

# Développement d'un algorithme d'affectation logements–bâtiments pour le paramétrage automatique de modèles énergétiques de quartiers résidentiels

Louis JOANNARD<sup>1</sup>, Victor MARTY-JOURJON<sup>1</sup>, Thomas BERTHOU<sup>1\*</sup>

*<sup>a</sup>Mines ParisTech, PSL Research University, Centre d'Efficacité énergétique des Systèmes (CES), 60 boulevard Saint Michel, 75272 Paris Cedex 06,*

*\*thomas.berthou@mines-paristech.fr*

---

*RESUME. Afin d'aider les acteurs du bâtiment à identifier des actions d'efficacité énergétique adaptées aux spécificités locales des territoires, des outils de simulation énergétique de quartiers et de villes sont développés. Pour que ces outils puissent modéliser des centaines de bâtiments de manière réaliste en un temps raisonnable, il est nécessaire de disposer d'algorithmes de paramétrage des modèles automatiques. Ce sont des «enrichisseurs» qui transforment les bases de données de description du territoire en paramètres d'entrée pour les modèles de quartier : systèmes, bâtis et occupants. En France, deux bases de données sont principalement utilisées : d'une part le recensement INSEE qui décrit les logements de chaque IRIS et d'autre part la BDTOPO qui renseigne la forme et l'usage des bâtiments. Cette étude propose d'évaluer un algorithme d'affectation des logements (décrits dans la base INSEE), aux bâtiments à usage résidentiel collectif (renseignés dans la BDTOPO). Les tests de validation de l'algorithme sont réalisés sur trois cas d'étude.*

*MOTS-CLÉS : modèles énergétiques de quartiers, enrichisseurs, algorithme d'affectation*

---

*ABSTRACT. In order to help building stakeholders to identify the energy efficiency actions most suited to the local specificities of territories, tools for energy simulation of districts and cities have been developed. To be able to simulate hundreds of buildings realistically and in a reasonable amount of time, automatic parameterization algorithms are required. These algorithms treat the information available in territorial databases to build input parameters for the districts models: systems, buildings and occupants. In France, two main databases are used: the INSEE census, which describes the dwellings in each IRIS, and the BDTOPO which provides information on the shape and use of buildings. This study proposes to evaluate a matching algorithm to allocate dwellings (described in the INSEE census) to the collective residential buildings described in the BDTOPO. Validation tests are carried out on three case studies.*

*KEYWORDS: district energy model, enrichment algorithms, matching algorithm*

---

## 1. INTRODUCTION

Le secteur du bâtiment consomme environ 44% de l'énergie finale (ADEME, 2013). Ce secteur clé de la transition énergétique est difficile à appréhender du fait de l'hétérogénéité des usages (résidentiel, tertiaire...), des énergies et systèmes utilisés (gaz, électricité, chaleur) et des techniques constructives. C'est également un secteur où de nombreuses solutions existent pour accélérer la transition énergétique : sobriété, rénovation, intégration d'énergie renouvelable, pilotage de la demande.

Afin d'aider les acteurs du bâtiment à identifier les actions d'efficacité énergétique les mieux adaptées aux spécificités locales des territoires, des outils de simulation énergétique de quartiers et de villes ont été développés (Hong et al., 2020), (Fonseca et al., 2016), (Berthou et al, 2019), (Riederer et al., 2015).

Pour que ces outils puissent modéliser des centaines de bâtiments de manière réaliste en un temps raisonnable, il est nécessaire de disposer d'algorithmes de paramétrage automatique. Ce sont des « enrichisseurs » qui transforment les bases de données de description du territoire en paramètres d'entrée pour les modèles de quartier : systèmes, bâtis et occupants.

En France, deux bases de données sont principalement utilisées pour alimenter les modèles énergétiques de quartier : le recensement (INSEE, 2019) et la BDTOPO (IGN, 2019). Le recensement fournit une liste des logements localisés à l'échelle de l'IRIS<sup>1</sup>, caractérisés par un grand nombre d'attributs (surface, années de construction, occupants, chauffage). La BDTOPO fournit les emprises au sol géolocalisées des bâtiments, caractérisés par des hauteurs, des usages et des années de construction. Pour profiter au mieux de ces informations, il est nécessaire, pour chaque IRIS, d'affecter les bons logements dans les bons bâtiments, en fonction des attributs communs des deux bases de données. Ce problème d'optimisation est complexe, car il est non convexe et présente un nombre de combinaison très élevé. De plus, les deux bases de données sont corrompues par de nombreuses informations fausses ou manquantes.

Cette étude propose donc un algorithme d'affectation des logements (décrits dans le recensement INSEE), aux bâtiments à usage résidentiel collectif décrit dans la BDTOPO. Les algorithmes classiques d'affectation simple comme l'algorithme Hongrois (Munkres, 1957) ou l'algorithme du mariage stable (Knuth, 1976), (Gale et Shapley, 1962) ont été écartés, car ils ne permettent pas d'intégrer certaines contraintes spécifiques au problème. Ainsi, un algorithme d'optimisation heuristique est proposé, il sera évalué sur trois cas d'étude :

- un quartier ad hoc pour lequel il existe solution optimale connue a priori,
- un second cas similaire au premier, mais comportant 20 % d'informations corrompues,
- un quartier réel situé à Dammarie-les-Lys (77).

## 2. PRESENTATION DE SMART-E

Smart-E est un modèle énergétique associé à une architecture informatique qui permet de simuler l'ensemble des consommations d'énergie liées aux bâtiments à l'échelle d'un territoire : ville, région, parc fictif ou réel (Berthou et al., 2015). La plateforme est alimentée par des bases de données locales et des statistiques nationales (concernant le bâti, les usagers, et les systèmes), dont principalement les bases de recensement INSEE. Ce paramétrage « automatique » des modèles permet de simuler rapidement des consommations d'énergie et d'évaluer des stratégies de maîtrise de la demande sur n'importe quel territoire en France.

Afin d'améliorer la justesse des simulations de Smart-E, l'objectif est d'améliorer la prise en compte de la base de données BDTOPO décrivant la forme et la localisation des bâtiments sur toute la France. Cette base de données étant hétérogène et comportant de nombreuses informations manquantes, il est nécessaire de développer un algorithme d'affectation robuste permettant d'associer de manière probabiliste ces deux bases.

Aujourd'hui, dans Smart-E les informations issues de la BDTOPO permettent d'associer des compacités ou des facteurs de masques solaires médians par IRIS et par typologie de bâtiments (ex.

---

<sup>1</sup> L'IRIS « Ilots Regroupés pour l'Information Statistique », constitue la maille de base de la diffusion infra communale, il correspond à un quartier ayant un contour identifiable ou villes de moins de 10 000 habitants

logements individuels, mitoyens ou collectifs). A l'aide d'un algorithme d'affectation logements-bâtiments, il sera possible d'attribuer une compacité ou un masque solaire spécifique à chaque logement.

Un autre intérêt de cet exercice est de pouvoir localiser les consommations (issues des simulations Smart-E) à l'adresse (ou au bâtiment) et non uniquement au quartier. Ceci permet d'élargir l'usage de modèles énergétique urbain à d'autres problématiques comme l'évaluation d'extension de réseaux d'énergies (ex. chaleur) et l'identification de phénomènes de congestions liés à la production locale : d'origine photovoltaïque pour l'électricité ou injection de biogaz pour le gaz naturel.

La question de l'attribution logement-bâtiment se pose principalement pour les logements collectifs, en effet pour les maisons individuelles il est possible d'associer directement les géométries issues de la BDTOPO aux surfaces issues du recensement avec l'aide des années de construction et des surfaces (à l'aide de l'algorithme des mariages stables par exemple). Il faudra tout de même proposer une méthode robuste face aux données manquantes ou obsolètes. La suite de cette étude ne traitera pas le cas des maisons individuelles.

### 3. DESCRIPTION DES BASES DE DONNEES

Le recensement LOGEMENT (« fichier détail recensement de la population ») décrit 40% logements localisés à l'IRIS (100% pour les communes de moins de 10 000 habitants) à l'aide d'environ 60 attributs qui décrivent le bâti (superficie, année de construction, énergie pour le chauffage...), les occupants (nombre, âge, statut d'occupation...) et l'environnement du logement (type d'iris, commune, poids statistique du logement...). Environ 8% de la base de données est mise à jour chaque année par l'envoi de questionnaires aux particuliers. Les réponses aux questions peuvent être approximatives (pour la surface par exemple) ou fausses (pour l'énergie principale de chauffage par exemple). Ces erreurs doivent être prises en compte lors de l'exploitation de la base de données, soit en les corrigeant lorsqu'elles sont détectables, soit en rendant les enrichisseurs robustes aux valeurs aberrantes.

La BDTOPO décrit précisément la forme (contour et hauteur) et la localisation de tous les bâtiments de France. Depuis 2019 et grâce à l'appareillement entre la BDTOPO et les fichiers fonciers, l'IGN a ajouté plusieurs attributs à chaque bâtiment :

- l'usage (résidentiel, commerciale, indifférencié...),
- l'année de construction,
- le nombre d'étages (pour le secteur résidentiel),
- le nombre de logements (pour le secteur résidentiel),
- un indice de qualité de l'appareillement.

Ces données sont utiles pour le paramétrage automatique des modèles énergétiques de quartier, mais doivent être utilisées avec prudence, car elles présentent des valeurs aberrantes ou des valeurs manquantes. La qualité varie en fonction des communes ce qui rend la méthode présentée ici encore peu généralisable.

Les attributs utilisés dans cette étude, issus de ces deux bases de données, sont présentés dans le tableau 1.

Base de données	attributs
BDTOPO	- Surface utile du bâtiment (estimée)
	- Année de construction
	- Nombre de logements
INSEE	- Surface des logements (par tranches)
	- Année de construction (par tranche)
	- Energie principale de chauffage

Tableau 1 : Attributs utilisés pour l'affectation logements-bâtiments

#### 4. CHOIX D'UN ALGORITHME D'AFFECTION

On souhaite répondre à la question suivante : comment répartir au plus proche de la réalité les différents logements décrits dans le recensement, dans les bâtiments décrits dans la BDTOPO ?

Il s'agit alors d'un problème d'affectation. On peut distinguer les problèmes d'affectation simples des problèmes d'affectation généralisés. Le problème d'affectation simple est un cas particulier qui consiste à associer deux à deux,  $n$  éléments d'un ensemble A à  $m$  éléments d'un ensemble B. Ce problème est rependu et peut notamment être résolu par l'algorithme des mariages stables ou l'algorithme hongrois. Cependant dans le cas des bâtiments résidentiels collectifs, chaque bâtiment peut contenir plusieurs logements et possèdent une contrainte de capacité, on parle alors d'un problème d'affectation généralisé.

Nous proposons d'utiliser un algorithme heuristique permettant de s'approcher rapidement d'une solution acceptable sans pour autant être optimale. La solution proposée est un algorithme d'optimisation par permutations aléatoires. A chaque permutation de  $n$  logements (d'un bâtiment à un autre), une fonction objectif est calculée et si la valeur de la fonction après permutation est réduite, la modification est conservée, sinon le système revient à son état initial. Le processus continue jusqu'à atteindre la condition d'arrêt de l'algorithme, ici correspondant à 10 min de temps de calcul. Avant la première itération, le parc de bâtiments est initialisé de manière aléatoire, avec le bon nombre de logements par bâtiment. Cette méthode d'optimisation a l'avantage d'être robuste aux imperfections des bases de données et de proposer des solutions acceptables dans un temps raisonnable. De plus, cet algorithme peut être enrichi simplement en ajoutant des paramètres supplémentaires à la fonction objectif.

La fonction « objectif » est présentée en équation 1, avec  $NbB$  le nombre de bâtiments dans l'IRIS. Elle est construite à partir de trois termes :

- La somme unitaire des écarts absolus entre les surfaces habitables issues de la BDTOPO et issues du recensement ( $t1 = \Delta_{surface}/NbB$ ),
- La somme unitaire des écarts absolus entre les années de construction issues de la BDTOPO et issues du recensement ( $t2 = \Delta_{année}/NbB$ ),
- La variance unitaire des énergies de chauffage ( $t3 = Var(chauffage)/NbB$ ), en effet on suppose que l'ensemble des logements d'un même bâtiment a une énergie principale de chauffage identique. Pour calculer la variance, nous associons la valeur « 1 » à tous les logements connectés aux chauffages urbains et la valeur « 2 » pour les logements chauffés à l'aide d'une autre énergie.

$$Obj = (1 + \Delta_{surface}/NbB) \times (1 + \Delta_{année}/NbB) \times (1 + Var(chauffage)/NbB) \quad (1)$$

## 5. DESCRIPTION DES CAS D'ETUDES

L'algorithme d'affectation est évalué à l'aide de trois cas d'études qui présentent une difficulté croissante.

Le premier cas d'étude est un quartier théorique ad hoc composé de 16 bâtiments résidentiels collectifs et de 400 logements. Il se rapproche d'un IRIS de taille moyenne. Il fait office de référence dans la mesure où il existe une solution unique connue pour laquelle la fonction objectif est égal à 1. Pour sa construction, un quartier parfaitement décrit a été séparé en deux bases de données distinctes, une première qui décrit les bâtiments similaire à la BDTPOPO et une seconde qui décrit les logements similaire à la base INSEE. De plus, les années de construction sont différentes pour chaque bâtiment. Le but de ce cas d'étude est de vérifier la capacité de l'algorithme à converger vers une solution proche de l'optimum global connu en un temps raisonnable.

Le second cas d'étude est proche du premier sauf que les données de description des logements ont été corrompues. Ainsi, 20% des années de construction des logements ont été remplacées de manière aléatoire par d'autres années présentes dans la base de données, la superficie de 20% des logements a été multipliée aléatoirement par des coefficients de 0,9 et 1,1 et l'attribut binaire correspondant à la catégorie de chauffage des logements a été inversé pour 20% des données. Ce second cas permet d'évaluer l'algorithme sur un quartier plus réaliste pour lequel une solution parfaite annulant la fonction objectif n'existe probablement pas.

Pour le troisième cas d'étude, nous avons utilisé les données réelles du recensement et de la BDTPOPO sur un IRIS de la commune de Dammarie-les-Lys. Cet IRIS, composé de 1370 logements et de 30 bâtiments (figure 1 et tableau 2), a été choisi pour plusieurs raisons. Il est composé à majorité de bâtiments résidentiels collectifs, les dates de construction de bâtiments présentent une variabilité satisfaisante ce qui va « aider » l'algorithme d'affectation, les attributs de la BDTPOPO sont correctement renseignés (dates, usages et nombre de logements) et enfin l'énergie de chauffage est principalement du chauffage urbain. Un prétraitement des données a également été appliqué:

- Les années de constructions sont renseignées par tranches homogènes dans les deux bases.
- Les surfaces d'immeubles (ie. surfaces de plancher multipliées par le nombre d'étages) ont été multipliées par 0.66, ce facteur, différent pour chaque IRIS, permet d'évaluer des surfaces habitables à partir de surface bâtie. Il est calculé pour que, dans un IRIS, la somme des surfaces habitables issues de la BDTPOPO soit égale à la somme des surfaces habitables issues du recensement.
- Le nombre de logements présents dans la BDTPOPO a été modifié afin qu'il corresponde à la valeur du recensement (un écart de 4.5 % avait été constaté).



Figure 1 : carte du cas d'étude N°3 issue de la BDTOPO.

Commune	Code de l'IRIS	Proportion d'appartements (%)	Surface Bâti BDTOPO (m <sup>2</sup> )	Nombre de logements BDTOPO	Surface habitable INSEE (m <sup>2</sup> )	Nombre de logements INSEE
Dammarielles-Lys	771520301	99	142512	1433	94388	1371.5

Tableau 2 : Description du quartier utilisé pour le cas d'étude 3

## 6. RESULTATS

Une première évaluation a été réalisée en faisant varier le nombre de logements permutés par itération (de 2 à 6). Dans tous les tests réalisés, une permutation de 2 logements par itération était systématiquement plus performant en termes de minimisation de la fonction objectif et de temps de calcul. La méthode proposée a été appliquée sur les trois cas d'étude. Les résultats sont résumés dans le tableau 3. La figure 2 présente un exemple de vitesse de convergence de l'algorithme d'affectation sur le cas d'étude 3.

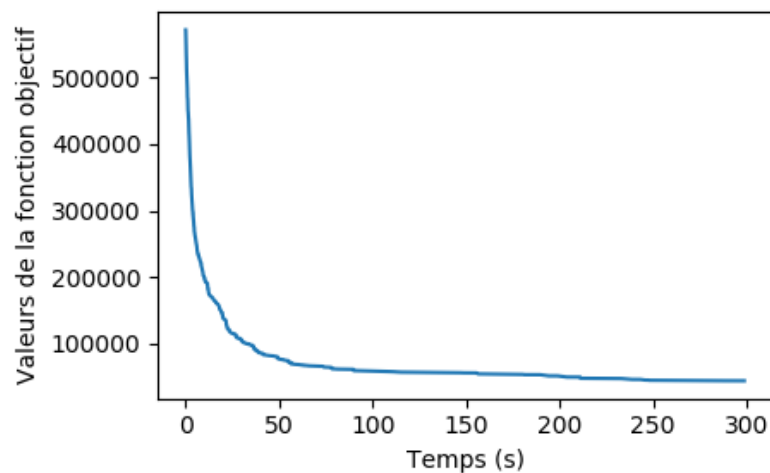


Figure 2: Exemple de vitesse de convergence de l'algorithme sur le quartier réel

	Valeur de la fonction Objectif à t = 0 min (initialisation)	Valeur de la fonction Objectif à t = 5 min (t1, t2, t3)	Valeur de la fonction Objectif à t = 10 min (t1, t2, t3)
Quartier fictif	467802	246 (1.75, 130, 1.08)	241 (1.6, 135, 1.1)
Quartier fictif corrompu	449231	669 (7.3, 82, 1.1)	600 (7.5, 75.9, 1.1)
Quartier réel	533955	47676 (257, 176, 1.1)	40970 (254, 155, 1.03)

*Tableau 3 : Evaluation de l'algorithme d'affectation sur les 3 cas d'études, entre parenthèses les valeurs des 3 termes de la fonction objectif*

Pour le cas d'étude 1, l'algorithme permet de diviser la valeur de la fonction objectif d'un facteur 2000 après 5 minutes de calcul. Cependant, même si une solution unique existe elle n'est pas identifiée et encore 25% des logements ne sont pas associé au bon bâtiment après 10 minutes de calcul. Pour le cas d'étude 2, l'algorithme permet de diviser la valeur de la fonction objectif par 670 en 5 minutes. Enfin, pour le cas d'étude 3, l'algorithme d'affectation permet de diviser la valeur de la fonction objectif par 13 en 10 minutes. Après 10 minutes de calcul, les solutions trouvées ne sont pas encore stabilité, mais les permutations acceptées (celles qui améliorent la fonction objectif) sont de plus en plus rares. Les résultats après 20 minutes de calcul ont montré de très faibles améliorations par rapport aux résultats à 10 minutes de calcul.

## 7. CONCLUSION ET PERSPECTIVES

Cette étude a permis d'évaluer un algorithme naïf d'affectation, sur 3 quartiers réalistes et propose une réponse simple à un problème complexe. L'efficacité de l'algorithme est satisfaisante puisqu'il permet d'identifier, sur un IRIS réel, une solution plus juste qu'une association totalement aléatoire (division par 13 de la fonction objectif) en 10 minutes de calcul sur un ordinateur de bureau et sans optimisation ou parallélisation du code informatique. Cet algorithme permettra d'améliorer les enrichisseurs automatiques de modèle énergétique de quartier et sera à terme intégré dans Smart-E. Il permettra un raffinement du calcul des surfaces déperditives et des masques solaires de chaque logement. Cependant il ne semble pas être encore assez fiable pour localiser précisément les consommations d'énergie au bâtiment. Il est nécessaire de faire des testes supplémentaires dans d'autres IRIS pour vérifier la généralité de ces résultats, notamment en choisissant des IRIS moins bien décrits et avec une faible variabilité dans les années de constructions et les énergies de chauffage.

D'autres algorithmes heuristiques pourraient être évalués comme les algorithmes génétiques ou les essais particuliers qui semblent particulièrement adaptés à ce type de problèmes. Aussi des méthodes existent pour transformer un problème d'affectation généralisé en problème d'affectation simple comme la méthode de relaxation Lagrangienne (Draghici, 2005). Il serait également intéressant de proposer des méthodes pour corriger les données aberrantes et manquantes avant l'optimisation afin de faciliter l'étape d'association logements-bâtiments.

## 8. BIBLIOGRAPHIE

ADEME (2013), Les chiffres clés du Bâtiment

BDTOPO (2019), Descriptif de contenu - version 3.0, IGN

Berthou, T., Duplessis, B., Rivière, P., Stabat, P., Casetta, D., & Marchio, D. (2015). Smart-E: A tool for energy demand simulation and optimization at the city scale. 14th International Conference of IBPSA - Building Simulation 2015

Berthou, T., Duplessis, B., Stabat, P., Rivière, P., & Dominique, M. (2019). Urban Energy Models Validation in Data Scarcity Context: Case of the Electricity Consumption in the French Residential Sector. 16th International Conference of the “International Building Performance Simulation Association”.

Draghici, Carmen, (2005), Modélisation et conception d’algorithmes pour la planification automatique du personnel de compagnies aériennes, thèse de doctorat, l’Institut National de Sciences Appliquées de Toulouse

Fichier Détail Recensement de la population, INSEE (2019) en ligne : <https://www.insee.fr/fr/information/4172214> (visite le 06/03/2020)

Fonseca, J. A., Nguyen, T. A., Schlueter, A., & Marechal, F. (2016). City Energy Analyst (CEA): Integrated framework for analysis and optimization of building energy systems in neighborhoods and city districts. *Energy and Buildings*, 113, 202–226.

Gale, D., Shapley L.S., (1962), College Admissions and the Stability of Marriage, *The American Mathematical Monthly*, Vol. 69

Hong, T., Chen, Y., Luo, X., Luo, N., & Lee, S. H. (2020). Ten questions on urban building energy modeling. *Building and Environment*, 168.

Knuth, Donal E. (1976), *Mariage stables et leurs relations avec d’autres problèmes combinatoires*, Presse de l’Université de Montréal

Munkres, J. (1957) Algorithms for the Assignment and Transportation Problems. *Journal of the Society for Industrial and Applied Mathematics*, 5, 32-38

Riederer, P., Partenay, V., Perez, N., Nocito, C., Trigance, R., & Guiot, T. (2015). Development of a simulation platform for the evaluation of district energy system performances. 14th International Conference of IBPSA - Building Simulation 2015, Conference Proceedings, 2499–2506.