d'un mur et a donné à cette occasion des performances médiocres. Il serait pertinent que la vitesse du vent, utilisée dans le modèle d'erreur, soit pondérée par une fonction de la distance séparant le capteur de l'obstacle le plus proche. Cela permettrait d'intégrer au modèle de correction des informations sur le contexte microclimatique propre à chaque site.

Avant d'utiliser les données de température (corrigées ou non) pour calibrer ou vérifier les modèles permettant d'estimer l'ICU, il pourrait être intéressant de vérifier que la température mesurée par une station de mesure donnée est bien représentative de celle de ses alentours proches. Pour cela, des campagnes de mesure mobile pourraient être mises en oeuvre à des périodes stratégiques de la journée et de l'année.

### Données géographiques

Les informations relatives au type de revêtement des surfaces imperméables et à la valeur des flux anthropiques émis sur le territoire ne sont pas utilisées comme données d'entrée dans les modèles proposés. Ces informations sont difficilement disponibles à l'échelle d'une agglomération mais il pourrait être intéressant d'évaluer l'influence de telles variables sur les valeurs d'ICU, relativement aux autres variables utilisées dans les modèles. Pour cela, des mesures de température d'air pourraient être réalisées pendant une période donnée pour une dizaine de sites, en veillant à ce qu'une moitié ait des surfaces urbaines relativement claires et l'autre des surfaces relativement foncées. En parallèle, la température de chacun des sites serait estimée à l'aide de nos modèles. L'écart moyen au modèle d'un type de site (clair ou foncé) permettrait d'évaluer si cette variable est vraiment une variable importante à prendre en compte relativement aux variables les plus significatives de notre modèle (telle que le *NDVI*). La même méthodologie pourrait être appliquée pour étudier l'influence des émissions anthropiques sur l'ICU : cinq sites seraient caractérisés par des émissions anthropiques fortes (trafic, rejets de climatisation, etc.), les cinq autres par des émissions anthropiques faibles (quartiers à basse consommation énergétique). Il serait également intéressant de mesurer la température d'air dans des quartiers dont les performances énergétiques sont du même ordre de grandeur mais dont la méthode d'isolation diffère. La nuit, un quartier dont les bâtiments sont isolés par l'extérieur sera potentiellement plus chaud qu'un quartier dont les bâtiments sont isolés par l'intérieur,

tandis que l'inverse aura tendance à se produire en journée.

## 6.2.2 Modification des méthodes et des modèles

### Méthode et modèle généraux

Les données de mesure utilisées pour la calibration des modèles ne sont pas très fiables. Elles ont été corrigées à partir d'un modèle d'erreur qui a été vérifié seulement pour un seul site de mesure. De plus, tous les sites du jeu de données de calibration ne respectent pas nécessairement la condition d'isotropie et toutes les stations ne sont pas nécessairement très représentatives du quartier dans lequel ils sont positionnés. Enfin, les sites ne sont pas représentatifs de la diversité géographique et climatique des quartiers de l'agglomération nantaise : peu sont très végétalisés, aucun n'est réellement proche de l'eau, aucun n'est situé dans une zone commerciale ou dans une zone fortement minéralisée et dépourvue de bâtiments. L'ensemble de ces éléments font que les modèles établis ne sont pas adaptés pour estimer la température de n'importe quelle zone d'une agglomération. Il serait donc pertinent de recalibrer les modèles à partir de la même méthode mais en utilisant des jeux de données plus fiables. Les jeux de données URBIO pourraient répondre à l'ensemble des contraintes concernant la qualité des mesures. Cependant, ils ne sont pas encore suffisamment conséquents temporellement pour permettre l'élaboration de modèles fiables. Il est donc nécessaire d'attendre d'avoir de plus gros volumes de données ou d'utiliser un jeu de données produit par d'autres acteurs que l'IRTSV. Avec ce nouveau jeu de données, l'indicateur diurne pourrait donner des résultats plus satisfaisants. La température diurne pourrait ainsi ne plus être considérée comme homogène comme c'était le cas au chapitre 5 lors de la reconstitution du signal.

Que ce soit pour la calibration ou pour la vérification, la station utilisée comme référence de température pour le calcul de l'indicateur *ITL* devrait avoir une température moyenne largement plus élevée ou plus basse que celle de toutes les autres stations, ce qui n'est pas le cas de tous les jeux de données utilisés. En effet, plus les stations ont une température moyenne proche de celle de la station de référence, plus la variance de l'indicateur *ITL* à expliquer est faible et donc plus les modèles sont imprécis. De plus, la décision de considérer en moyenne une station comme plus fraîche ou plus chaude que la station de référence a des répercussions notables sur les résultats : si elle est considérée comme légèrement plus chaude mais qu'elle est en réalité légèrement plus froide (erreur d'estimation du paramètre  $\alpha_i$  de l'équation 3.8), la modélisation de son comportement temporel sera exactement contraire à son comportement réel (sa température sera estimée plus chaude que la station de référence lorsque le contraire

sera attendu et vice-versa).

### Méthodes et modèles temporels

Pour chaque jour, il serait intéressant de calculer le nombre de journées écoulées depuis les dernières précipitations et d'intégrer cet indicateur dans les modèles temporels comme facteur de pondération des variables explicatives. En effet, un sol perméable n'a pas du tout les mêmes caractéristiques selon s'il est sec ou humide. En journée, un sol perméable sec sera associé à un flux de chaleur latente faible et à une faible effusivité thermique. Dans ce cas, sa température peut augmenter très vite sous l'action du soleil et refroidir très rapidement la nuit. Au contraire, un sol perméable humide aura une température plus faible en journée et une effusivité thermique telle qu'il refroidira plus doucement la nuit (Spronken-Smith et Oke 1999). La vitesse de refroidissement de l'air d'un site de référence, souvent placé dans un contexte rural et constitué de surfaces naturelles perméables, sera donc plus ou moins élevée selon le niveau de sécheresse du sol. Ainsi, entre deux journées ayant les mêmes conditions météorologiques mais à concentration d'eau différente dans les sols, l'ICU nocturne d'un quartier contenant peu de sols perméables pourra fortement varier.

Pour modéliser l'indicateur *ITLT*, des relations non-linéaires sont utilisées, pouvant rendre complexe l'extrapolation des modèles à des villes situées dans d'autres zones climatiques. En effet, il est possible que dans ces conditions, certaines variables météorologiques sortent de l'intervalle de valeur utilisé pour la calibration (on pensera notamment à la vitesse du vent qui pourra être plus élevée ou à la température de l'air qui pourra être plus basse ou plus élevée). L'utilisation des fonctions quadratiques ou exponentielles qui n'a dans ce cas pas de justification physique, peut donner lieu à des estimations erronées car ces fonctions ne sont pas extrapolables. Afin de gagner en robustesse (s'ils devaient être appliqués à des villes situées dans d'autres zones climatiques), les modèles non-linéaires actuellement utilisés pour modéliser temporellement l'ICU pourraient être remplacés par des modèles linéaires.

Les indicateurs *ITLS* sont des indicateurs moyennés qui caractérisent le microclimat d'un site donné. Or, ces moyennes sont réalisées à partir d'un volume de données peu important : 4 années de données sont utilisées au lieu des 30 années habituellement utilisées en climatologie. Cependant, dans le cadre de l'étude de l'ICU, même si de tels volumes de données microclimatiques existaient, il ne serait pas recommandé d'utiliser les 30 années pour calibrer les modèles car la nature des sols d'un site urbain et la morphologie de ses bâtiments peuvent évoluer drastiquement pendant un tel intervalle de temps. Il serait malgré tout oppor-

tun de vérifier pour chaque variable météorologique que les quatre années utilisées sont bien représentatives climatiquement des trentes dernières années écoulées.

### Méthodes et modèles spatiaux

Plusieurs problèmes découlent de l'utilisation du *NDVI* comme indicateur explicatif des variations de l'ICU. Pour commencer, deux hypothèses majeures ont été retenues pour justifier le lien entre *NDVI* et bilan d'énergie d'une surface.

- nous avons considéré une relation linéaire entre *NDVI* et évapotranspiration, ce qui n'est pas forcément physiquement défendable. Soumis à une température donnée et pour une même activité chlorophyllienne, les végétaux n'évapotranspirent pas nécessairement la même quantité d'eau. De plus, à propriétés thermiques presque équivalentes, deux types de surfaces non végétalisées peuvent avoir des valeurs de réflectance différentes dans le rouge et le proche infra-rouge. Leur valeur de *NDVI* sera donc différente alors qu'aucun des deux ne transpire. Il serait donc intéressant de fixer à une valeur donnée le *NDVI* d'une surface dès lors qu'elle est inférieure à une valeur seuil (par exemple en-dessous de zéro, la valeur fixée serait 0).
- les images qui permettent de calculer le *NDVI* ont été prises lors d'une journée printanière. Or, le comportement de la végétation n'est pas constant tout au long de l'année : toutes les espèces n'entrent ou ne sortent pas de leur dormance à la même période. De plus, toutes ne sont pas sensibles de la même manière à des périodes de sécheresse prolongée, à des vagues de chaleur intenses ou à des refroidissements brusques (celles qui sont les plus sensibles peuvent végéter voire mourir). Utiliser des images satellitaires pour chaque saison ou chaque mois permettrait donc d'avoir une information plus précise des variations temporelles de l'évapotranspiration de la végétation.

De plus, l'indicateur *NDVI* est difficilement appréhendable. Il n'est pas associé directement à une densité de végétation haute ou basse. Il serait donc intéressant de traduire cet indicateur en une information plus concrète. Enfin, le *NDVI* est un indicateur qui répond seulement à une quantité de chlorophylle contenue par les surfaces. D'autres indicateurs de végétation pondèrent cette quantité par d'autres facteurs telles que la réflectance du sol (*modified soil-adjusted vegetation index - MSAVI*) ou la morphologie du végétal (*enhanced vegetation index - EVI*). Ces indicateurs pourraient donc être utilisés en substitution du *NDVI* afin de déterminer s'ils permettent d'estimer plus précisément les valeurs d'ICU.

Quelle que soit la période d'acquisition des données météorologiques (entre 2010 et 2016), les données géographiques (BD Topographique et images satellitaires) utilisées datent de 2011. Il serait intéressant de vérifier si l'utilisation de données géographiques mieux adaptées à la période d'acquisition de chaque jeu de données modifie drastiquement les résultats.

La méthode d'identification des zones les plus ouvertes, utilisée pour calculer la distance d'un point à la périphérie de la ville, peut être critiquée.

- Elle ne prend pas en compte la hauteur des bâtiments qui peuvent modifier grandement le facteur de vue du ciel.
- Elle ne considère pas le type de sol des zones ouvertes vers le ciel : indifféremment, les zones ouvertes sont assimilées à des zones de fraîcheur. Or en ville, les grands espaces ouverts peuvent correspondre à des parkings ou à des cimetières fortement minéralisés, qui en été, sont associés à des sources de chaleur plutôt qu'à des sources de fraîcheur ;
- Dans le cas où le vent souffle fort, sa direction peut avoir une grande influence sur la direction de propagation de la fraîcheur produite par un espace ouvert (Zhang 2016). La distance à la périphérie pourrait donc être calculée frontalement à la direction du vent. Cela implique que l'indicateur de distance à la périphérie varie selon la direction du vent.

La taille de l'unité spatiale de référence choisie pour le calcul des indicateurs géographiques a été fixée à partir des résultats de la littérature (cercle de 500 m de rayon). Une étude pourrait être menée pour identifier la taille de l'unité spatiale donnant les meilleures estimations des indicateurs *ITL*. Pour cela, la méthode qui a permis d'identifier la taille des zones d'influence des stations de référence *MF\_angers* et *MF\_LRY*, illustrée en section 4.5.2, pourrait être employée.

### **6.3 Mise en oeuvre de la méthodologie pour d'autres zones urbaines**

La méthodologie proposée dans cette thèse a été pensée afin d'être appliquée à d'autres villes de France métropolitaine. Utiliser directement les modèles spatiaux et temporels issus de ce travail n'est pas conseillé pour des territoires ou des zones du territoire contenant d'importantes surfaces d'eau. Les villes situées près des montagnes ou de la mer sont également des territoires où les résultats ne seront probablement pas les meilleurs. Il est fortement recommandé de mettre

en place des campagnes de mesure de température afin de vérifier les estimations produites par les modèles :

- dans des zones clés de l'agglomération ;
- pour des conditions météorologiques particulières.

Dans l'idéal, il faudrait appliquer l'ensemble de la méthodologie à ces territoires afin de calibrer de nouveaux modèles. Dans ce cas, les recommandations suivantes doivent être mises en oeuvre :

- des jeux de données de température utilisant au moins une dizaine de sites de mesure et plusieurs années de données doivent être utilisés ;
- les équipements de mesure devront être le moins possible sensibles au rayonnement solaire ;
- il faudra veiller à ce que les sites de mesure soient bien représentatifs de l'hétérogénéité géographique et climatique de l'agglomération ;
- des indicateurs spécifiques à ces territoires, comme l'élévation d'un site par rapport au niveau de la mer ou par rapport à ses alentours proches, pourront être utilisés ;
- certaines propositions d'amélioration techniques proposées en section 6.2 pourront éventuellement être testées.

Certaines collectivités souhaiteront peut-être, pour estimer l'ICU de leur territoire, utiliser d'autres bases de données géographiques que celles proposées dans cette thèse (BD Topographique de l'IGN et images satellitaires *RapidEye*). Ce sera par exemple le cas de villes qui ne sont pas situées sur le territoire français. Dans ce cas et comme précédemment, les modèles pourront être utilisés. Cependant, il est conseillé d'au moins réaliser des campagnes de mesure de température pour vérifier la précision des estimations, et dans l'idéal de recalibrer les modèles à partir des nouvelles données géographiques et d'un jeu de données météorologiques.



# Bibliographie

- Alonso, M., M. Fidalgo et J. Labajo (2007). “The urban heat island in Salamanca (Spain) and its relationship to meteorological parameters”. In : *Climate Research* 34.1, p. 39 (cf. p. [52](#), [72](#), [107](#), [179](#)).
- Anderson, S. (2002). “An evaluation of spatial interpolation methods on air temperature in Phoenix, AZ”. In : *Department of Geography, Arizona State University Tempe*, p. 85287–0104 (cf. p. [56](#)).
- Armstrong, J. (1974). “Temperature differences between two ground-level sites and a roof site in Southampton”. In : *Meteorological Magazine* 103, p. 360–368 (cf. p. [63](#)).
- Balázs, B., J. Unger, T. Gál, Z. Sümeghy, J. Geiger et S. Szegedi (2009). “Simulation of the mean urban heat island using 2D surface parameters : empirical modelling, verification and extension”. In : *Meteorological Applications* 16.3, p. 275–287 (cf. p. [57](#), [65](#), [68](#), [77](#), [79–81](#), [84](#), [89](#), [107](#), [114](#), [207](#)).
- Bernabé, A., M. Musy, H. Andrieu et I. Calmet (2015). “Radiative properties of the urban fabric derived from surface form analysis : A simplified solar balance model”. In : *Solar Energy* 122, p. 156–168 (cf. p. [202](#)).
- Best, M., C. Grimmond et M. G. Villani (2006). “Evaluation of the urban tile in MOSES using surface energy balance observations”. In : *Boundary-Layer Meteorology* 118.3, p. 503–525 (cf. p. [48](#)).
- Blazejczyk, K., G. Jendritzky, P. Bröde, D. Fiala, G. Havenith, Y. Epstein, A. Psikuta et B. Kampmann (2013). “An introduction to the universal thermal climate index (UTCI)”. In : (cf. p. [35](#)).
- Bocher, E. et G. Petit (2012). “OrbisGIS : Geographical Information System designed by and for research”. In : *Innovative Software Development in GIS*, p. 23–66 (cf. p. [209](#)).
- Böhm, R. (1998). “Urban bias in temperature time series—A case study for the city of Vienna, Austria”. In : *Climatic Change* 38.1, p. 113–128 (cf. p. [51](#)).



- Bouyer, J. (2009). “Modélisation et simulation des microclimats urbains—Étude de l’impact de l’aménagement urbain sur les consommations énergétiques des bâtiments”. Thèse de doct. Université de Nantes (cf. p. 42, 45).
- Cegibat, G. (2016). *Outil de calcul des DJU*. URL : <https://cegibat.grdf.fr/simulateur/calcul-dju> (visité le 31/08/2016) (cf. p. 272).
- Champiat, C. (2009). “Identifier les îlots de chaleur urbains pour réduire l’impact sanitaire des vagues de chaleur”. In : *Environnement, Risques & Santé* 8.5, p. 399–411 (cf. p. 35).
- Chen, L., E. Ng, X. An, C. Ren, M. Lee, U. Wang et Z. He (2012). “Sky view factor analysis of street canyons and its implications for daytime intra-urban air temperature differentials in high-rise, high-density urban areas of Hong Kong : a GIS-based simulation approach”. In : *International Journal of Climatology* 32.1, p. 121–136 (cf. p. 57, 65, 84, 107, 108, 114, 175).
- Cheng, X., D. Su, D. Li, L. Chen, W. Xu, M. Yang, Y. Li, Z. Yue et Z. Wang (2014). “An improved method for correction of air temperature measured using different radiation shields”. In : *Advances in Atmospheric Sciences* 31.6, p. 1460–1468 (cf. p. 125, 142, 146).
- Delhomme, J. (1978). “Kriging in the hydrosociences”. In : *Advances in water resources* 1.5, p. 251–266 (cf. p. 56).
- Desthieux, G. (2004). “Utilisation d’indicateurs spatiaux dans un processus participatif de diagnostic de quartier à Saint-Jean”. In : *EPLF, Genève* (cf. p. 74).
- Erell, E. et T. Williamson (2007). “Intra-urban differences in canopy layer air temperature at a mid-latitude city”. In : *International Journal of Climatology* 27.9, p. 1243–1256 (cf. p. 73, 107, 173, 179, 195).
- Erell, E., V. Leal et E. Maldonado (2005). “Measurement of air temperature in the presence of a large radiant flux : an assessment of passively ventilated thermometer screens”. In : *Boundary-layer meteorology* 114.1, p. 205–231 (cf. p. 125).
- Fenner, D., F. Meier, D. Scherer et A. Polze (2014). “Spatial and temporal air temperature variability in Berlin, Germany, during the years 2001–2010”. In : *Urban Climate* 10, p. 308–331 (cf. p. 67, 77, 105, 107, 173).
- Gal, T., F. Lindberg et J. Unger (2009). “Computing continuous sky view factors using 3D urban raster and vector databases : comparison and application to urban climate”. In : *Theoretical and applied climatology* 95.1-2, p. 111–123 (cf. p. 107).
- Garstang, M., P. Tyson et G. Emmitt (1975). “The structure of heat islands”. In : *Reviews of Geophysics* 13.1, p. 139–165 (cf. p. 51).

- Grimmond, C. et T. Oke (2002). “Turbulent heat fluxes in urban areas : observations and a local-scale urban meteorological parameterization scheme (LUMPS)”. In : *Journal of Applied Meteorology* 41.7, p. 792–810 (cf. p. 48).
- Grimmond, C. et al. (2009). “Urban surface energy balance models : model characteristics and methodology for a comparison study”. In : *Meteorological and Air Quality Models for Urban Areas*. Springer, p. 97–123 (cf. p. 43, 47).
- Grimmond, C. et al. (2010). “The international urban energy balance models comparison project : first results from phase 1”. In : *Journal of applied meteorology and climatology* 49.6, p. 1268–1292 (cf. p. 48, 49, 51).
- Gros, A. (2013). “Modélisation de la demande énergétique des bâtiments à l’échelle d’un quartier”. Thèse de doct. Université de La Rochelle (cf. p. 35, 44).
- Hart, M. et D. Sailor (2009). “Quantifying the influence of land-use and surface characteristics on spatial variability in the urban heat island”. In : *Theoretical and Applied Climatology* 95.3-4, p. 397–406 (cf. p. 65, 71, 73, 77, 80, 82, 107, 175, 179).
- Hathway, E. et S. Sharples (2012). “The interaction of rivers and urban form in mitigating the Urban Heat Island effect : A UK case study”. In : *Building and Environment* 58, p. 14–22 (cf. p. 80, 236).
- Hinkel, K. M., F. E. Nelson, A. E. Klene et J. H. Bell (2003). “The urban heat island in winter at Barrow, Alaska”. In : *International Journal of Climatology* 23.15, p. 1889–1905 (cf. p. 56).
- Hjort, J. et M. Marmion (2008). “Effects of sample size on the accuracy of geomorphological models”. In : *Geomorphology* 102.3, p. 341–350 (cf. p. 82).
- Hjort, J., J. Suomi et J. Käyhkö (2011). “Spatial prediction of urban–rural temperatures using statistical methods”. In : *Theoretical and applied climatology* 106.1-2, p. 139–152 (cf. p. 56, 57, 68, 69, 82–84, 107, 108).
- Höppe, P. (1999). “The physiological equivalent temperature—a universal index for the biometeorological assessment of the thermal environment”. In : *International journal of Biometeorology* 43.2, p. 71–75 (cf. p. 35).
- Howard, L. (1818). *The climate of London*. T. 1. W. Phillips, sold also by J. et A. Arch (cf. p. 35, 41, 42, 44).
- Lacombe, M., D. Bousri, M. Leroy et M. Mezred (2011). *WMO Field Intercomparison of Thermometer Screens/Shields and Humidity Measuring Instruments, Ghardaia, Algeria, November 2008–October 2009* (cf. p. 60, 93, 125).

- Barnett, A., D. Hatton et D. Jones (1998a). *Recent Changes in Thermometer Screen Design and their Impact* (cf. p. 60).
- ISO (2007). *Meteorology – Air temperature measurements – Test methods for comparing the performance of thermometer shields/screens and defining important characteristics*. Standard. Geneva, CH (cf. p. 137).
- Jauregui, E. (1991). “Influence of a large urban park on temperature and convective precipitation in a tropical city”. In : *Energy and buildings* 15.3, p. 457–463 (cf. p. 79, 208).
- Jochner, S. et A. Menzel (2015). “Urban phenological studies—Past, present, future”. In : *Environmental Pollution* 203, p. 250–261 (cf. p. 35).
- Joerin, F. et P. Rondier (2007). “Indicateurs et décision territoriale : Pourquoi ? quand ? comment”. In : *Sénécal G. et al.-Les indicateurs socio-territoriaux et les métropoles-Presses del’Université Laval, Québec, Canada*, p. 9–36 (cf. p. 74).
- Karl, T. R., H. F. Diaz et G. Kukla (1988). “Urbanization : Its detection and effect in the United States climate record”. In : *Journal of Climate* 1.11, p. 1099–1123 (cf. p. 50).
- Knorr, W., I. Prentice, J. House et E. Holland (2005). “Long-term sensitivity of soil carbon turnover to warming”. In : *Nature* 433.7023, p. 298–301 (cf. p. 42).
- Krüger, E. et B. Givoni (2007). “Outdoor measurements and temperature comparisons of seven monitoring stations : Preliminary studies in Curitiba, Brazil”. In : *Building and environment* 42.4, p. 1685–1698 (cf. p. 68, 72, 73, 77, 78, 107, 175).
- Lacombe, M. (2010). “Results of the WMO intercomparison of thermometer screens/shields and hygrometers in hot desert conditions”. In : *TECO-2010–WMO Technical Conference on Meteorological, Environmental Instruments et Methods of Observation, Helsinki, Finland, 30 August–1 September 2010*, [http://www.wmo.int/pages/prog/www/IMOP/publications/IOM-104 TECO-2010/3 4 Lacombe France. pdf](http://www.wmo.int/pages/prog/www/IMOP/publications/IOM-104%20TECO-2010/3%20Lacombe%20France.pdf) (cf. p. 93, 113, 125, 132).
- László, E. et S. Szegedi (2012). “Impacts of Some Surface Parameters on Urban Heat Island Development”. In : (cf. p. 79, 107).
- Lauwaet, D., K. De Ridder, S. Saeed, E. Brisson, F. Chatterjee, N. van Lipzig, B. Maiheu et H. Hooyberghs (2016). “Assessing the current and future urban heat island of Brussels”. In : *Urban Climate* 15, p. 1–15 (cf. p. 194).
- Le Bras, J. (2015). “Le microclimat urbain à haute résolution : mesures et modélisation”. Thèse de doct. Toulouse 3 (cf. p. 65, 73, 193, 195).
- Leconte, F. (2014). “Caractérisation des îlots de chaleur urbain par zonage climatique et mesures mobiles : cas de Nancy”. Thèse de doct. Université de Lorraine (cf. p. 41, 45, 53, 55).

- Lelovics, E., J. Unger et T. Gál (2014). “Design of an urban monitoring network based on local climate zone mapping and temperature pattern modelling”. In : *Climate Research* 60, p. 51–62 (cf. p. 55, 56).
- l'énergie, P. membres de l'agence internationale pour (2014). *Key world energy statistics* (cf. p. 70).
- Maby, J. (2003). “Approche conceptuelle et pratique des indicateurs en géographie”. In : *Objets et indicateurs géographiques, Avignon, A. Barthélémy*, p. 16–41 (cf. p. 74, 201).
- Malys, L. (2012). “Évaluation des impacts directs et indirects des façades et des toitures végétales sur le comportement thermique des bâtiments”. Thèse de doct. PhD thesis] Ecole Centrale de Nantes (France) (cf. p. 35, 44).
- Masson, V., C. Grimmond et T. Oke (2002). “Evaluation of the Town Energy Balance (TEB) scheme with direct measurements from dry districts in two cities”. In : *Journal of applied meteorology* 41.10, p. 1011–1026 (cf. p. 48).
- Matzarakis, A., F. Rutz et H. Mayer (2010). “Modelling radiation fluxes in simple and complex environments : basics of the RayMan model”. In : *International Journal of Biometeorology* 54.2, p. 131–139 (cf. p. 149, 167).
- McKendry, I. G. (2003). “Applied climatology”. In : *Progress in Physical Geography* 27.4, p. 597–606 (cf. p. 47).
- Meier, F., D. Fenner, T. Grassmann, B. Jänicke, M. Otto et D. Scherer (2015). “Challenges and benefits from crowd-sourced atmospheric data for urban climate research using Berlin, Germany, as testbed”. In : (cf. p. 37).
- Météo-France, D. d. l. C. (2005). *Fiche méthode Degrés Jours*. URL : %5C(<http://climatheque.meteo.fr/Docs/DJC-methode.pdf%5C>) (visité le 07/10/2016) (cf. p. 269).
- Miguet, F. (2000). “Paramètres physiques des ambiances architecturales : Un modèle numérique pour la simulation de la lumière naturelle dans le projet urbain”. Thèse de doct. Ecole Nationale Supérieure d'Architecture de Nantes (France) (cf. p. 128).
- Monteith, J. et M. Unsworth (2007). *Principles of environmental physics*. Academic Press (cf. p. 149).
- Nakamura, R. et L. Mahrt (2005). “Air temperature measurement errors in naturally ventilated radiation shields”. In : *Journal of Atmospheric and Oceanic Technology* 22.7, p. 1046–1058 (cf. p. 125, 142, 146).

- Nakamura, Y. et T. Oke (1988). “Wind, temperature and stability conditions in an east-west oriented urban canyon”. In : *Atmospheric Environment (1967)* 22.12, p. 2691–2700 (cf. p. 63, 92, 122, 123).
- Niachou, K., I. Livada et M. Santamouris (2008). “Experimental study of temperature and air-flow distribution inside an urban street canyon during hot summer weather conditions—Part I : Air and surface temperatures”. In : *Building and Environment* 43.8. Hybrid Ventilation In The Urban Environment, p. 1383–1392. ISSN : 0360-1323. DOI : \ (<http://dx.doi.org/10.1016/j.buildenv.2007.01.039> \). URL : %5C(<http://www.sciencedirect.com/science/article/pii/S0360132307000418>%5C) (cf. p. 122).
- Oke, T. (1973). “City size and the urban heat island”. In : *Atmospheric Environment (1967)* 7.8, p. 769–779 (cf. p. 50).
- (1976). “The distinction between canopy and boundary-layer urban heat islands”. In : *Atmosphere* 14.4, p. 268–277 (cf. p. 69, 73, 107, 173, 175, 179, 196).
- (1987). *Boundary layer climates. 2nd.* Methuen, 289p (cf. p. 35, 42–44, 46, 47, 51, 68, 70, 72, 149).
- (2004). *Initial guidance to obtain representative meteorological observations at urban sites.* T. 81. World Meteorological Organization Geneva (cf. p. 42, 45, 62, 63, 65, 79, 81, 97, 113, 122, 186).
- Oke, T. et G. B. Maxwell (1975). “Urban heat island dynamics in Montreal and Vancouver”. In : *Atmospheric Environment (1967)* 9.2, p. 191–200 (cf. p. 66).
- Peel, M. C., B. L. Finlayson et T. A. McMahon (2007). “Updated world map of the Köppen-Geiger climate classification”. In : *Hydrology and earth system sciences discussions* 4.2, p. 439–473 (cf. p. 89).
- Pelgrum, H. et W. Bastiaanssen (1996). “An Intercomparison of Techniques to Determine the Area-Averaged Latent Heat Flux from Individual in Situ Observations : A remote Sensing Approach Using the European Field Experiment in a Desertification-Threatened Area Data”. In : *Water Resources Research* 32.9, p. 2775–2786 (cf. p. 204).
- Perez, R., R. Seals et J. Michalsky (1993). “All-weather model for sky luminance distribution—preliminary configuration and validation”. In : *Solar energy* 50.3, p. 235–245 (cf. p. 128).

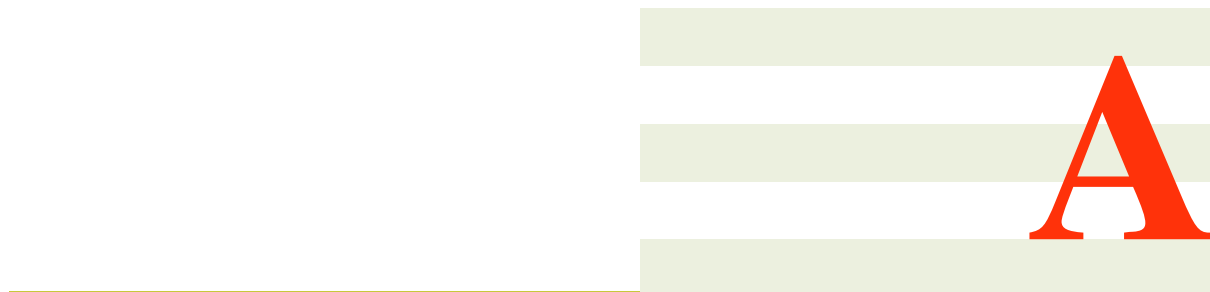
- Petralli, M., L. Massetti, G. Brandani et S. Orlandini (2014). “Urban planning indicators : useful tools to measure the effect of urbanization and vegetation on summer air temperatures”. In : *International Journal of Climatology* 34.4, p. 1236–1244 (cf. p. [52](#), [69](#), [77](#), [107](#), [175](#), [201](#)).
- Poupard, M. (2013). *Stratégie d'adaptation au changement climatique dans le Grand-Ouest* (cf. p. [275](#)).
- Zhang, X. (2016). “Étude des parcs comme dispositifs d'atténuation de l'îlot de chaleur urbain”. Mém.de mast. Institut de Recherche des Sciences et Techniques de la Ville (cf. p. [292](#)).
- Revel, D., H.-M. Füßel et A. Jol (2012). *Climate change, impacts and vulnerability in Europe 2012* (cf. p. [35](#)).
- Robitaille, É. et J. Laguë (2009). *Indicateurs géographiques de l'environnement bâti et de l'environnement des services influant sur l'activité physique, l'alimentation et le poids corporel*. Direction du développement des individus et des communautés, Institut national de santé publique du Québec (cf. p. [74](#)).
- Rouse Jr, J., R. Haas, J. Schell et D. Deering (1974). “Monitoring vegetation systems in the Great Plains with ERTS”. In : *NASA special publication* 351, p. 309 (cf. p. [203](#)).
- Runnalls, K. et T. Oke (2000). “Dynamics and controls of the near-surface heat island of Vancouver, British Columbia”. In : *Physical Geography* 21.4, p. 283–304 (cf. p. [67](#), [173](#), [175](#)).
- Sailor, D. (2011). “A review of methods for estimating anthropogenic heat and moisture emissions in the urban environment”. In : *International Journal of Climatology* 31.2, p. 189–199. ISSN : 08998418. DOI : [10.1002/joc.2106](https://doi.org/10.1002/joc.2106) (cf. p. [71](#)).
- Santamouris, M., N. Papanikolaou, I. Koronakis, I. Livada et D. Asimakopoulos (1999). “Thermal and air flow characteristics in a deep pedestrian canyon under hot weather conditions”. In : *Atmospheric Environment* 33.27, p. 4503–4521 (cf. p. [122](#)).
- Scherer, D., U. Fehrenbach, T. Lakes, S. Lauf, F. Meier et C. Schuster (2014). “Quantification of heat-stress related mortality hazard, vulnerability and risk in Berlin, Germany”. In : *DIE ERDE—Journal of the Geographical Society of Berlin* 144.3-4, p. 238–259 (cf. p. [35](#)).
- Sims, R., R. Hodas et T. Rinke (2009). *Cities, towns and renewable energy Yes in my front yard* (cf. p. [70](#)).
- Spronken-Smith, R. et T. Oke (1998). “The thermal regime of urban parks in two cities with different summer climates”. In : *International journal of remote sensing* 19.11, p. 2085–2104 (cf. p. [79](#), [208](#)).
- (1999). “Scale modelling of nocturnal cooling in urban parks”. In : *Boundary-Layer Meteorology* 93.2, p. 287–312 (cf. p. [80](#), [290](#)).

- Steenefeld, G., S. Koopmans, B. Heusinkveld et N. Theeuwes (2014). “Refreshing the role of open water surfaces on mitigating the maximum urban heat island effect”. In : *Landscape and Urban Planning* 121, p. 92–96 (cf. p. 68, 81, 107, 236).
- Stewart, I. D. (2011). “Redefining the urban heat island”. In : (cf. p. 41, 44, 53, 54, 81).
- Suomi, J., J. Hjort et J. Käyhkö (2012). “Effects of scale on modelling the urban heat island in Turku, SW Finland”. In : *Clim Res* 55, p. 105–118 (cf. p. 52, 69, 77, 79, 81, 85, 89, 107, 201).
- Svensson, M., I. Eliasson et B. Holmer (2002). “A GIS based empirical model to simulate air temperature variations in the Goteborg urban area during the night”. In : *Climate Research* 22.3, p. 215–226 (cf. p. 52, 68, 69, 72, 73, 83, 108, 173, 179, 195, 255).
- Szymanowski, M. et M. Kryza (2011). “Application of geographically weighted regression for modelling the spatial structure of urban heat island in the city of Wroclaw (SW Poland)”. In : *Procedia Environmental Sciences* 3, p. 87–92 (cf. p. 56).
- Szymanowski, M., M. Kryza et al. (2009). “GIS-based techniques for urban heat island spatialization”. In : *Climate research (Open Access for articles 4 years old and older)* 38.2, p. 171 (cf. p. 56).
- Taesler, R. (1980). *Studies of the development and thermal structure of the urban boundary layer in Uppsala*. Meteorologiska institutionen [Uppsala universitet] (cf. p. 63).
- Taniguchi, M., W. C. Burnett et G. D. Ness (2008). “Integrated research on subsurface environments in Asian urban areas”. In : *Science of the Total Environment* 404.2, p. 377–392 (cf. p. 42).
- Tannier, C., I. Thomas, G. Vuidel et P. Frankhauser (2011). “A Fractal Approach to Identifying Urban Boundaries.” In : *Geographical Analysis* 43.2, p. 211–227 (cf. p. 207, 212).
- Tucker, C. J. (1979). “Red and photographic infrared linear combinations for monitoring vegetation”. In : *Remote sensing of Environment* 8.2, p. 127–150 (cf. p. 203).
- Van der Meulen, J. et T. Brandsma (2008). “Thermometer screen intercomparison in De Bilt (The Netherlands), Part I : Understanding the weather-dependent temperature differences”. In : *International Journal of Climatology* 28.3, p. 371–387 (cf. p. 133).
- Voogt, J. A. et C. Grimmond (2000). “Modeling surface sensible heat flux using surface radiative temperatures in a simple urban area”. In : *Journal of Applied Meteorology* 39.10, p. 1679–1699 (cf. p. 48).

- Watson, R. T. et D. L. Albritton (2001). *Climate change 2001 : Synthesis report : Third assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press (cf. p. 35).
- Wilby, R. L. et G. L. Perry (2006). “Climate change, biodiversity and the urban environment : a critical review based on London, UK”. In : *Progress in Physical Geography* 30.1, p. 73–98 (cf. p. 35).
- Yan, H., S. Fan, C. Guo, F. Wu, N. Zhang et L. Dong (2014). “Assessing the effects of landscape design parameters on intra-urban air temperature variability : The case of Beijing, China”. In : *Building and environment* 76, p. 44–53 (cf. p. 65, 69, 77–79, 107, 175).
- Yokobori, T., S. Ohta et al. (2009). “Effect of land cover on air temperatures involved in the development of an intra-urban heat island”. In : *Climate research (Open Access for articles 4 years old and older)* 39.1, p. 61 (cf. p. 52, 53, 65, 69, 77, 107).

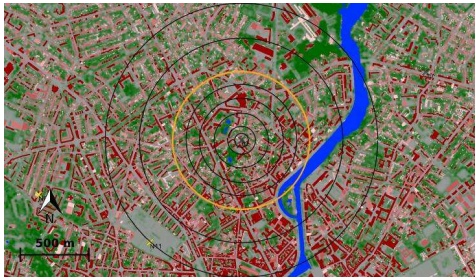






## **Contexte géographique de chaque site de mesure**

FIGURE A.1 – Contexte géographique des sites du jeu de données  $JDD_{calib}$ . Surface rouge clair à rouge foncé : emprise de bâtiments du plus bas au plus haut - surface grise à surface verte : surface imperméabilisée à surface de végétation dense - surface bleue : surface d'eau - croix : station de mesure



(a) Site 02



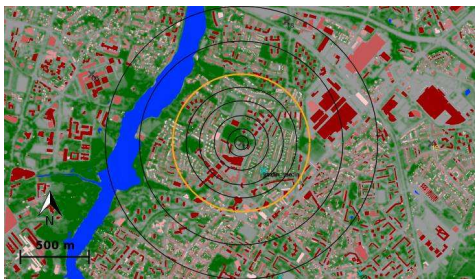
(b) Site 04



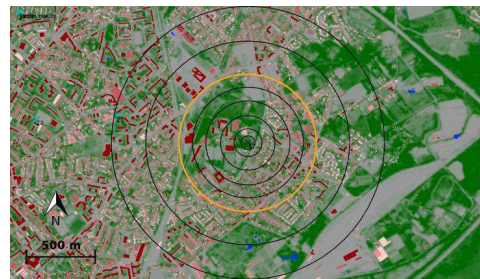
(c) Site 06



(d) Site 10



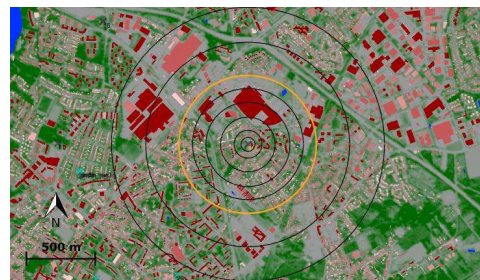
(e) Site 11



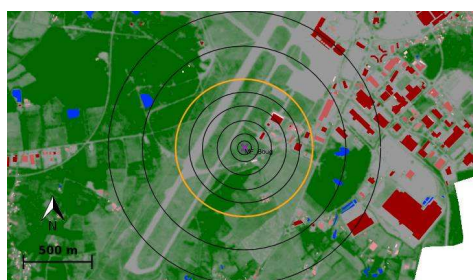
(f) Site 17



(g) Site 18



(h) Site 19



(i) Site MF\_Boug

FIGURE A.2 – Contexte géographique des sites du jeu de données  $JDD_{temp}$ . Surface rouge clair à rouge foncé : emprise de bâtiments du plus bas au plus haut - surface grise à surface verte : surface imperméabilisée à surface de végétation dense - surface bleue : surface d'eau - croix : station de mesure



(a) Site 02



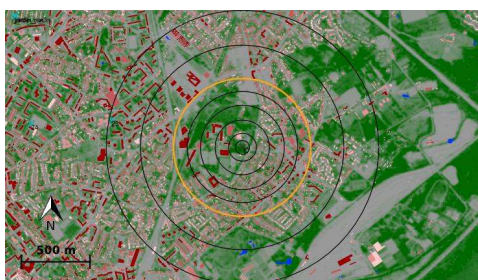
(b) Site 04



(c) Site 06



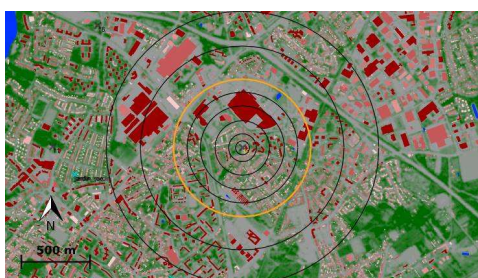
(d) Site 11



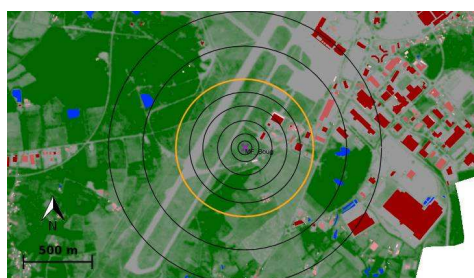
(e) Site 17



(f) Site 18



(g) Site 19



(h) Site MF\_Boug

FIGURE A.3 – Contexte géographique des sites du jeu de données  $JDD_{spat}$ . Surface rouge clair à rouge foncé : emprise de bâtiments du plus bas au plus haut - surface grise à surface verte : surface imperméabilisée à surface de végétation dense - surface bleue : surface d'eau - croix : station de mesure



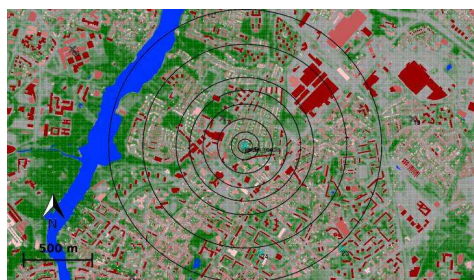
(a) Site 23



(b) Site 24



(c) Site 25



(d) Site jardin\_jardin



(e) Site jardin\_rue

FIGURE A.4 – Contexte géographique des sites du jeu de données  $JDD_{spattemp}$ . Surface rouge clair à rouge foncé : emprise de bâtiments du plus bas au plus haut - surface grise à surface verte : surface imperméabilisée à surface de végétation dense - surface bleue : surface d'eau - croix : station de mesure



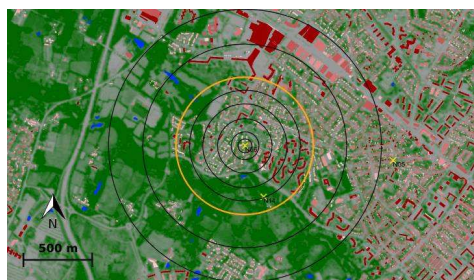
(a) Site N02



(b) Site N03



(c) Site N05



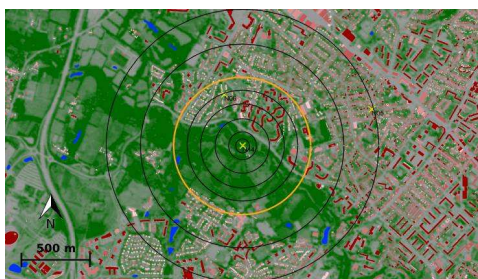
(d) Site N06



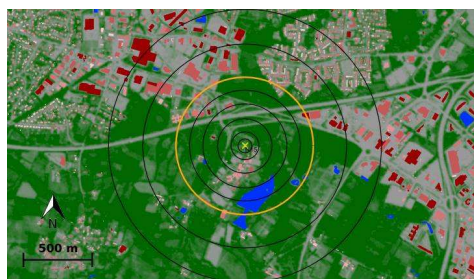
(e) Site N11



(f) Site N12



(g) Site N14



(h) Site N15

FIGURE A.5 – Contexte géographique des sites du jeu de données  $JDD^{(Angers)_{extrapol}}$ . Surface rouge clair à rouge foncé : emprise de bâtiments du plus bas au plus haut - surface grise à surface verte : surface imperméabilisée à surface de végétation dense - surface bleue : surface d'eau - croix : station de mesure



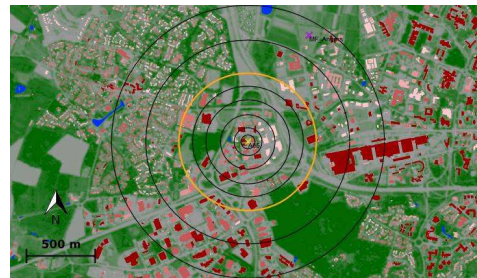
(a) Site A02



(b) Site A03



(c) Site A04



(d) Site A05



(e) Site A06



(f) Site A07

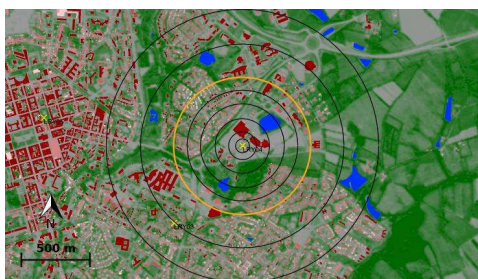
FIGURE A.6 – Contexte géographique des sites du jeu de données  $JDD_{extrapol}^{(Angers)}$ . Surface rouge clair à rouge foncé : emprise de bâtiments du plus bas au plus haut - surface grise à surface verte : surface imperméabilisée à surface de végétation dense - surface bleue : surface d'eau - croix : station de mesure



(a) Site LRY02



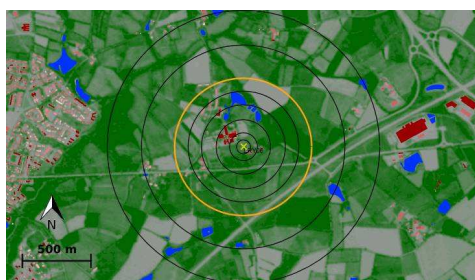
(b) Site LRY03



(c) Site LRY04



(d) Site LRY05



(e) Site LRY06



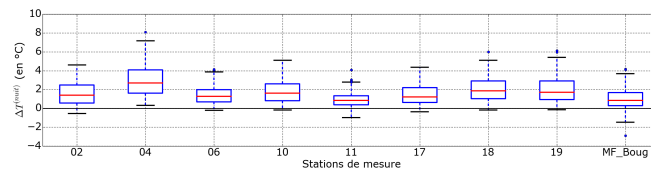
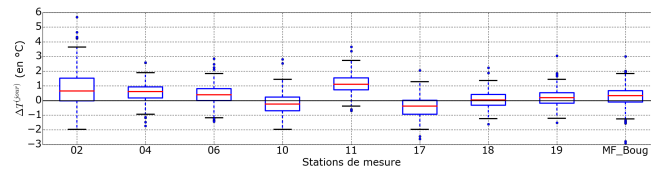




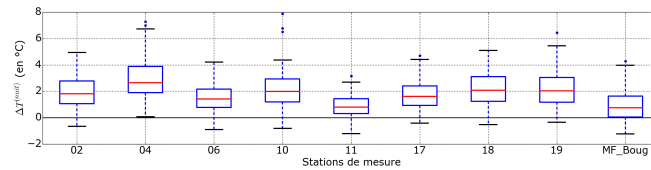
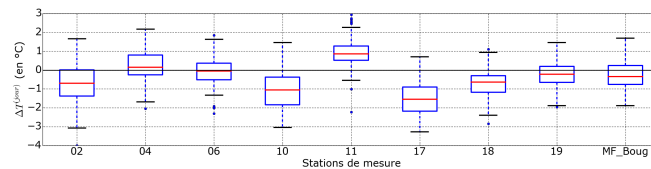
**B**

**Données climatiques de chaque site de  
mesure**

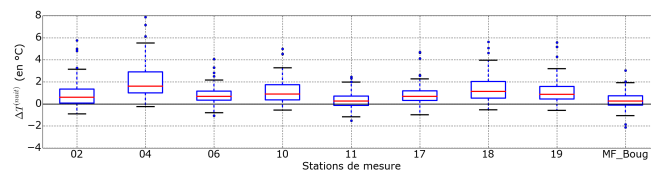
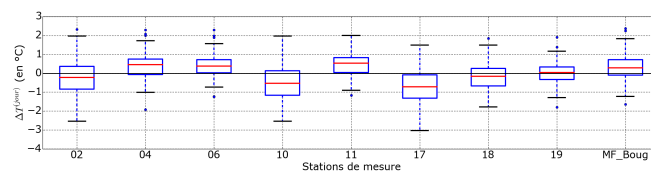
FIGURE B.1 – Données climatiques des sites du jeu de données  $JDD_{calib}$ . Boîte : ses limites hautes et basses correspondent aux centiles 0,25 et 0,75 - moustaches : ses valeurs basses et hautes correspondent aux centiles 0,25 et 0,75 auxquelles ont été respectivement soustraits ou additionnés l'interquartile  $1,5 \cdot \text{int}_{0,25}^{0,75}(\Delta T_i)$  - points bleus isolés : ce sont valeurs extrêmes supérieures aux limites des moustaches - segment rouge vertical : valeur médiane



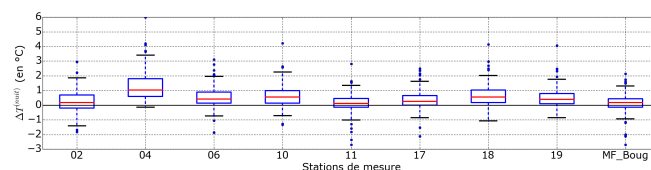
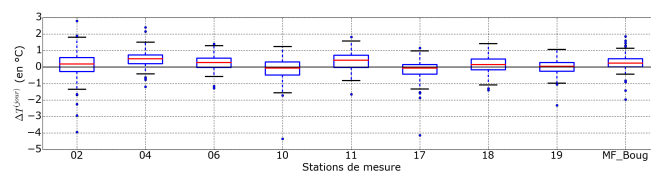
(a) Printemps



(b) Été

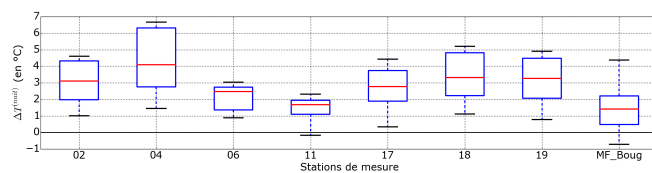
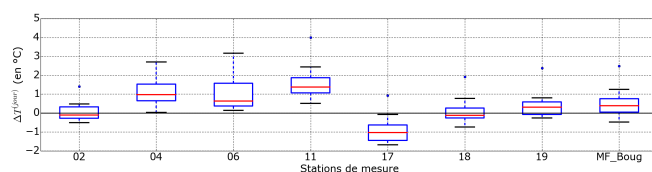


(c) Automne

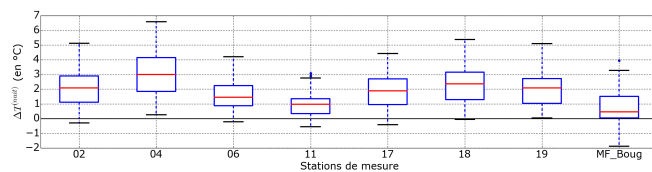
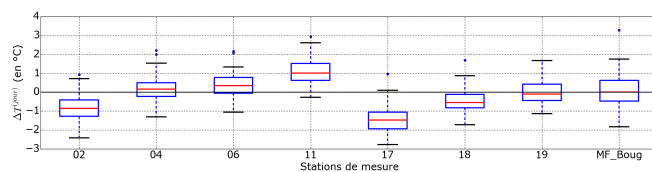


(d) Hiver

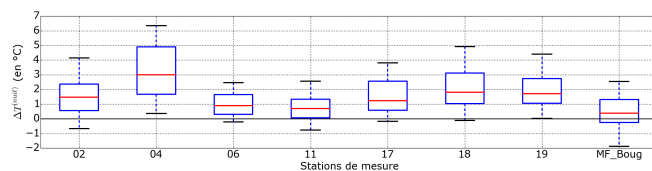
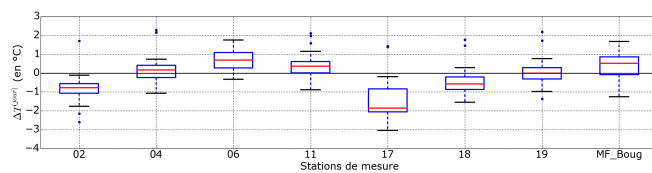
FIGURE B.2 – Données climatiques des sites du jeu de données  $JDD_{temp}$ . Boîte : ses limites hautes et basses correspondent aux centiles 0,25 et 0,75 - moustaches : ses valeurs basses et hautes correspondent aux centiles 0,25 et 0,75 auxquelles ont été respectivement soustraits ou additionnés l'interquartile  $1,5 \cdot \text{int}_{0,25}^{0,75}(\Delta T_i)$  - points bleus isolés : ce sont valeurs extrêmes supérieures aux limites des moustaches - segment rouge vertical : valeur médiane



(a) Printemps

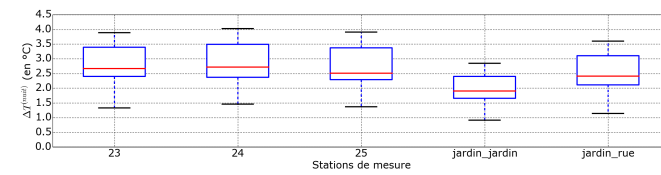
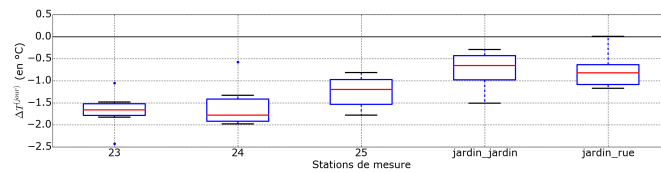


(b) Été

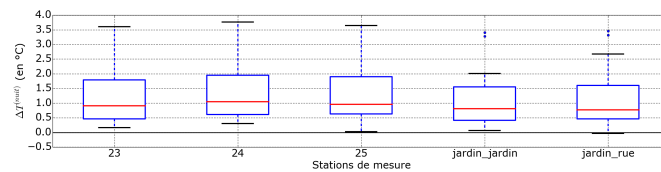
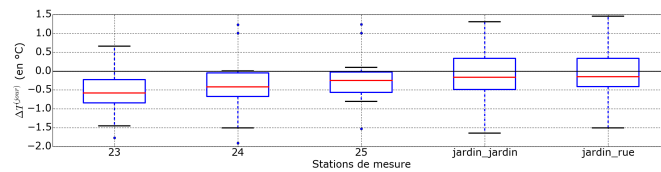


(c) Automne

FIGURE B.3 – Données climatiques des sites du jeu de données  $JDD_{spat}$ . Boîte : ses limites hautes et basses correspondent aux centiles 0,25 et 0,75 - moustaches : ses valeurs basses et hautes correspondent aux centiles 0,25 et 0,75 auxquelles ont été respectivement soustraits ou additionnés l'interquartile  $1,5 \cdot \text{int}_{0,25}^{0,75}(\Delta T_i)$  - points bleus isolés : ce sont valeurs extrêmes supérieures aux limites des moustaches - segment rouge vertical : valeur médiane

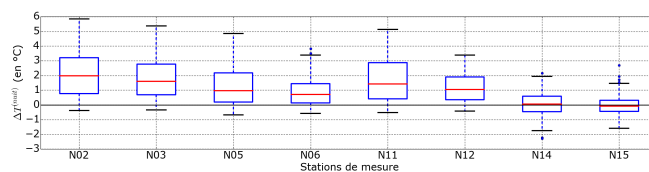
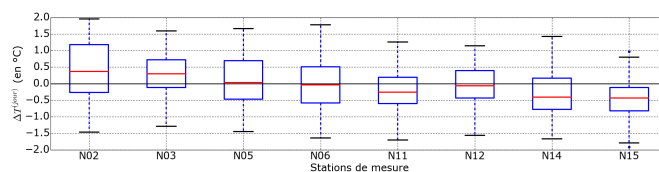


(a) Printemps

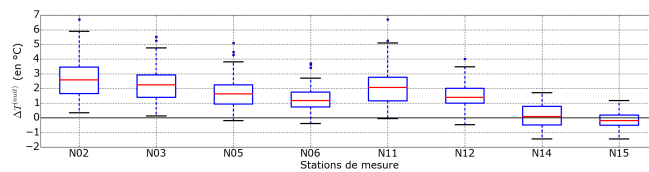
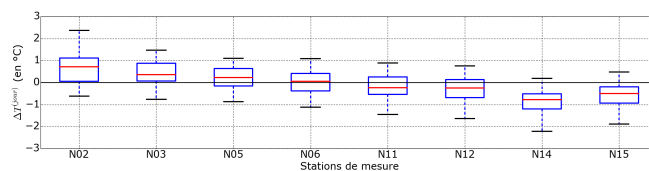


(b) Été

FIGURE B.4 – Données climatiques des sites du jeu de données  $JDD_{spattemp}$ . Boîte : ses limites hautes et basses correspondent aux centiles 0,25 et 0,75 - moustaches : ses valeurs basses et hautes correspondent aux centiles 0,25 et 0,75 auxquelles ont été respectivement soustraits ou additionnés l'interquartile  $1,5 \cdot \frac{0,75 - 0,25}{\Delta T_i}$  - points bleus isolés : ce sont valeurs extrêmes supérieures aux limites des moustaches - segment rouge vertical : valeur médiane

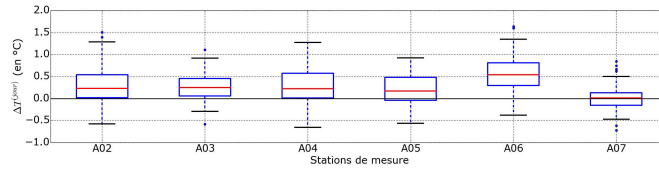


(a) Printemps

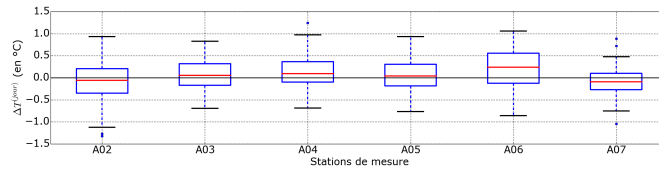


(b) Été

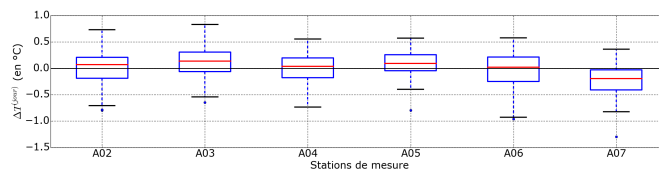
FIGURE B.5 – Données climatiques des sites du jeu de données  $JDD_{extrapol}^{(LRY)}$ . Boîte : ses limites hautes et basses correspondent aux centiles 0,25 et 0,75 - moustaches : ses valeurs basses et hautes correspondent aux centiles 0,25 et 0,75 auxquelles ont été respectivement soustraits ou additionnés l'interquartile  $1,5 \cdot \text{int}_{0,25}^{0,75}(\Delta T_i)$  - points bleus isolés : ce sont valeurs extrêmes supérieures aux limites des moustaches - segment rouge vertical : valeur médiane



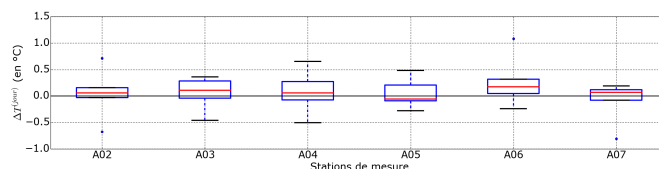
(a) Printemps



(b) Été

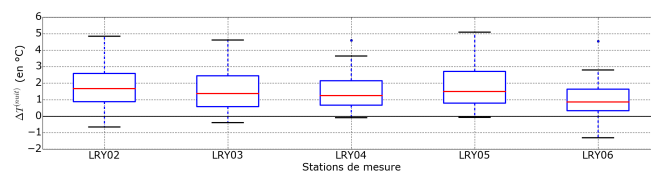
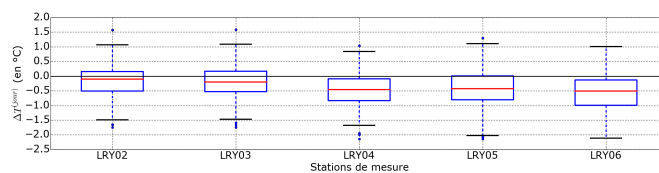


(c) Automne

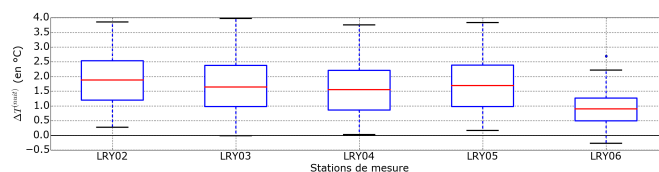
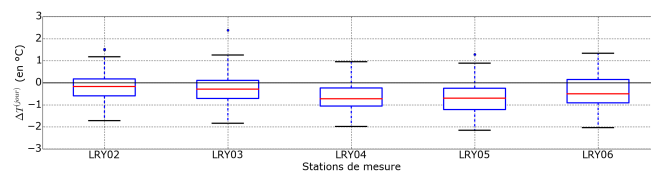


(d) Hiver

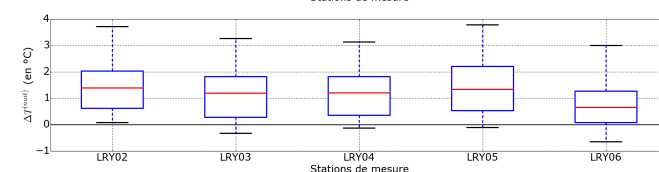
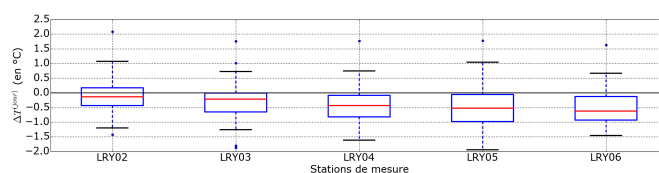
FIGURE B.6 – Données climatiques des sites du jeu de données  $JDD_{extrapol}^{(LRY)}$ . Boîte : ses limites hautes et basses correspondent aux centiles 0,25 et 0,75 - moustaches : ses valeurs basses et hautes correspondent aux centiles 0,25 et 0,75 auxquelles ont été respectivement soustraits ou additionnés l'interquartile  $1,5 \cdot \text{int}_{0,25}^{0,75}(\Delta T_i)$  - points bleus isolés : ce sont valeurs extrêmes supérieures aux limites des moustaches - segment rouge vertical : valeur médiane



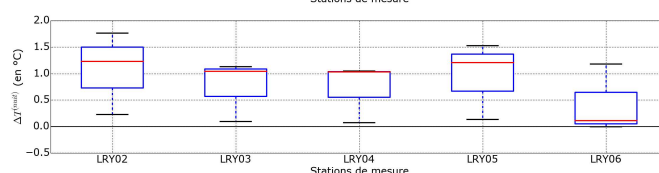
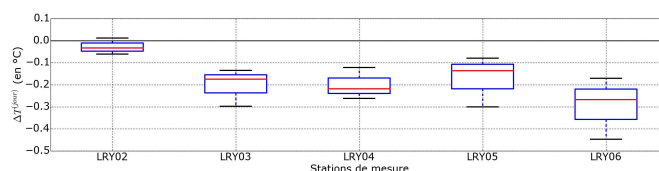
(a) Printemps



(b) Été



(c) Automne



(d) Hiver







**Jérémy BERNARD**

**Signature géographique et météorologique des variations spatiales et temporelles de la température de l'air au sein d'une zone urbaine**

**Geographic and weather signature of the spatial and temporal air temperature variations within an urban area**

## Résumé

Ce travail de thèse propose une modélisation des différences de température d'air observées au sein d'une agglomération. Elle repose sur des formulations simples établies à partir de l'observation des interactions qui existent entre la formation d'un microclimat, son contexte géographique et les conditions météorologiques. Cinq réseaux de capteurs de température d'air sont utilisés, permettant de qualifier les microclimats de trois agglomérations du grand ouest français : Nantes, Angers et La Roche-sur-Yon. Un état de l'art des méthodes d'analyse du climat urbain est réalisé au chapitre 1. Le chapitre 2 est consacré à la présentation et à l'analyse de chacun des réseaux de capteurs. Une méthode de correction des données erronées est également proposée. Au chapitre 3, les différences de température mesurées entre les sites sont expliquées à partir de régressions linéaires multiples utilisant des indicateurs géographiques. Ces indicateurs, descriptifs des formes urbaines et de la typologie des sols de chaque site, sont calculés au sein d'un cercle de rayon de 500 m centré sur chaque station. L'intensité de l'îlot de chaleur urbain est modélisé à partir d'une régression linéaire multiple utilisant des variables météorologiques telles que la vitesse du vent ou le rayonnement solaire comme variables explicatives. Préalablement à la calibration, les données sont triées de différentes manières afin d'élaborer plusieurs modèles. Le plus performant est identifié au chapitre 4 en confrontant les valeurs mesurées par les cinq réseaux de capteurs à celles estimées par chacun des modèles. Au chapitre 5, trois exemples d'application possible du travail réalisé sont présentées.

## Mots clés

îlot de chaleur urbain, observation, cartographie, modélisation, analyse de données

## Abstract

This thesis presents a method to model air temperature differences within a conurbation . It relies on the analysis of the existing relationships between urban microclimate, geographical context and weather conditions. To do so, five air temperature networks located in three western France conurbations are utilized : Nantes, Angers and La Roche-sur-Yon. A state of the art of the methods used to analyze urban climate is realized in chapter 1. In Chapter 2, air temperature networks are presented and analyzed and a methodology to correct biased data is proposed. In Chapter 3 air temperature differences observed between the network sites are modelled from multiple linear regressions, using weather variables such as wind speed or solar radiation as explicative variables. Before the calibration step, data is sorted in a variety of ways in order to elaborate several models. The most performant model is identified in chapter 4, confronting data observed using the five temperature networks to data estimated from each of the models. Chapter 5 illustrates the results of the work through three practical applications.

## Key Words

urban heat island, observation, mapping, modelisation, data analyse