

UVCI
Université Virtuelle
de Côte d'Ivoire

Ministère de l'Enseignement
Supérieur et de la
Recherche Scientifique

Organisation
des Nations Unies
pour l'éducation,
la science et la culture

COLLOQUE SCIENTIFIQUE INTERNATIONAL

Technologies IoT, Big data, IA, Blockchain et protection des données et de la vie privée dans les sociétés connectées

SOUS-THEME 1
Intelligence Artificielle et Big Data

SOUS-THEME 2
Traitement du Signal et de l'Image

SOUS-THEME 3
IoT, Cybersécurité et Blockchain

26 - 27
Octobre 2023

Siège de l'UVCI
et en ligne

Plus d'info ———
✉ colloquevirtuel@uvci.edu.ci
🌐 <https://colloque.uvci.edu.ci/iot>

Partenaires

CONSEIL AFRICAIN
ET MALGACHE POUR
L'ENSEIGNEMENT SUPÉRIEUR

AUF

EA4it

Titre de la communication :

Apports de la boîte à outils Advanced Data Analysis (AdvDatAna) pour le traitement automatisé des bases de données météorologiques : cas du parc de la Lopé.

Donald Romarick ROTIMBO MBOUROU¹; Djes-Frézy BILENGA MOUKODOUMA^{1,2}; Christiane ATTEKE NKOULEMBENE³; Christophe DENIS^{2,4}

1. *Laboratoire de Recherche Multidisciplinaire en Environnement (LARME), Université des Sciences et Techniques de Masuku (USTM), Département de Chimie, Mbaya, Route Nationale 1, BP 901 Franceville, GABON;
Téléphone +24174932936 (W)/ +24160246524, Courriel (rotimboromarick@yahoo.fr);*
2. *Unité de Modélisation Mathématique et Informatique des systèmes complexes (UMMISCO), UMI 209, 32 Avenue Henri Varagnat, 93143 Bondy Cedex FRANCE;*
3. *Département de Biologie, Université des Sciences et Techniques de Masuku (USTM), Mbaya, Route Nationale 1, BP 901 Franceville, GABON;*
4. *Laboratoire Informatique de Paris 6 (LIP6), LIP6 Boîte courrier 169 Tour 26, Couloir 26-00, 2^è étage 4 place Jussieu 75252 Paris Cedex 05 France.*

PLAN

INTRODUCTION

*II. AdvDatAna appliquée à l'étude
des données météorologiques
du parc de la Lopé*

CONCLUSION

I. Présentation de la boîte à outils

*Advanced Data Analysis - AdvDatAna
(Analyse Avancée des Données)*

*III. Recommandations générales sur
AdvDatAna et les données
météorologiques*

Traitement de bases de données (ou données personnelles) :

- Synoptique bien déterminée (Fig.);
- Quelques prérequis :
 - Connexion internet ponctuelle/permanente (téléchargement des données);
 - **Ordinateur + processeur correct (environnement de travail);**
 - **Logiciels gratuits/payants (traitement des données).**
 - Disque dur interne/externe (stockage des données);

Carences diverses par "les usagers non experts" :

- Prise en main des interfaces du logiciel de traitement de données choisi (langue, fonctionnement, aide, FAQs, ...);
- Création des routines de traitement (Statistique, Traitement du Signal) faisant généralement appel à de la programmation informatique;**
- Affichage et/ou enregistrement des résultats pour les grands jeux de données.

Est-il possible pour "les usagers non experts " de combler leurs carences grâce à des outils intuitifs



Choix et téléchargement de la base de données en ligne (ou création personnelle)

A1



Importation dans un logiciel de traitement (matriciel, statistique, ...)

A2



Création des routines de traitement (fonctions, modèles, ...)

A3



Exécution des routines de traitement (simulation)

A4



Génération des résultats (données brutes, graphiques, ...)

A5

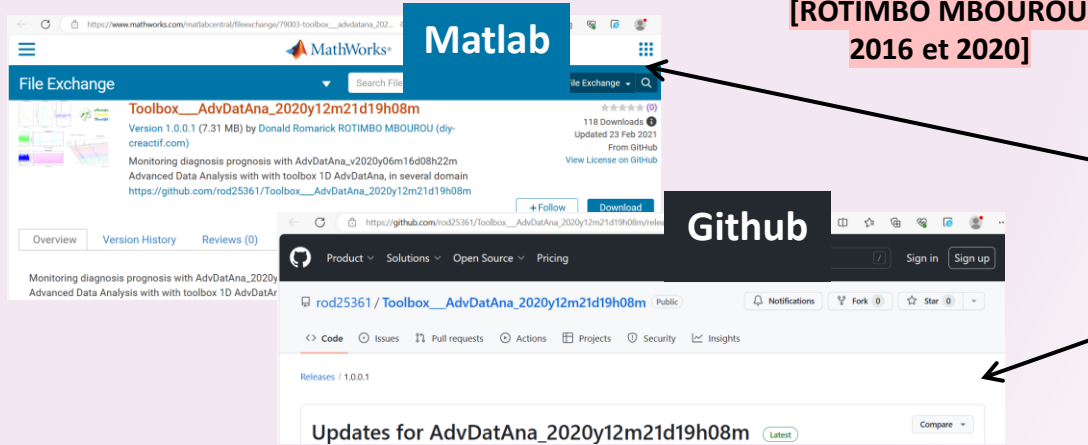


Enregistrement des résultats dans des divers fichiers

A6

Fig. : Synoptique général menant des données aux résultats.

Advanced Data Analysis : Analyse Avancée des Données.



[ROTIMBO MBOUROU
2016 et 2020]

Téléchargez le 1^{er} fichier **Toolbox__AdvDatAna_2020y12m21d19h08m-1.0.0.1.zip** depuis l'un des 2 dépôts officiels **Matlab File Exchange** (*inscription en ligne requise pour le téléchargement*) ou **Github** :

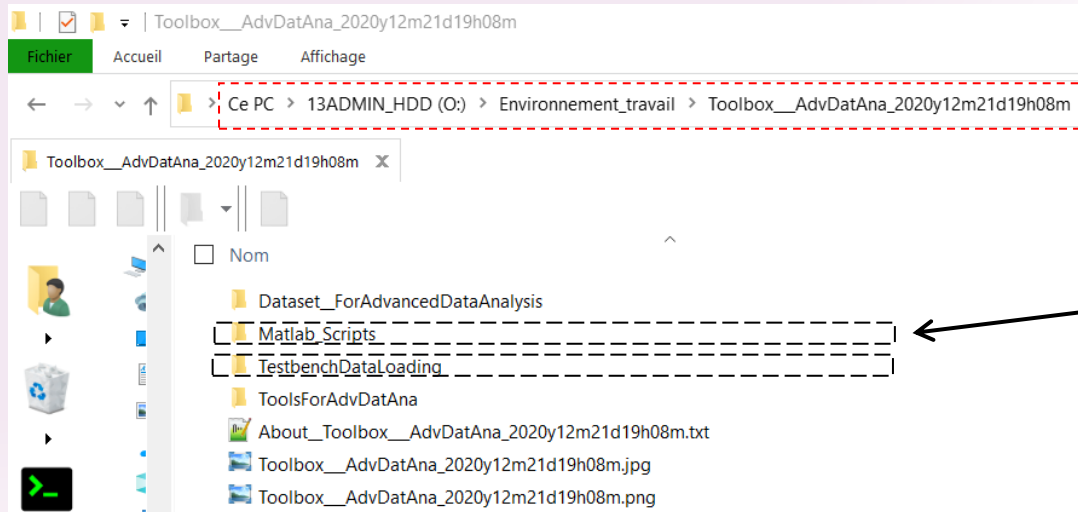
https://www.mathworks.com/matlabcentral/fileexchange/79003-toolbox_advdatana_2020y12m21d19h08m

https://github.com/rod25361/Toolbox__AdvDatAna_2020y12m21d19h08m/releases/tag/1.0.0.1

B1



AdvDatAna : Arborescence des fichiers.



Dézippez le fichier téléchargé et accédez au dossier **Toolbox__AdvDatAna_2020y12m21d19h08m-1.0.0.1**. S'il n'est pas déjà dézippé, dézippez le 2^{ème} fichier

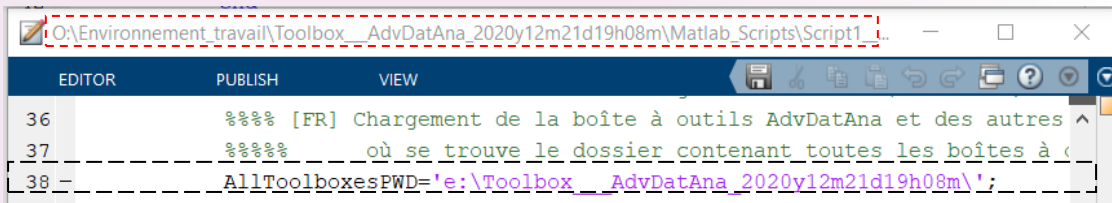
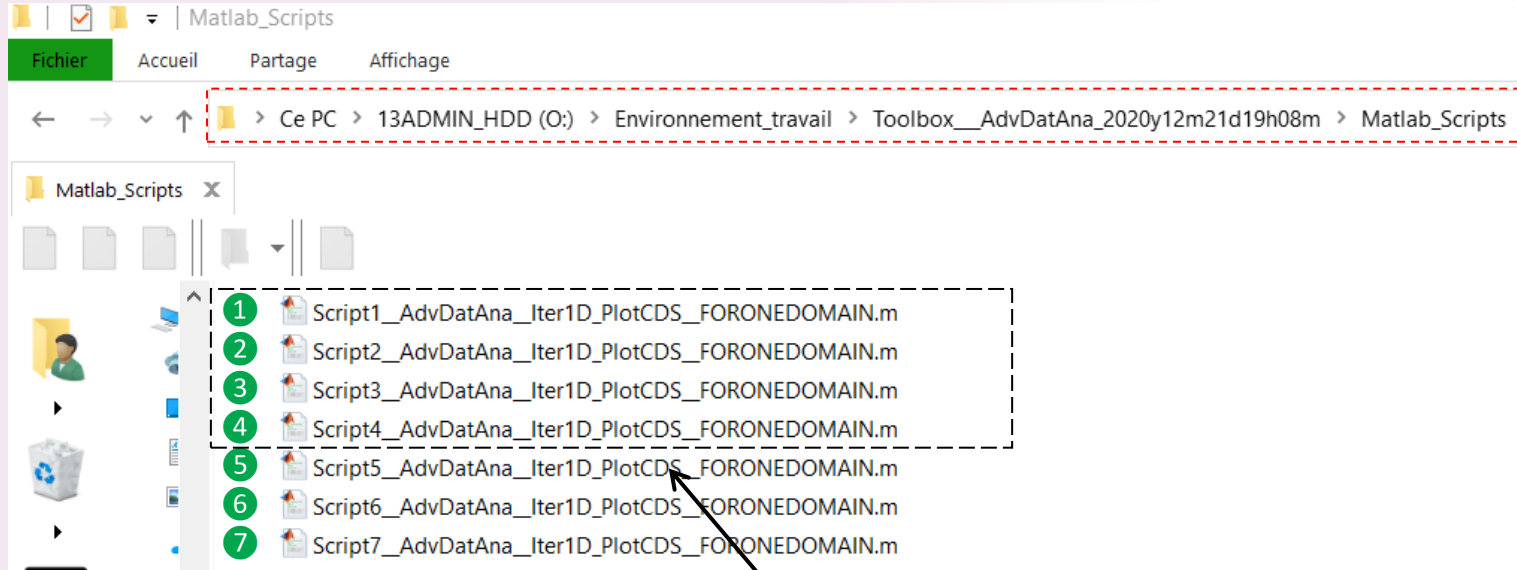
Toolbox__AdvDatAna_2020y12m21d19h08m.rar en l'extrayant directement dans votre dossier de travail (dans notre cas **o:\Environnement_travail**). Accédez enfin au dossier **Matlab_Scripts**. Dans la suite, nous considérons que le dossier **Matlab_Scripts** est directement contenu dans le dossier **o:\Environnement_travail\Toolbox__AdvDatAna_2020y12m21d19h08m**.

B2



N.B. : Il est capital d'utiliser le logiciel **7zip** (bouton Extraire) pour toutes opérations de dézippage.

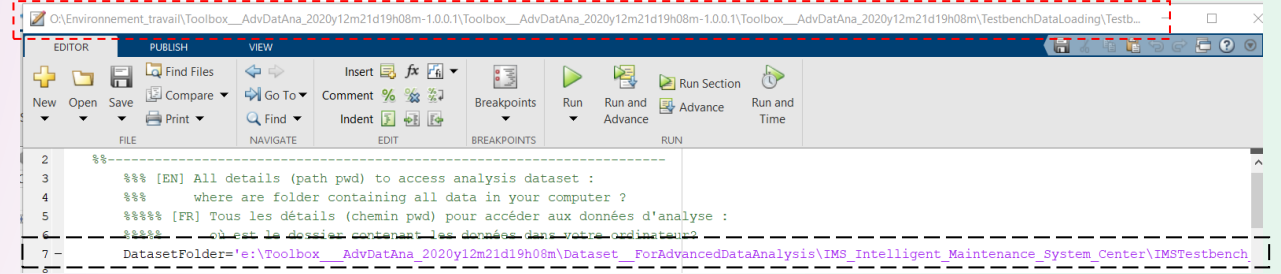
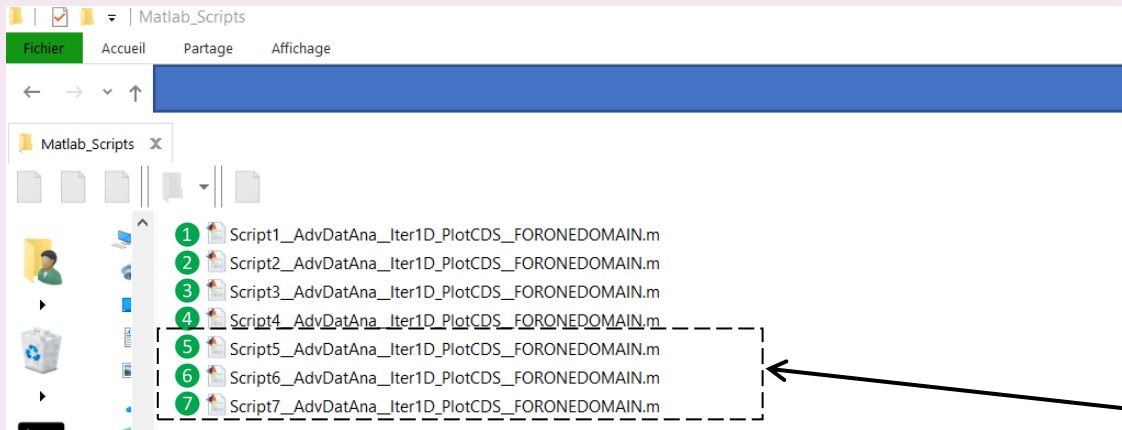
Fig. : Synoptique général d'installation de AdvDatAna (1/2).



Testez en 1^{er} les scripts ① à ④. Ouvrez alors les fichiers scripts .m concernés avec Matlab (**double-clic** ou **clic-droit Open**). Changez la ligne 38 telle que :
 AllToolboxesPWD='e:\Toolbox__AdvDatAna_2020y12m21d19h08m\';
soit réécrite et devienne :
 AllToolboxesPWD='o:\Environnement_travail\Toolbox__AdvDatAna_2020y12m21d19h08m\';

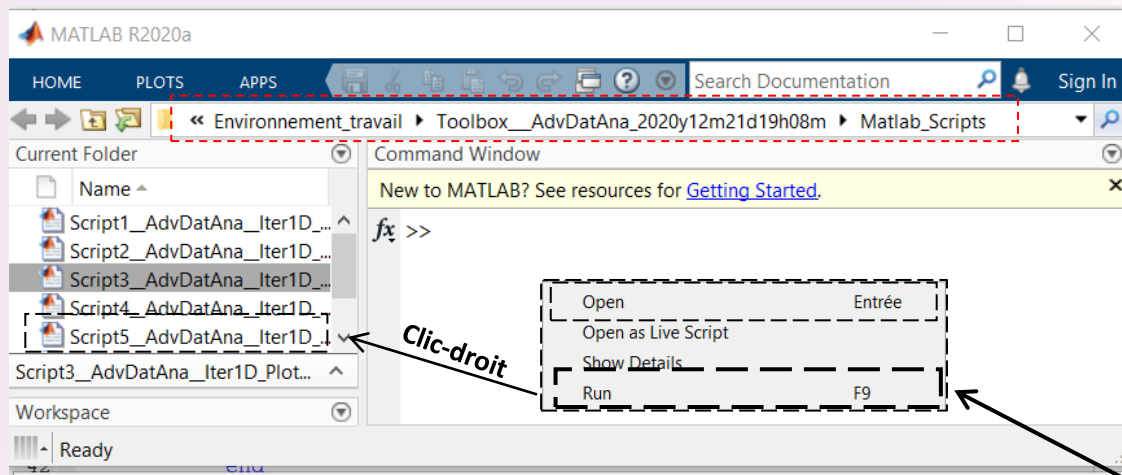
N.B. : Les étapes **B1** à **B3** + **B5** sont nécessaires pour exécuter les scripts Matlab ① à ④.

Fig. : Synoptique général d'installation de AdvDatAna (2/2).



Testez en 2nd les scripts 5 à 7. Ouvrez alors au préalable le fichier script **TestbenchDataLoading__IMS_Intelligent_Maintenance_System_Center.m** (situé dans le dossier **TestbenchDataLoading**) avec Matlab et changez la ligne 7 de sorte que :

DatasetFolder='e:\Toolbox__AdvDatAna_2020y12m21d19h08m\Dataset__ForAdvancedDataAnalysis\IMS_Intelligent_Maintenance_System_Center\IMSTestbench__DatasetForTest';
soit réécrite et devienne :
DatasetFolder='o:\Environnement_travail\Toolbox__AdvDatAna_2020y12m21d19h08m\Dataset__ForAdvancedDataAnalysis\IMS_Intelligent_Maintenance_System_Center\IMSTestbench__DatasetForTest';



Depuis la console Matlab, cliquez-droit devant le script à tester, suivi d'un clic-gauche sur le bouton Run ou d'un clic sur la touche F9 (clavier d'ordinateur).

N.B. : Les étapes **B1** à **B5** sont nécessaires pour exécuter les scripts Matlab 5 à 7.

Fig. : Synoptique général d'installation de AdvDatAna (2/2).

D1 Script1__AdvDatAna__Iter1D_PlotCDS__FORONEDOMAIN.m :

Chargement de 4 signaux Matlab quelconques et affichage des représentations temporelles (option **CalcOnTimeOnTime**, à la **ligne 152**) sur 2 graphiques 2D (distincts) et 2 graphiques 3D (superposés), aux échelles personnalisée (échelle propre ou normalisation à 1) ou identique.

D2 Script2__AdvDatAna__Iter1D_PlotCDS__FORONEDOMAIN.m :

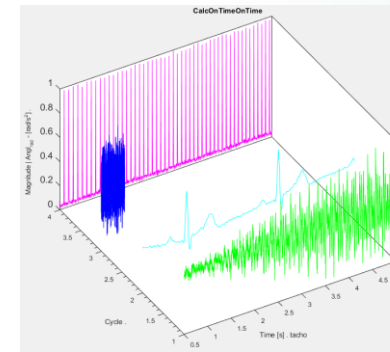
Même résultat que dans **D1**. Les changements principaux sont aux **lignes 150 et 157**, où nous affichons cette fois-ci le spectre d'enveloppe carrée en linéaire (option **CalcOnTimeOnSpectrumASes + DomainLogarithmLin**).

D3 Script3__AdvDatAna__Iter1D_PlotCDS__FORONEDOMAIN.m

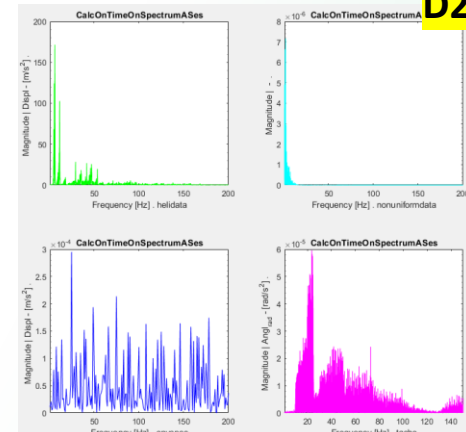
Génération de 3 signaux simulés de cinématique connue et différente (Fe, Fc) et affichage (2D et 3D) du spectre en dB grâce aux options **CalcOnTimeOnSpectrumA + DomainLogarithmLog** (**lignes 151 et 159**). Nous pouvons activer le suivi graphique d'un nombre d'harmoniques de fréquences désirées avec des marqueurs (**ligne 136**).

D4 Script4__AdvDatAna__Iter1D_PlotCDS__FORONEDOMAIN.m :

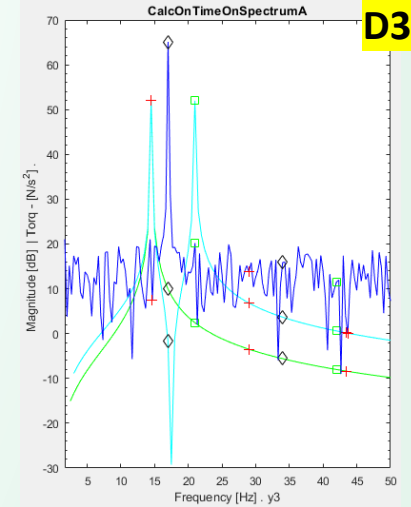
Génération aléatoire de 8 signaux simulés et affichage (2D et 3D) du spectre en linéaire grâce aux options **CalcOnTimeOnSpectrumA + DomainLogarithmLog** (**lignes 133 et 140**).



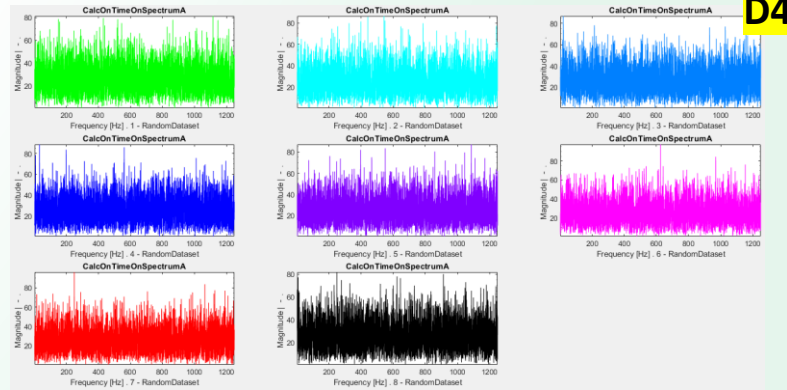
D1



D2



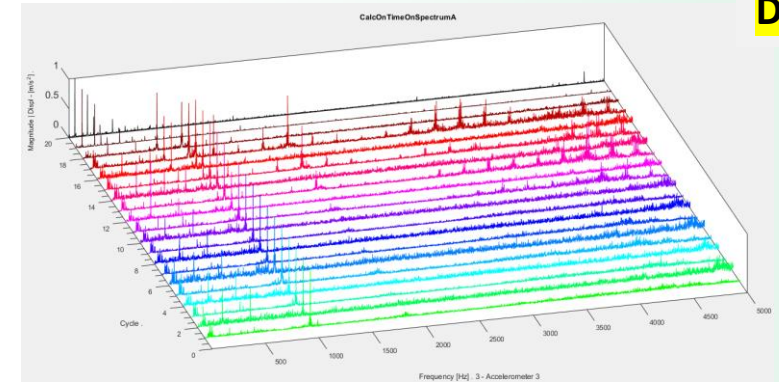
D3



D4

D5 Script5__AdvDatAna__Iter1D_PlotCDS__FORONEDOMAIN.m :

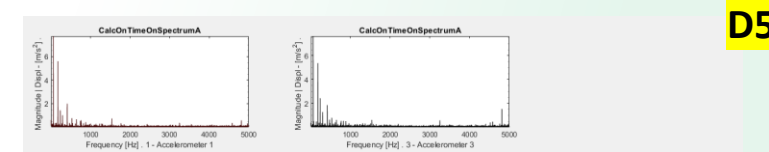
Chargement de 10 fichiers textes (extension .txt) issus du banc de test de roulements de la NASA IMS Center [LEE et al. 2007]. Toutes les options de chargements sont détaillées dans fichier script **TestbenchDataLoading__IMS_Intelligent_Maintenance_System_Center.m** (situé dans le dossier **TestbenchDataLoading**). Les 10 fichiers sont stockés dans **o:\Environnement_travail\Toolbox__AdvDatAna_2020y12m21d19h08m\Dataset__ForAdvancedDataAnalysis\IMS_Intelligent_Maintenance_System_Center\IMSTestbench__DatasetForTest**. Nous avons entre autres un affichage 3D des 10 signaux chargés (option **CalcOnTimeOnSpectrumA + DomainLogarithmLog**). Les graphiques restent affichés en fin de procédure. Ils sont enregistrés dans le dossier indiqué à la **ligne 11** du fichier **TestbenchDataLoading__IMS_Intelligent_Maintenance_System_Center.m**.



D5

D6 Script6__AdvDatAna__Iter1D_PlotCDS__FORONEDOMAIN.m :

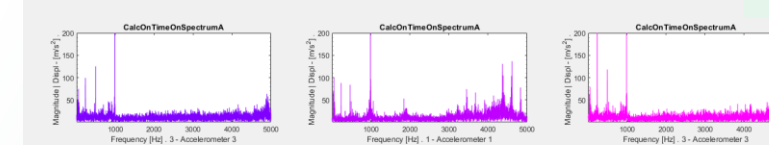
Même résultat que dans **D5**. L'enregistrement des graphiques est en mode invisible.



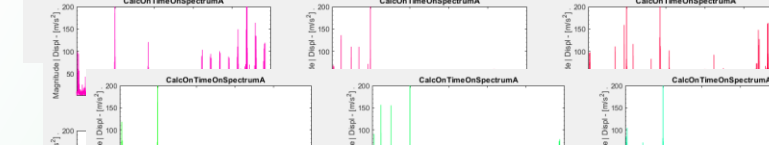
D5

D7 Script7__AdvDatAna__Iter1D_PlotCDS__FORONEDOMAIN.m :

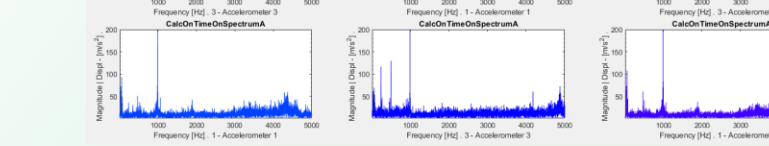
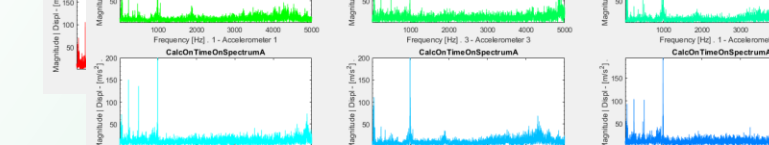
Même résultat que dans **D5**. L'enregistrement des graphiques est en mode semi-visible.



D5



D5



Data	C'est la variable qui stockera l'ensemble de vos données, de type matrice ou Cell.
DataType	Il est possible d'initialiser une matrice avec ses données ou un variable de type Cell. Alors que la matrice est figée et implique que tous les signaux aient la même longueur, l'élément Cell peut prendre des signaux de taille variable.
DataMarker	Il est possible de suivre certaines valeurs notamment dans le domaine fréquentiel. Un choix du marqueur, de sa couleur et du nombre d'harmoniques recherchés suffisent.
SamplingFrequency	C'est ici que la fréquence d'échantillonnage est mentionné. Il est impératif de la convertir en Hz.
AnalysisDomain	Les analyses de signaux peuvent se faire en temporel, fréquentiel, cepstrale ou même dans le domaine de l'analyse d'enveloppe : CalcOnTimeOnTime , CalcOffAngularOnCepstrumR , ...
DomainLogarithm	L'échelle devra toujours rester en linéaire excepté lorsque l'on cherche à travailler en dB, surtout dans les représentations fréquentielles.
XAxisLimits, YAxisLimits, ZAxisLimits	En modifiant l'un de ces vecteurs, nous pouvons réaliser des zooms et améliorer de ce fait la résolution des images.
PlotBox, PlotXGrid, PlotYGrid, PlotYMinorTick	Les graphiques peuvent être encadrés. Il est également intéressant d'activer la graduation selon l'axe des abscisses ou ordonnées.
PlotZoom	Tous les graphiques sont mis à la même échelle avec l'option Global. Local est dédié aux zooms spécifiques à chaque graphique. Le WaterfallRescale correspond au 3D. Super permet d'avoir toutes les figures sur un même graphique. Assembling est une autre option à découvrir.
SubplotValues	Il peut être parfois nécessaire d'afficher tous ses graphiques sur 1 ligne à 3 colonne [1 3] ou 1 colonne à 4 lignes [4 1]. Vous avez 5 signaux et pouvez aussi les afficher sur 2 lignes et 3 colonnes [2 3].

LegendLanguage	Les légendes peuvent être mis en Français ou en Anglais.
LegendNumbering	Les graphiques peuvent être numérotés dans l'ordre croissant (1 ^{er} chargé, 1 ^{er} numéroté).
DataRecordType	Si non activé, aucun enregistrement de données n'est possible.
DataRecordFormat	Les données peuvent être enregistrés suivant différents formats de fichiers (Mat, Txt, Csv, ...).
LegendNumbering	Les graphiques peuvent être numérotés dans l'ordre croissant (1 ^{er} chargé, 1 ^{er} numéroté).
PlotRecordType	Si non activé, aucun enregistrement de graphiques n'est possible.
PlotRecordFormat	Les données peuvent être enregistrés suivant différent format de graphiques. (Png, Bmp, Jpg, ...).
VisiblePlot	Il est possible d'afficher les graphiques ou de les cacher. Les cacher lorsque l'enregistrement automatisé est activé est pratique. Vous pouvez disposer librement de votre machine pour faire autre chose (difficile lors d'un affichage intempestif de graphiques).
AdvDatAna__Create1D_PlotCDS	Vous pouvez sélectionner cette variable et le raccourci CTRL + D. Vous accéderez à l'ensemble des options exhaustives par défaut activées par AdvDatAna, avant modifications de l'usagé.
AdvDatAna__Iter1D_PlotCDS__2020y12m21d19h08m	Vous pouvez sélectionner cette variable et le raccourci CTRL + D. Vous accéderez à l'ensemble des options exhaustives non présentés ici. Vous pourrez découvrir tous les domaines d'analyse (exemple)

AdvDatAna a été développée avec un ensemble de garde-fou. Si vous faites une erreur dans la saisie d'une option, un message d'erreur essaiera de vous proposer les meilleures alternatives.

```

57 fs1 = 400; % [EN] Sampling frequency / [FR] Fréquence d'échantillonnage
58 length_t1=2; % [EN] time duration / [FR] Durée temporelle
59 t1 = 0:1/fs1:length_t1; % [EN] Time vector / [FR] Vecteur temporel
60 f1 = 14.5; % [EN] Tracking frequency / [FR] Fréquence à surveiller
61 y1 = sin(2*pi*f1*t1); % [EN] Generated signal 1/ [FR] Signal généré 1
62
63 fs2 = 400;
64 length_t2=2;
65 t2 = 0:1/fs2:length_t2;
66 f2 = 21;
67 y2 = y1+ sin(2*pi*f2*t2);
68
69 fs3 = 1200;
70 length_t3=3;
71 t3 = 0:1/fs3:length_t3;
72 f3 = 17;
73 y3 = sin(2*pi*f3*t3) + randn(size(t3))/12;
74
75 DatasetFolderForResultsSaving=['e:\SinusDataset'];
76 LoadedClassicalData={y1,y2,y3};
77 Number_Dataset=size(LoadedClassicalData,1);
78 LoadedSamplingFrequency={fs1,fs2,fs3};
79
80 Main_TextSensorsNames={'y1','y2','y3'};
81 Main_TextSensorsType__Temp='SensorTypeTorque';
82 Main_TextSensorsType= repmat({Main_TextSensorsType__Temp},1,Number_Dataset);
    
```

Fig. : Zone de création des vecteurs "signaux" (signaux à étudier, leur fréquence d'échantillonnage et les noms donnés à chacun), argument d'entrée de AdvDatAna.

```

85 %%% [EN] Toolbox AdvDatAna integrates many options.
86 %%% You can discover it when you will read below function
87 %%% [FR] La boîte à outils AdvDatAna possède de nombreuses options.
88 %%% Consulter la fonction suivante pour les découvrir
89 AdvDatAna_ListID_PlotCDS();
90 %
91 %-----
92 %%% [EN] Initialization of all default parameters
93 %%% Below parameters come to modify these default parameters
94 %%% You can see below function to discover default configuration
95 %%% [FR] Initialisation de tous les paramètres par défaut.
96 %%% Les initialisations qui suivent viennent modifier
97 %%% ces paramètres par défaut. Lisez la fonction suivante
98 %%% pour en savoir plus
99 [PlotDisplay_AdvDatAna_IterID_PlotCDS,PlotRecord_AdvDatAna_IterID_PlotCDS,
100 AdvDatAna_CreateID_PlotCDS];
101 %-----
102 %
103 %%% [EN] We cannot explain all parameters in details. Different definitions
104 %%% are similar in Matlab.
105 %%% [FR] Nous ne pourrons pas tout vous expliquer car les noms
106 %%% des paramètres sont très clairs. Il ressemble à ceux de Matlab.
107 [PlotDisplay_AdvDatAna_IterID_PlotCDS,PlotRecord_AdvDatAna_IterID_PlotCDS,
108 AdvDatAna_CreateID_PlotCDS];
109
110 %*****
111 %%% [EN] We cannot explain all parameters in details. Different definitions
112 %%% are similar in Matlab.
    
```

Fig. : Des paramètres de préparation de l'analyse à venir sont initialisés ici, entre autres actions.

```

124 PlotDisplay_AdvDatAna_IterID_PlotCDS.DataCDS.DataType='DataTypeCellID';
125 %PlotDisplay_AdvDatAna_IterID_PlotCDS.DataCDS.DataType='DataTypeMatrixID';
126 %*****
127 %%% [EN] Variable LoadedClassicalData containing imported data comes from
128 %%% function TestbenchDataLoading_IMS_Intelligent_Maintenance_System_Center
129 %%% [FR] La variable LoadedClassicalData contenant les données a été importée grâce
130 %%% à la fonction TestbenchDataLoading_IMS_Intelligent_Maintenance_System_Center
131 PlotDisplay_AdvDatAna_IterID_PlotCDS.DataCDS.Data=LoadedClassicalData;
132 %*****
133 PlotDisplay_AdvDatAna_IterID_PlotCDS.DataCDS.DataMarker='DataMarkerOn';
134 % PlotDisplay_AdvDatAna_IterID_PlotCDS.DataCDS.DataMarker='DataMarkerOff';
135 %*****
136 PlotDisplay_AdvDatAna_IterID_PlotCDS.DataCDS.DataMarkerValues=...
137 {{f1,8,'DataMarkerValues','DataMarkerColorsRed'},...
138 {f2,10,'DataMarkerValues','DataMarkerColorsGreen'},...
139 {f3,5,'DataMarkerValues','DataMarkerColorsBlack'}};
140 %*****
141 %%% [EN] Variable LoadedSamplingFrequency containing imported sampling frequency comes from
142 %%% function TestbenchDataLoading_IMS_Intelligent_Maintenance_System_Center
143 %%% [FR] La variable LoadedSamplingFrequency contenant les fréquences
144 %%% d'échantillonnages a été importée grâce à la fonction TestbenchDataLoading_IMS_Inte
145 PlotDisplay_AdvDatAna_IterID_PlotCDS.DataCDS.SamplingFrequency=LoadedSamplingFrequency;
146 %*****
147 PlotDisplay_AdvDatAna_IterID_PlotCDS.DataCDS.DataThreshold='DataThresholdCumvar';
148 % PlotDisplay_AdvDatAna_IterID_PlotCDS.DataCDS.DataThreshold='DataThresholdOff';
149 % PlotDisplay_AdvDatAna_IterID_PlotCDS.DataCDS.DataThreshold='DataThresholdMovingmed';
150 %*****
    
```

Fig. : Une des zone de modification des paramètres par défaut.

Paramètres de la zone d'étude :

- Pays : Gabon ;
- Parc national : La Lopé;
- Superficie : 4 910 km²;
- Sous-zones :
 - Forêt : (coordonnées 11.605E, -0.206N);
 - Savane : (coordonnées GPS : 11.605E, -0.201N);

[TUFIN et al. 2019, BUSH et al. 2017, BUSH et al. 2018, BUSH 2018]

Données météorologiques :

- Période d'acquisition : 1984 à 2018;
- Type de mesures :
 - Humidité (Humidity) - 1 fichier csv de 534 Ko;
 - Ensoleillement (Solar) - 1 fichier csv de 64 Ko;
 - Précipitations (Rainfall) - 1 fichier csv de 619 Ko;
 - **Température (Temperature)** - 1 fichier csv de 1,596 Mo;
 - Vent (Wind) - 1 fichier csv de 72 Ko.



Fig. : Le Gabon et ses 13 parcs nationaux (source internet).

Informations liées aux données de températures :

- Appareil de mesure :
 - Savane : Thermomètre numérique Max/Min (Taylor 1441)
 - Forêt : Thermomètre manuel Max/Min placé à 1,5 m au-dessus du sol, sous une canopée fermée;
- Fréquence de mesure : A intervalle quotidien (régulier);
- Nom du fichier : **Temperature_daily_v2018-03-13.csv**;
- Unité : ° C.

Champs	Valeur
Année	1984, 1985, 1986, ... ,2016, 2017, 2018
Type	Min [12744 lignes], Max [12794 lignes]
Method	HOBO [3724 lignes], MaxMin Thermometer (digital) [1177 lignes], MaxMin Thermometer (manual) [6103 lignes] , SKYE [1350 lignes], TinyTag [11808 lignes], VantagePro [1376 lignes]
Site	Saline [13 941 lignes] , SEGC [11597 lignes]

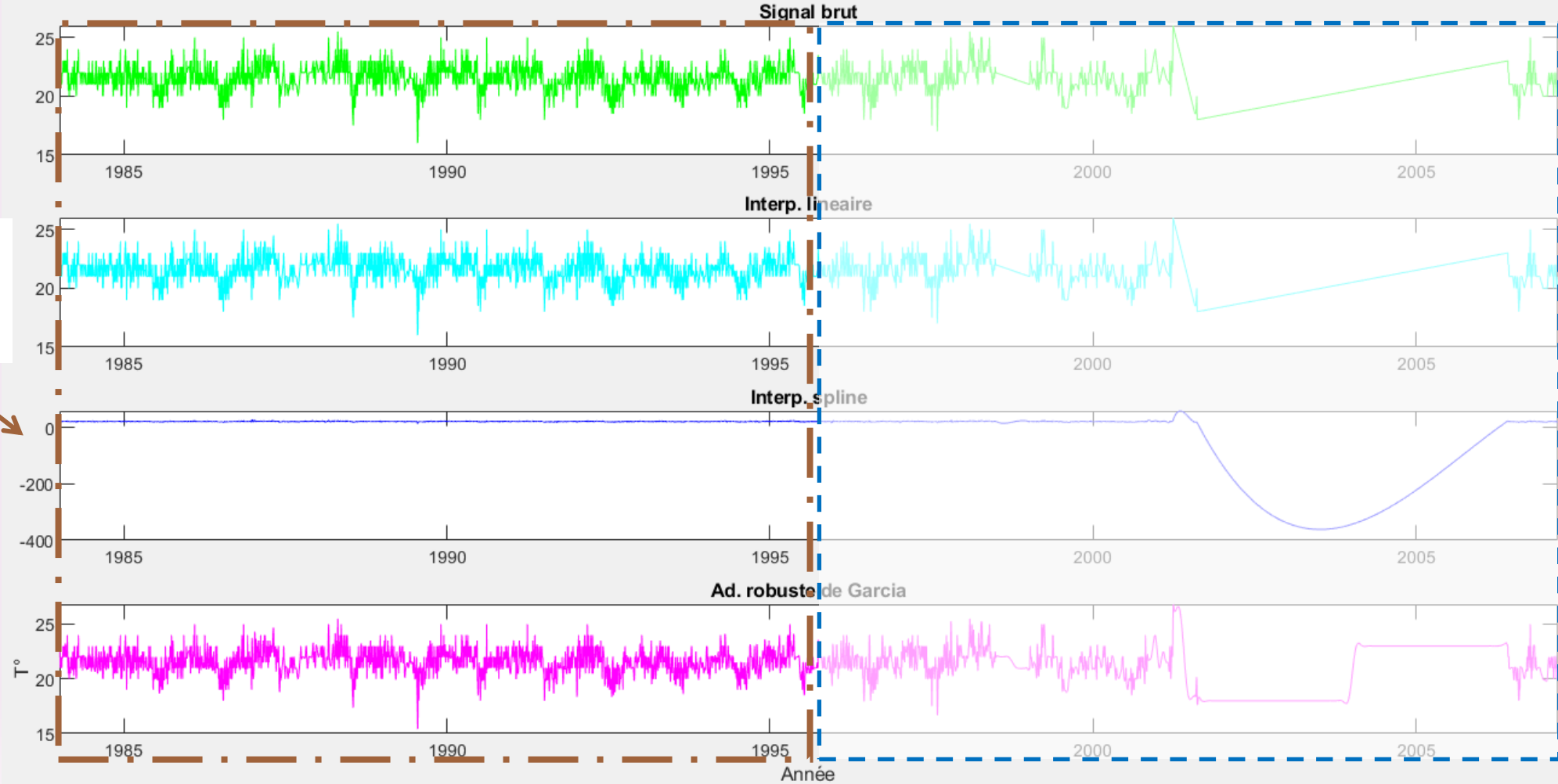
Fig. : Champs (colonnes) et valeurs.

	X	Date	Temperature_c	Type	Method	Site	Comments
1	1	03/01/1984	23	Min	MaxMin Thermometer (manual)	Saline	
2	2	05/01/1984	23	Min	MaxMin Thermometer (manual)	Saline	
3	3	06/01/1984	22	Min	MaxMin Thermometer (manual)	Saline	
4	4	08/01/1984	21	Min	MaxMin Thermometer (manual)	Saline	
5	5	09/01/1984	22	Min	MaxMin Thermometer (manual)	Saline	
6	6	11/01/1984	21	Min	MaxMin Thermometer (manual)	Saline	
7	7	12/01/1984	21	Min	MaxMin Thermometer (manual)	Saline	
8	8	13/01/1984	22	Min	MaxMin Thermometer (manual)	Saline	
9	9	15/01/1984	21	Min	MaxMin Thermometer (manual)	Saline	
10	10	17/01/1984	22	Min	MaxMin Thermometer (manual)	Saline	

Fig. : Vue zoomée du fichier **Temperature_daily_v2018-03-13.csv** composé originellement de 25538 lignes (jours) et 8 colonnes (informations diverses).

Conditions de travail : Données (**Brutes**), Année (**1984-1996**), Adjacence (**1x13**), Type (**Min**), Method (**MaxMin Thermometer manual**), Site (**Saline**)

La Région B est peu fiable à cause de la présence importante de données manquantes, difficiles à corriger par algorithme.



Région A
(peu
bruitée)

Région B
(très
bruitée)

Fig. : Représentations temporelles des températures minimales globales dans la période de 1982 à 2007, signal brut et ses versions reconstruites (signal brut interpolé linéairement, avec des splines ou adoucis avec l'algorithme de GARCIA [WANG et al. 2012, GARCIA 2010 et 2023]).

Conditions de travail : Données (**Brutes**), Année (**1984-1996**), Adjacence (**1x13**), Type (**Min**), Method (**MaxMin Thermometer manual**), Site (**Saline**)

La Région A, plus fiable, montre que la correction des données manquantes peu avoir un un impact sensible sur le signal reconstruit.

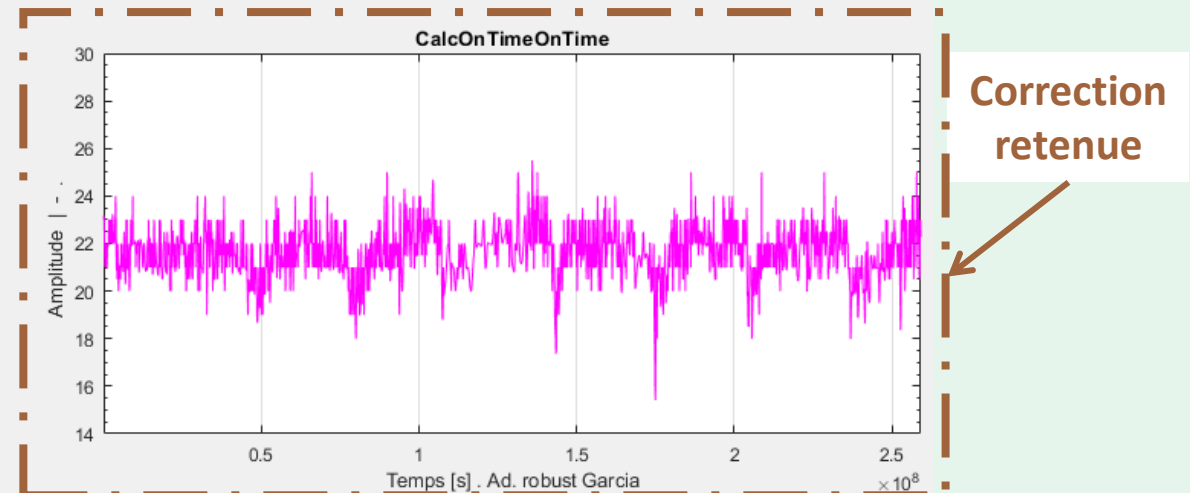
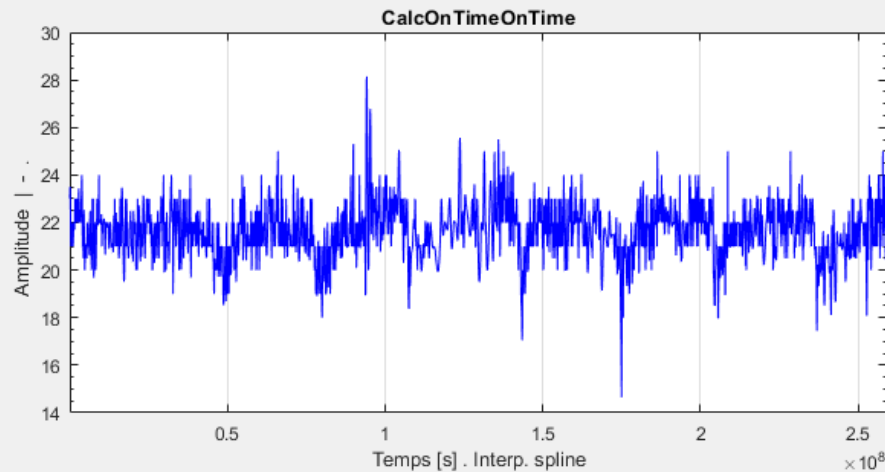
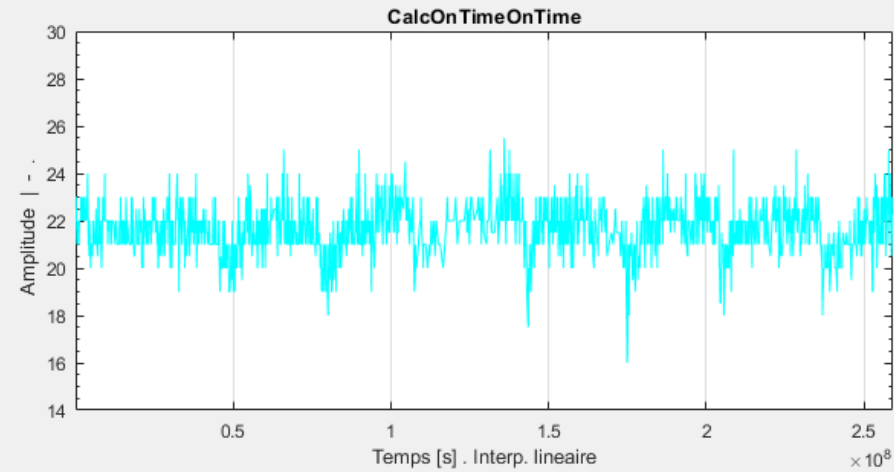
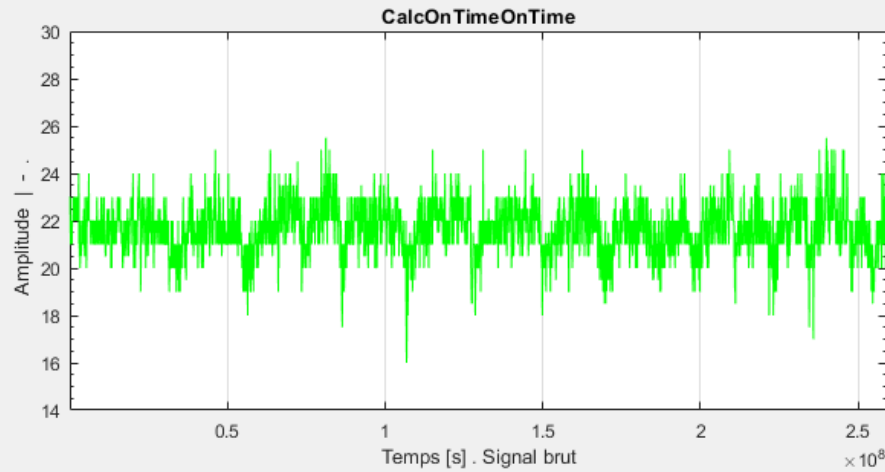
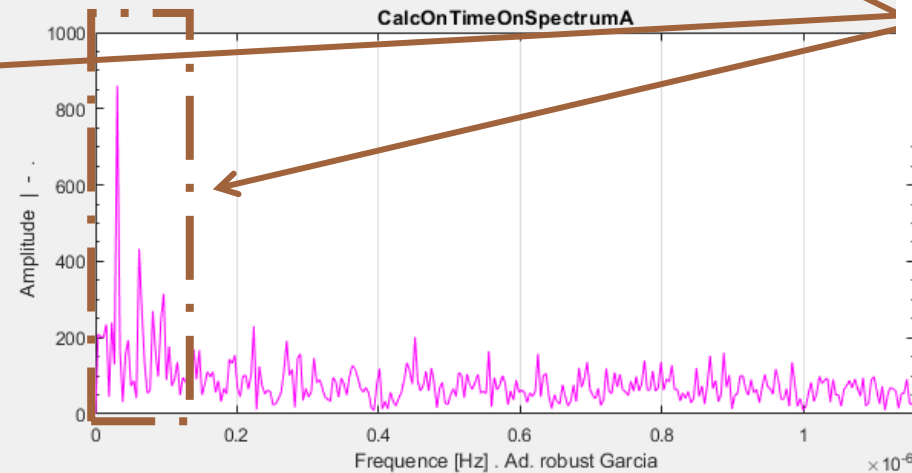
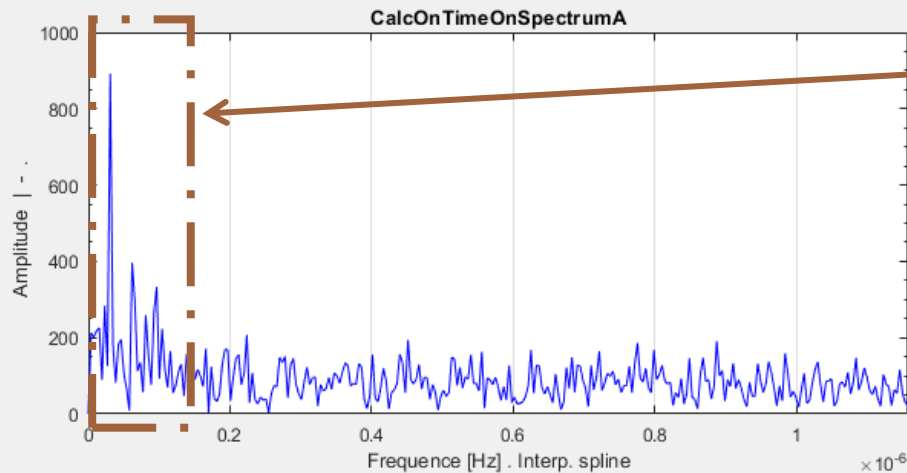
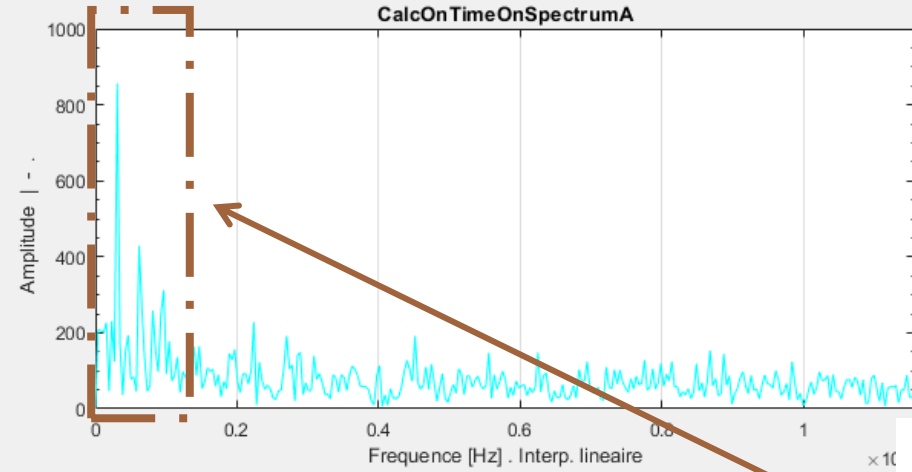
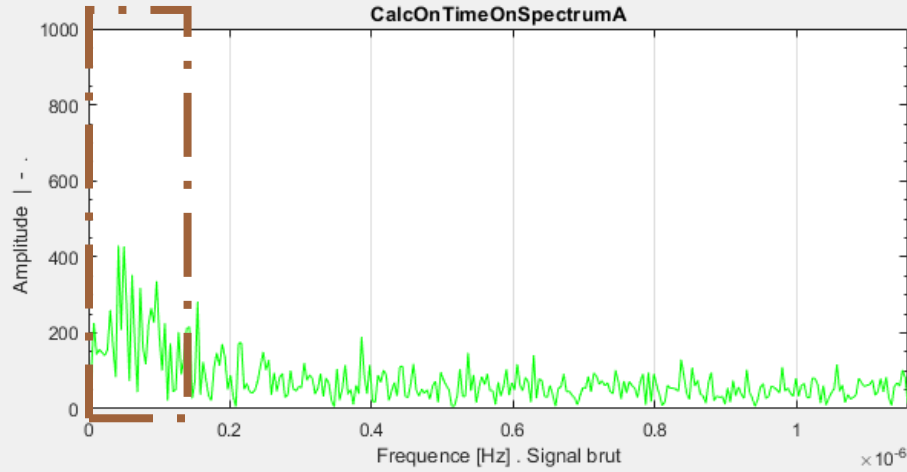


Fig. : Représentations temporelles des températures minimales globales (signal brut, interpolation linéaire, interpolation spline et adoucissement robuste de GARCIA [WANG et al. 2012, GARCIA 2010 et 2023]), au-delà d'une région parsemée de valeurs manquantes.

Conditions de travail : Données (**Brutes**), Année (**1984-1996**), Adjacence (**1x13**), Type (**Min**), Method (**MaxMin Thermometer manual**), Site (**Saline**)

La Région A, plus fiable, montre que la correction des données manquantes peu avoir un un impact sensible sur le signal reconstruit.



Apports de la correction des données manquantes

Fig. : Représentations spectrales des températures minimales globales (signal brut, interpolation linéaire, interpolation spline et adoucissement robuste de GARCIA [WANG et al. 2012, GARCIA 2010 et 2023]), au-delà d'une région parsemée de valeurs manquantes.

Conditions de travail : Données (**Brutes**), Année (**1984-1992**), Adjacence (**1x9**), Type (**Min**), Method (**MaxMin Thermometer manual**), Site (**Saline**)

Conditions de travail correspondant à la série temporelle ayant le plus de données, à savoir 3074 lignes (jours) de 1984 à 2007.

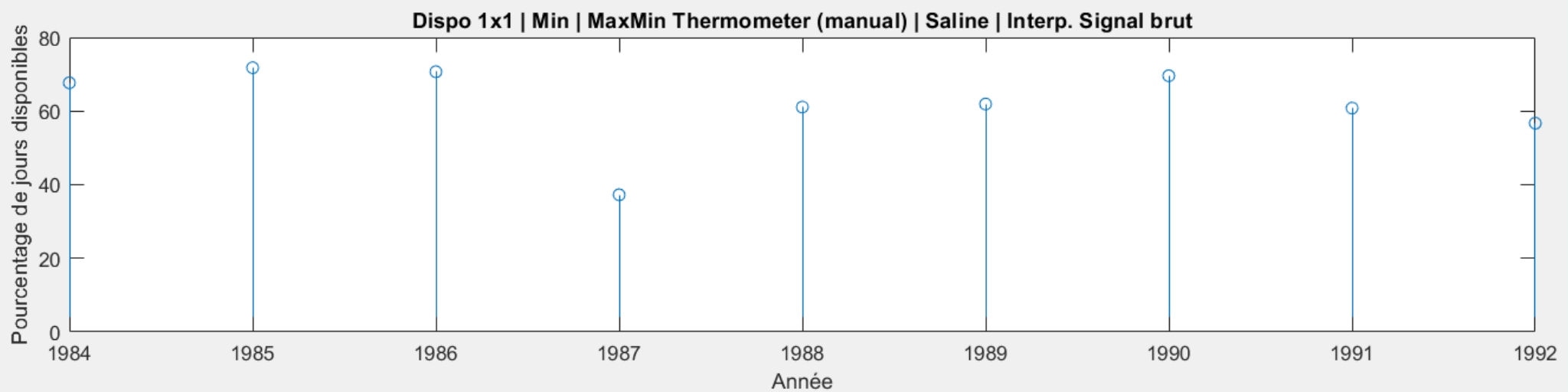
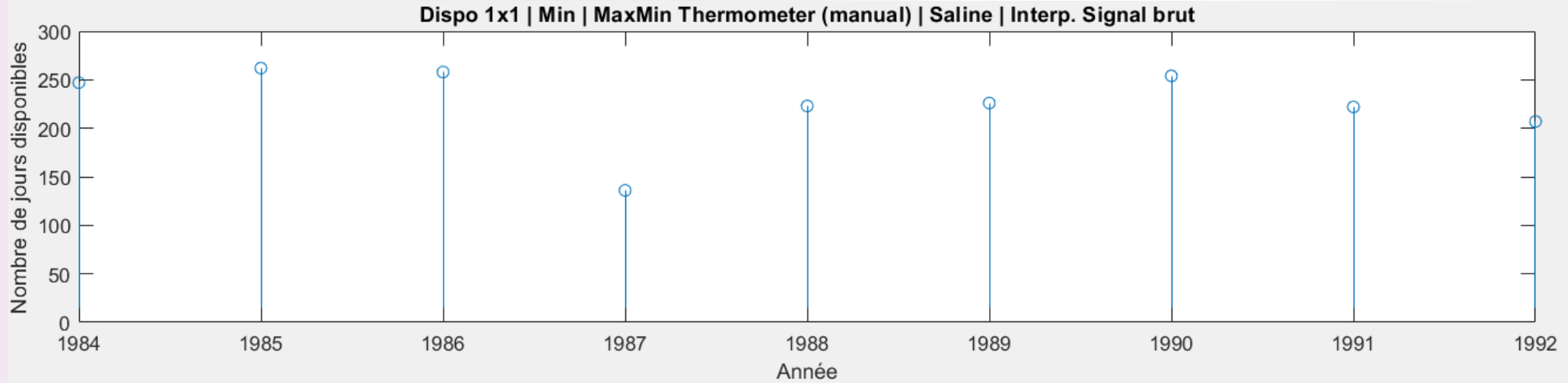


Fig. : Disponibilité des températures minimales en fonction de l'année (en jour en haut, en pourcentage en bas), dans la période de 1984 à 1992.

Conditions de travail : Données (**GARCIA**), Année (**1984-1992**), Adjacence (**1x9**), Type (**Min**), Method (**MaxMin Thermometer manual**), Site (**Saline**)

Conditions de travail correspondant à la série temporelle ayant le plus de données, à savoir 3074 lignes (jours) de 1984 à 2007.

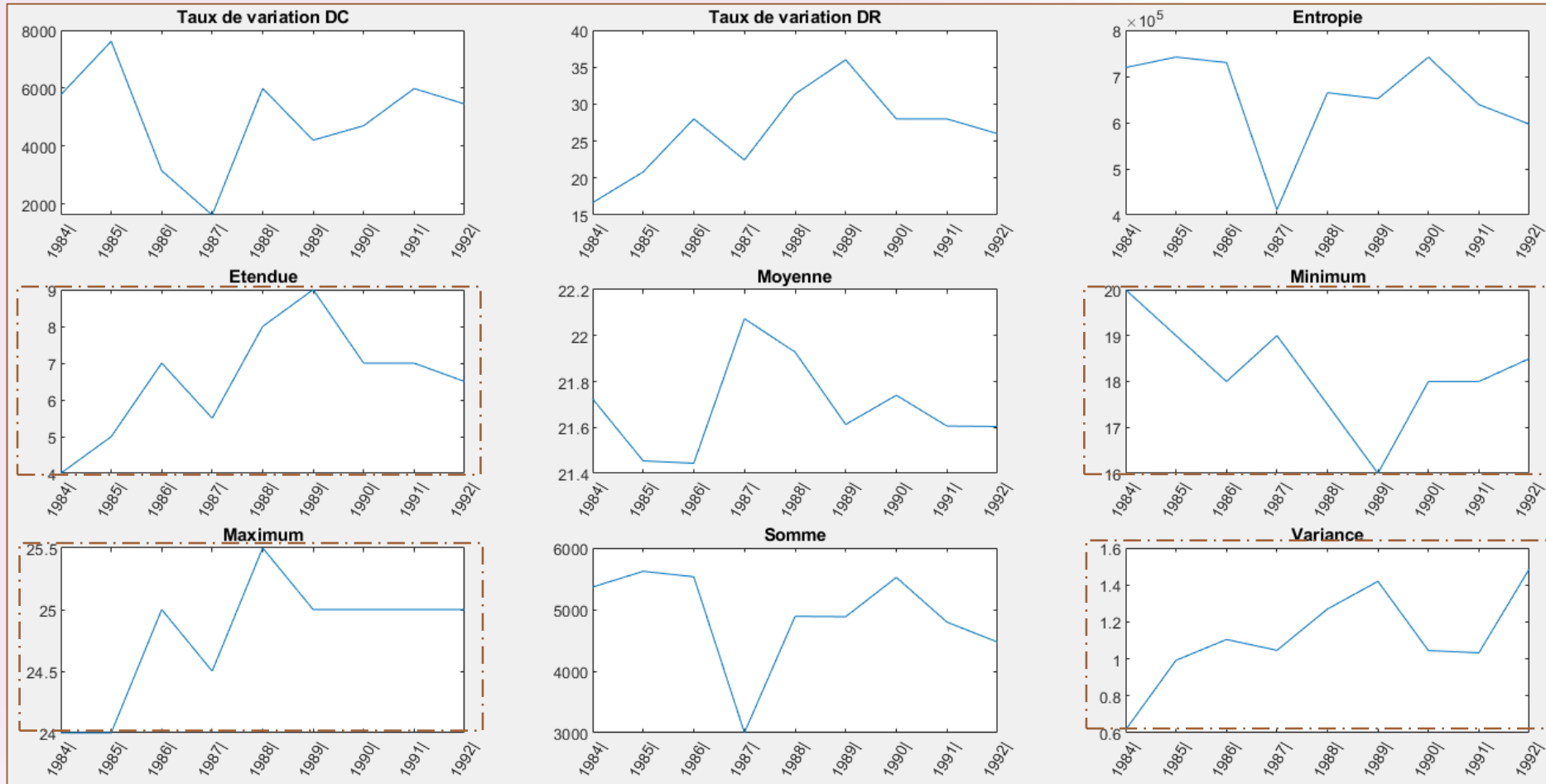


Fig. : Quelques paramètres statistiques calculés sur les températures minimales, dans la période de 1984 à 1992.

1985 et 1988 sont des années très différentes des autres (**Taux de variation DC**).

Les signaux sont plus désordonnés qu'en 1987 (**Entropie**).

L'écart entre le minimum et le maximum semble augmenté (**Etendue**) : étendue de l'étendue = $9 - 4 = 5$.

La **Moyenne** indique de faibles variations de température d'une année à l'autre sauf en 1987.

La valeur de température la plus basse a été obtenue en 1989 (**Minimum**) et la plus élevée en 1988 (**Maximum**).

1987 est l'année ayant eu les températures les plus faibles (**Somme**).

Les écarts de températures dans une même année semble croître (**Variance**).

Conditions de travail : Données (**GARCIA**), Année (**1984-1992**), Adjacence (**1x9**), Type (**Min**), Method (**MaxMin Thermometer manual**), Site (**Saline**)
 Conditions de travail correspondant à la série temporelle ayant le plus de données, à savoir 3074 lignes (jours) de 1984 à 2007.

Fréquence (Hz)	$F_{\text{trimestre}} = 1,286 \cdot 10^{-7}$	$F_{\text{semestre}} = 6,43 \cdot 10^{-8}$	$F_{\text{an}} = 3,215 \cdot 10^{-8}$	$F_{\text{2ans}} = 1,6075 \cdot 10^{-8}$
Harmonique n°1 (Hz)	$2,572 \cdot 10^{-7}$	1,2860e-07	$6,43 \cdot 10^{-8}$	$3,215 \cdot 10^{-8}$
Harmonique n°2 (Hz)	$3,858 \cdot 10^{-7}$	1,9290e-07	$9,6451 \cdot 10^{-8}$	$4,8225 \cdot 10^{-8}$
Période (s)	7 776 000	1 555 200	31 104 000	62 208 000
Symbole	*	□	*	□

$$T_e = 24 \times 60 \times 60 = 86400 \text{ s}$$

$$\Leftrightarrow F_e = 1/T_e = 1,1574 \cdot 10^{-5} \text{ Hz}$$

$$F1_{\text{car}} = 5,119 \cdot 10^{-8} \text{ Hz} \quad (\Leftrightarrow 7,5367 \text{ mois})$$

$$F2_{\text{car}} = \text{Harmonique 2 de } F_{\text{an}} \quad (\Leftrightarrow 4 \text{ mois})$$

$$F3_{\text{car}} = \text{Harmonique 2 de } F_{\text{trimestre}} \quad (\Leftrightarrow 1 \text{ mois})$$

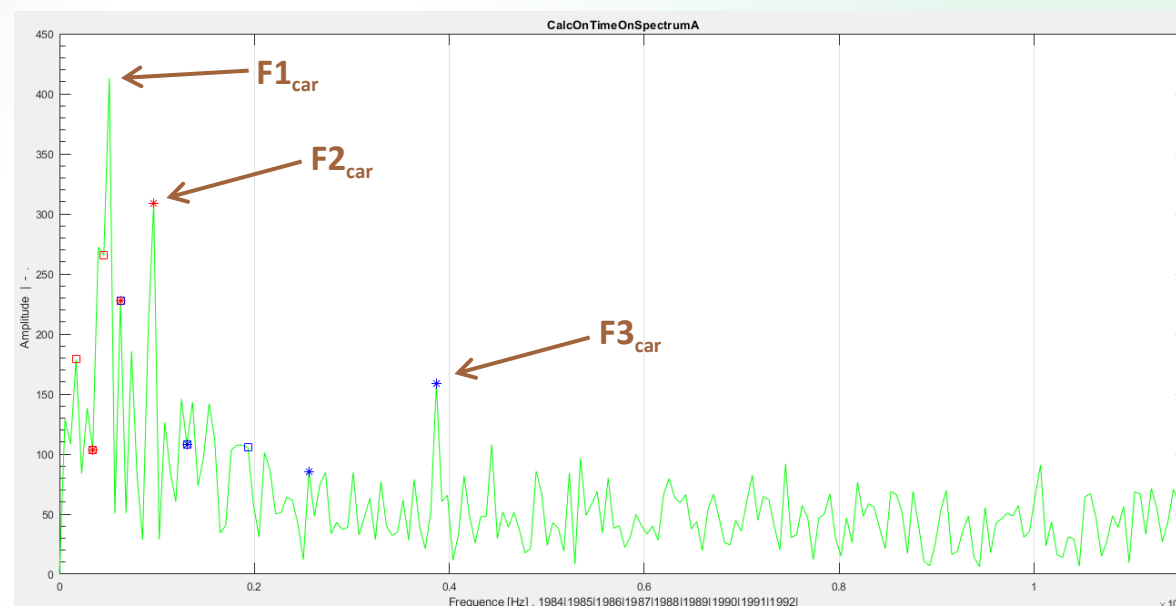
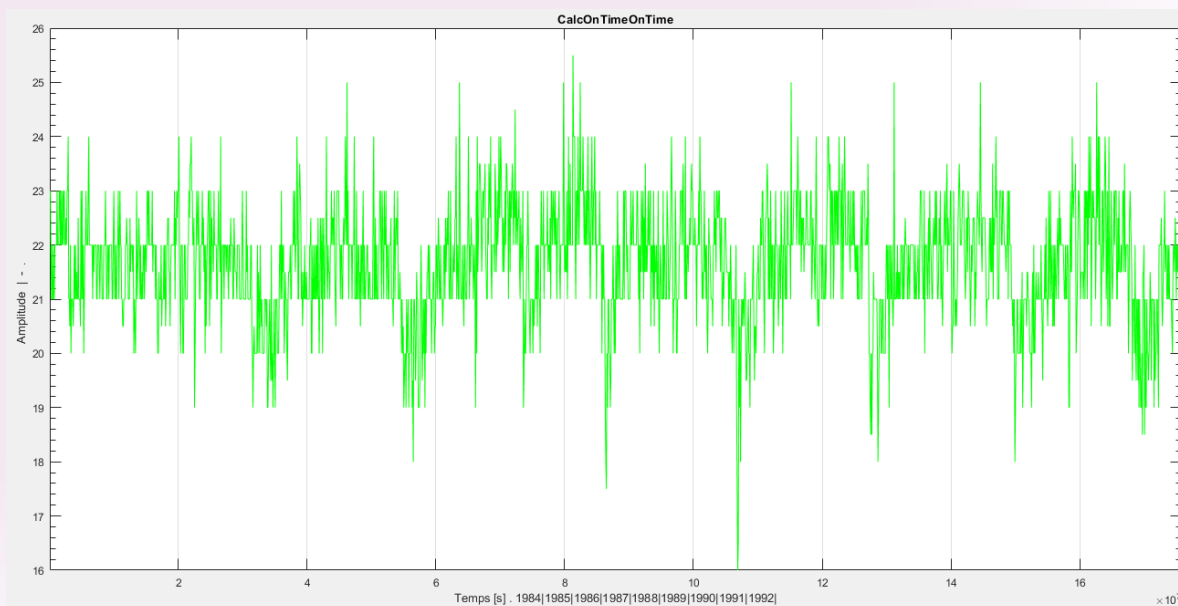
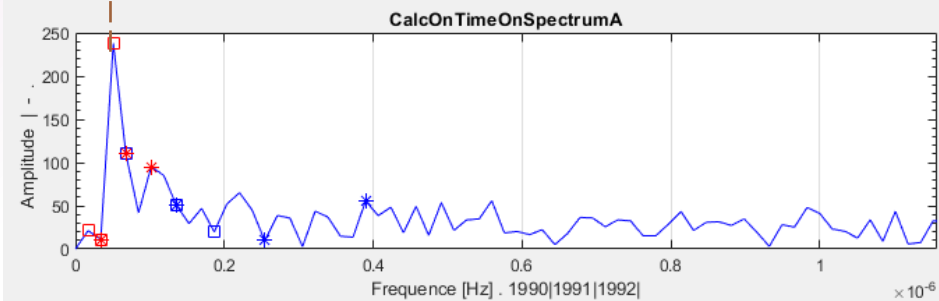
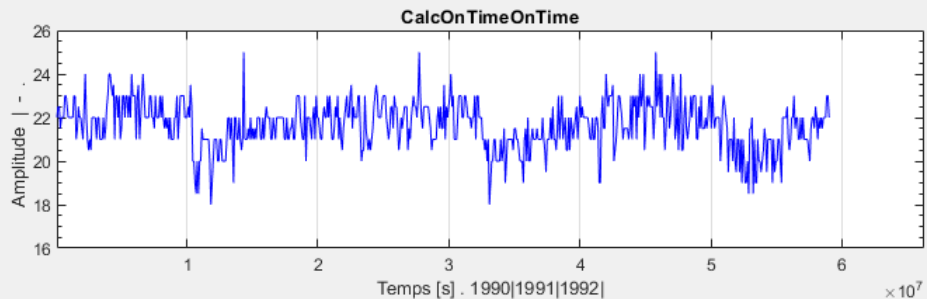
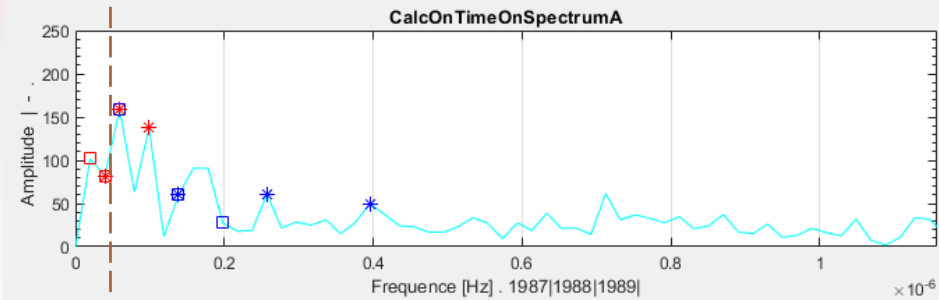
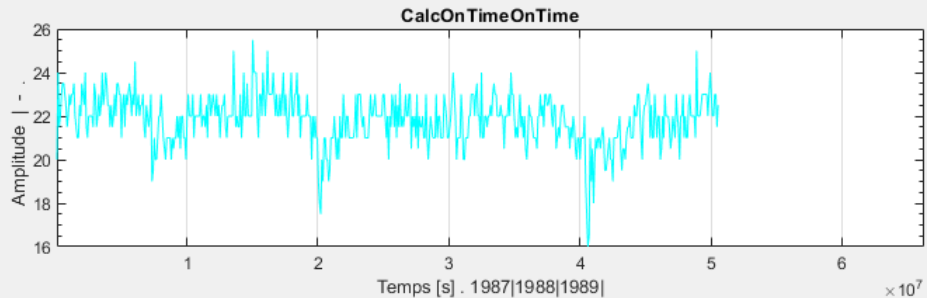
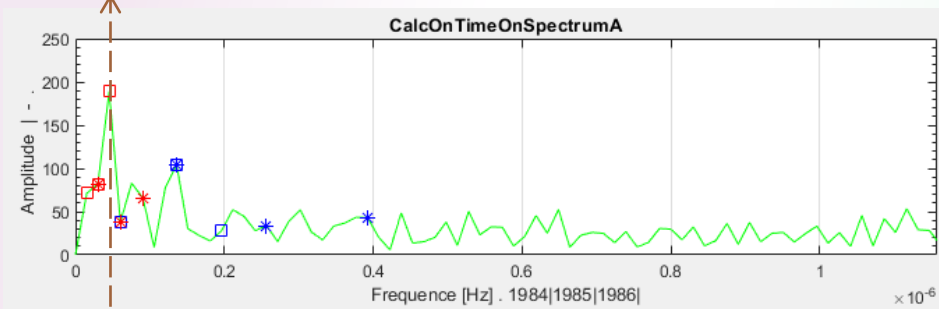
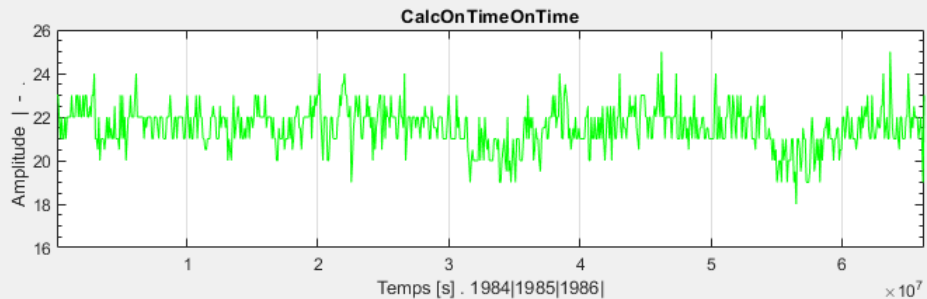


Fig. : Représentations temporelle (gauche) et spectrale (droite) des températures minimales, allant de 1984 à 1992 (9 années \Leftrightarrow 1x9).

Conditions de travail : Données (**GARCIA**), Année (**1984-1992**), Adjacence (**3x3**), Type (**Min**), Method (**MaxMin Thermometer manual**), Site (**Saline**)

Conditions de travail correspondant à la série temporelle ayant le plus de données, à savoir 3074 lignes (jours) de 1984 à 2007.

$$F1_{car} = 4,86 \cdot 10^{-8} \text{ Hz } (\Leftrightarrow 7,9383 \text{ mois})$$

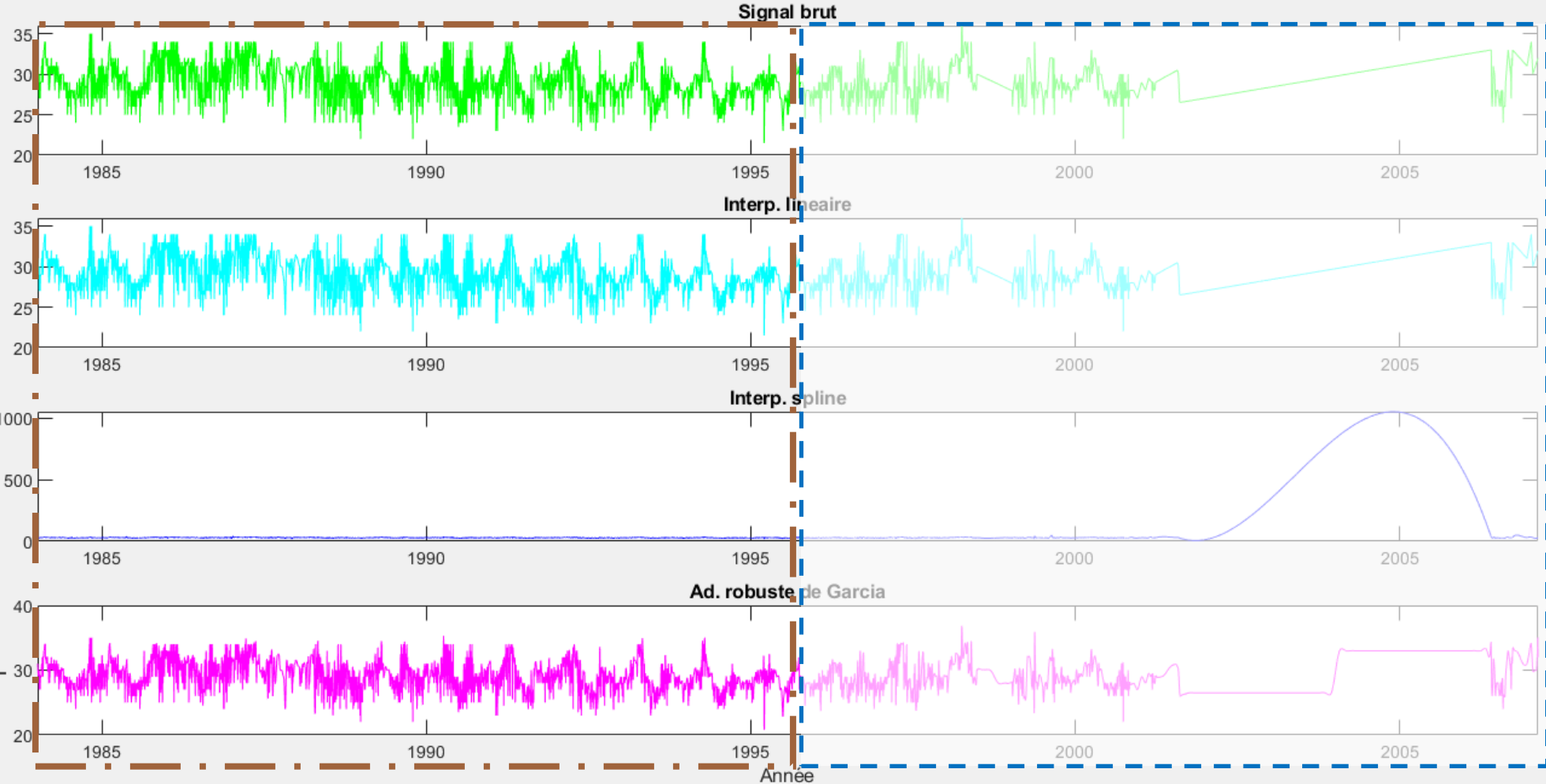


Fréquence (Hz)	Symbole
$F_{trimestre} = 1,286 \cdot 10^{-7}$	*
$F_{semestre} = 6,43 \cdot 10^{-8}$	□
$F_{an} = 3,215 \cdot 10^{-8}$	*
$F_{2ans} = 1,6075 \cdot 10^{-8}$	□

Fig. : Représentations temporelle (gauche) et spectrale (droite) des températures minimales, allant de 1984 à 1992 (3 vecteurs contenant 3 années consécutives $\Leftrightarrow 3x3$).

Conditions de travail : Données (**Brutes**), Année (**1984-1996**), Adjacence (**1x13**) , Type (**Max**), Method (**MaxMin Thermometer manual**), Site (**Saline**)

La Région B est peu fiable à cause de la présence importante de données manquantes, difficiles à corriger par algorithme.



Région A
(peu
bruitée)

Région B
(très
bruitée)

Fig. : Représentations temporelles des températures maximales globales dans la période de 1982 à 2007 : signal brut et ses versions reconstruites (signal brut interpolé linéairement, avec des splines ou adoucis avec l'algorithme de GARCIA [WANG et al. 2012, GARCIA 2010 et 2023]).

Conditions de travail : Données (**Brutes**), Année (**1984-1992**), Adjacence (**1x9**), Type (**Max**), Method (**MaxMin Thermometer manual**), Site (**Saline**)

Conditions de travail correspondant à la série temporelle ayant le plus de données, à savoir 3029 lignes (jours) de 1984 à 2007.

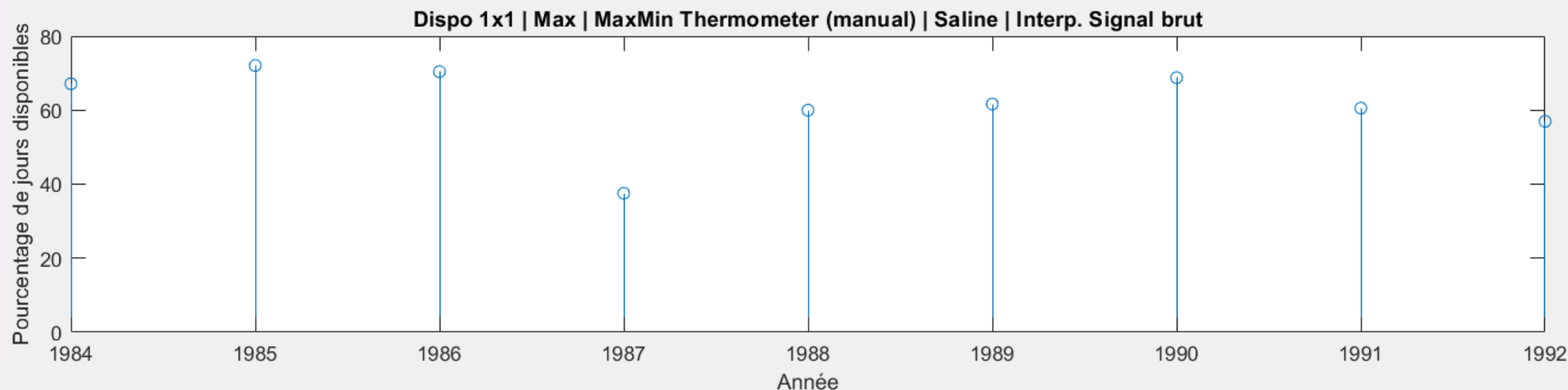
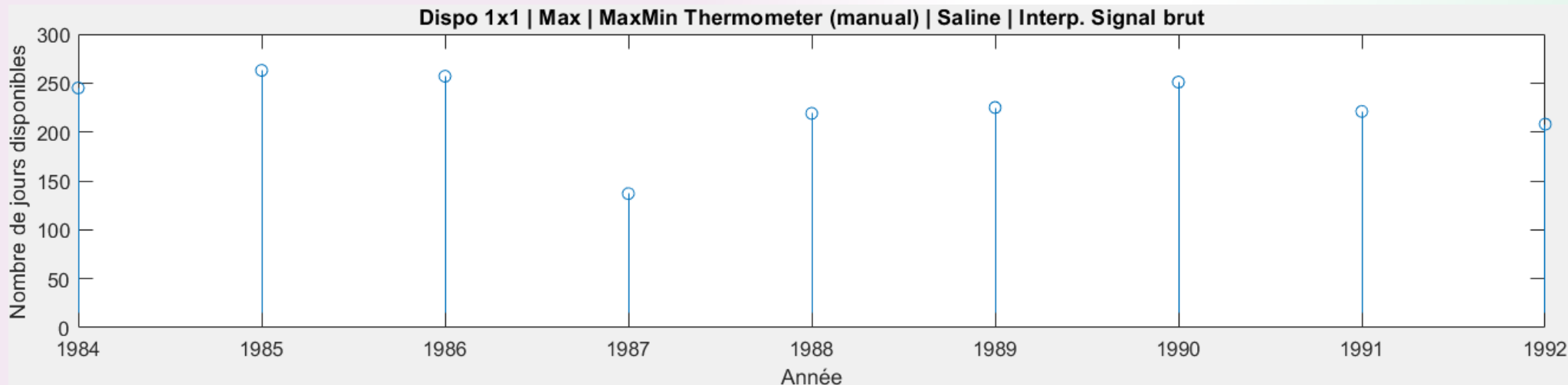


Fig. : Disponibilité des températures maximales en fonction de l'année (en jour en haut, en pourcentage en bas), dans la période de 1984 à 1992.

Conditions de travail : Données (**GARCIA**), Année (**1984-1992**), Adjacence (**1x9**), Type (**Max**), Method (**MaxMin Thermometer manual**), Site (**Saline**)
 Conditions de travail correspondant à la série temporelle ayant le plus de données, à savoir 3029 lignes (jours) de 1984 à 2007.

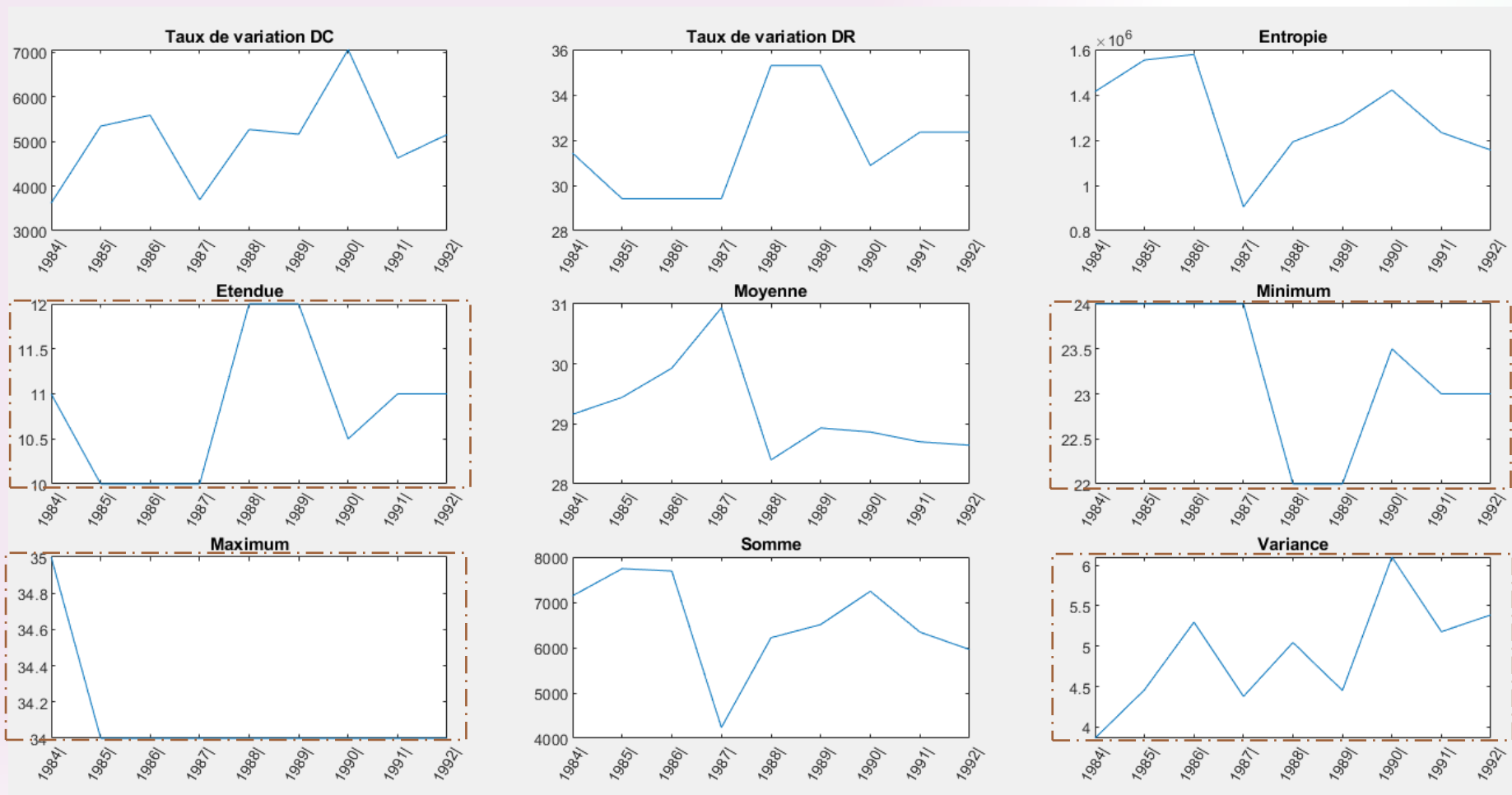


Fig. : Quelques paramètres statistiques calculés sur les températures maximales, dans la période de 1984 à 1992.

1986 et 1990 sont des années très différentes des autres (**Taux de variation DC**).

Les signaux sont plus désordonnés qu'en 1987 (**Entropie**).

L'écart entre le minimum et le maximum semble augmenté (**Etendue**) : étendue de l'étendue = 12-10=2.

La **Moyenne** indique de faibles variations de température d'une année à l'autre sauf en 1987.

La température la plus basse a été obtenue en 1988 et 1989 (**Minimum**) et la plus élevée de 1985 à 1992 (**Maximum**).

1987 est l'année ayant eu les températures les plus faibles (**Somme**).

Les écarts de températures dans une même année semble croître (**Variance**).

Conditions de travail : Données (**GARCIA**), Année (**1984-1992**), Adjacence (**1x9**), Type (**Max**), Method (**MaxMin Thermometer manual**), Site (**Saline**)
 Conditions de travail correspondant à la série temporelle ayant le plus de données, à savoir 3029 lignes (jours) de 1984 à 2007.

Fréquence (Hz)	$F_{\text{trimestre}} = 1,286 \cdot 10^{-7}$	$F_{\text{semestre}} = 6,43 \cdot 10^{-8}$	$F_{\text{an}} = 3,215 \cdot 10^{-8}$	$F_{2\text{ans}} = 1,6075 \cdot 10^{-8}$
Harmonique n°1 (Hz)	$2,572 \cdot 10^{-7}$	1,2860e-07	$6,43 \cdot 10^{-8}$	$3,215 \cdot 10^{-8}$
Harmonique n°2 (Hz)	$3,858 \cdot 10^{-7}$	1,9290e-07	$9,6451 \cdot 10^{-8}$	$4,8225 \cdot 10^{-8}$
Période (s)	7 776 000	1 555 200	31 104 000	62 208 000
Symbole	*	□	*	□

$$T_e = 24 \times 60 \times 60 = 86400 \text{ s}$$

$$\Leftrightarrow F_e = 1/T_e = 1,1574 \cdot 10^{-5} \text{ Hz}$$

$$F_{1_{\text{car}}} = 5,141 \cdot 10^{-8} \text{ Hz} \quad (\Leftrightarrow 7,5 \text{ mois})$$

$$F_{2_{\text{car}}} = \text{Harmonique 2 de } F_{\text{an}} \quad (\Leftrightarrow 4 \text{ mois})$$

$$F_{3_{\text{car}}} = \text{Harmonique 2 de } F_{\text{trimestre}} \quad (\Leftrightarrow 2,7 \text{ mois})$$

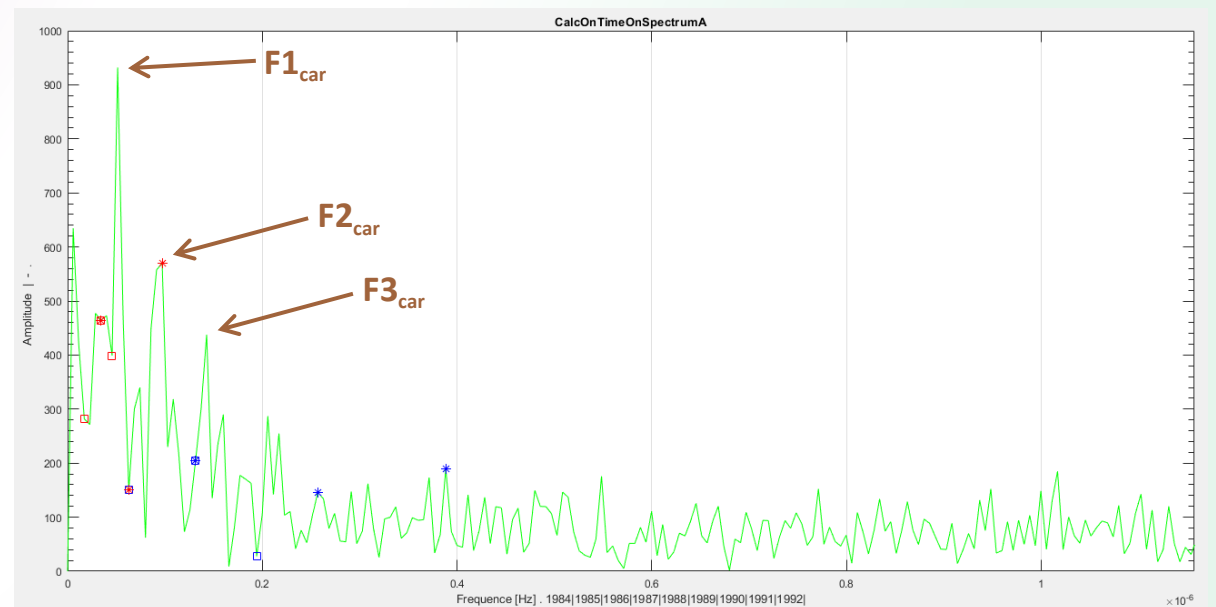
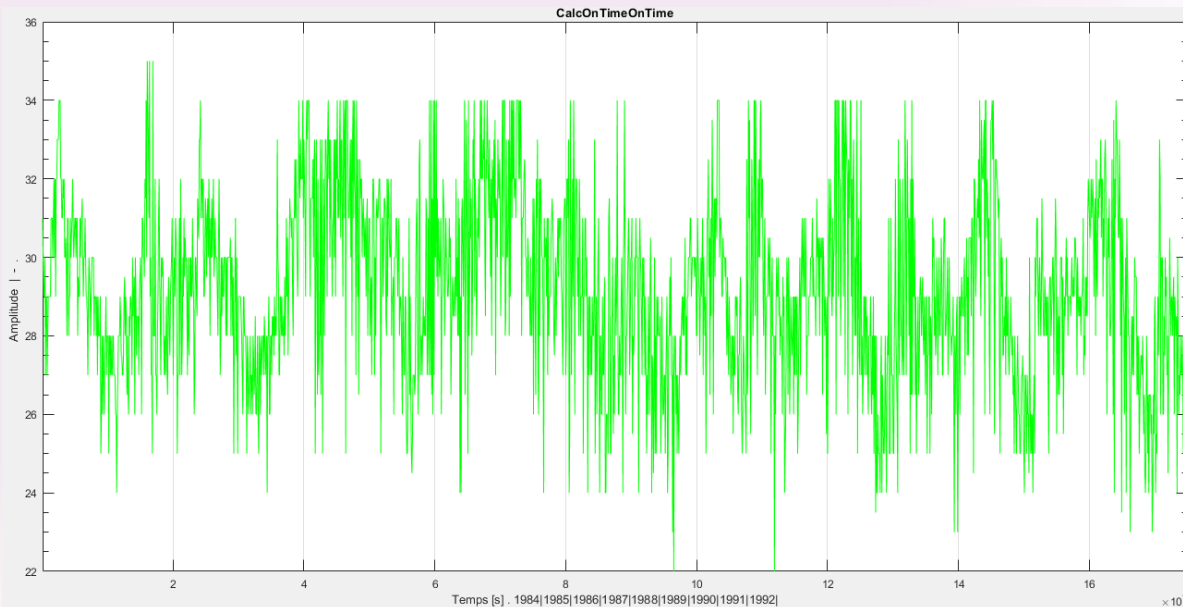
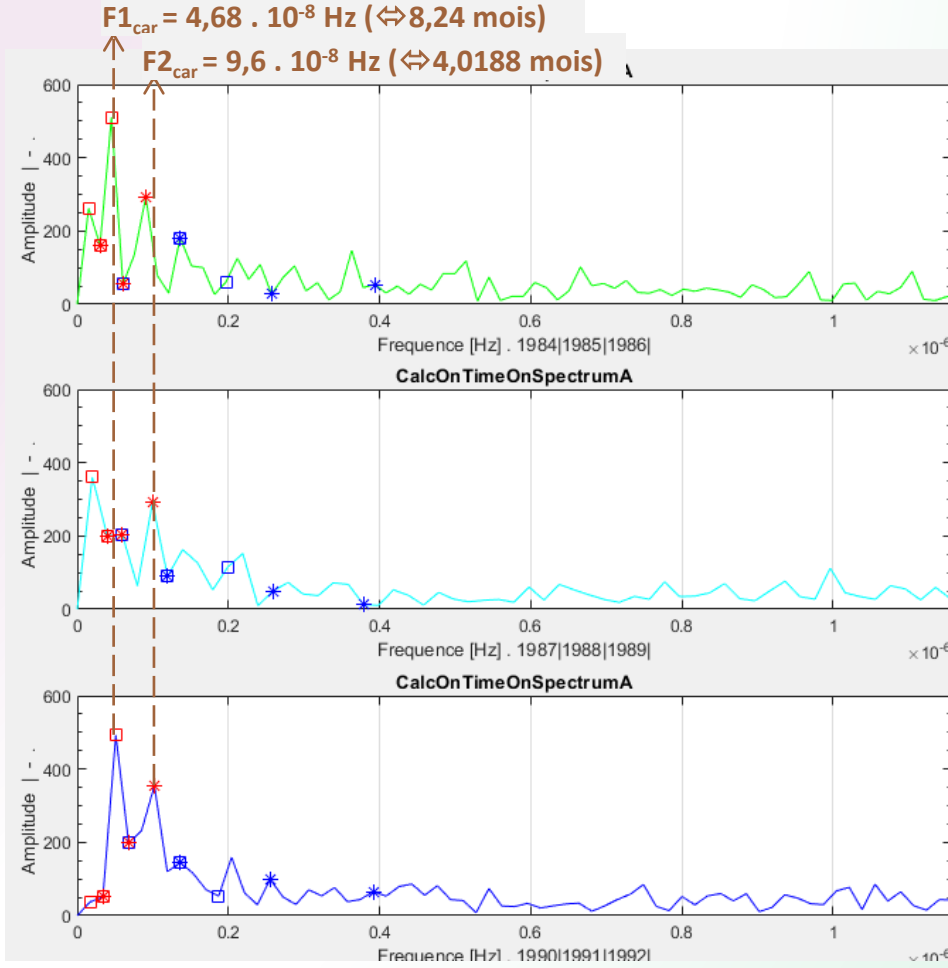
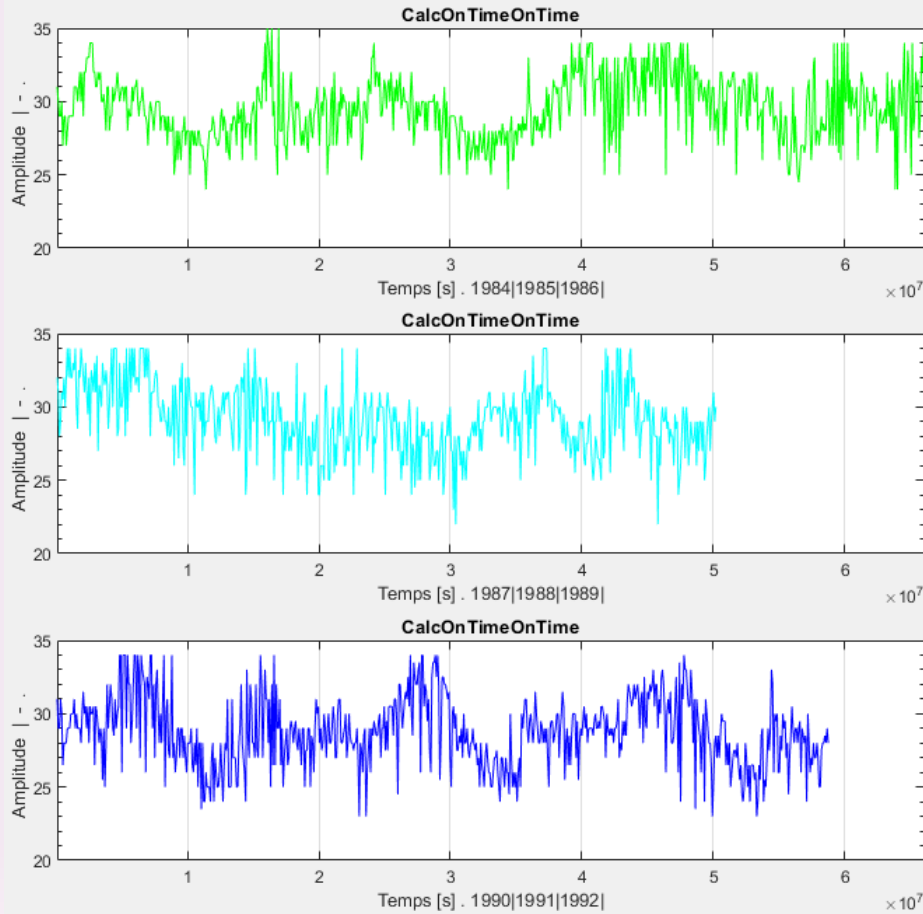


Fig. : Représentations temporelle (gauche) et spectrale (droite) des températures maximales, allant de 1984 à 1992 (9 années \Leftrightarrow 1x9).

Conditions de travail : Données (**GARCIA**), Année (**1984-1992**), Adjacence (**3x3**), Type (**Max**), Method (**MaxMin Thermometer manual**), Site (**Saline**)
 Conditions de travail correspondant à la série temporelle ayant le plus de données, à savoir 3029 lignes (jours) de 1984 à 2007.



Fréquence (Hz)	Symbole
$F_{\text{trimestre}} = 1,286 \cdot 10^{-7}$	*
$F_{\text{semestre}} = 6,43 \cdot 10^{-8}$	□
$F_{\text{an}} = 3,215 \cdot 10^{-8}$	*
$F_{2\text{ans}} = 1,6075 \cdot 10^{-8}$	□

Fig. : Représentations temporelle (gauche) et spectrale (droite) des températures maximales, allant de 1984 à 1992 (3 vecteurs contenant 3 années consécutives ↔ 3x3).

Quelques bases de données de références :

[BILENGA MOUKODOUMA et al. 2023]

- NCDC NOAA GSOD : <ftp://ftp.ncdc.noaa.gov/pub/data/g sod> (en général sans nom utilisateur et mot de passe);
- NCDC NOAA GHCN : <ftp://ftp.ncdc.noaa.gov/pub/data/ghcn> (en général sans nom utilisateur et mot de passe);
- TAMSAT : <https://www.tamsat.org.uk>;

Téléchargement des données :

- Le téléchargement de 1000 fichiers manuellement étant fastidieux, le mieux est de chercher si le fournisseur n'a pas indiqué les informations FTP, nécessaire pour exploiter des logiciels de téléchargement automatisés comme FileZilla Client. L'aspiration du site web du fournisseur (HTTrack) est aussi une alternative à condition de demander l'autorisation;
- Disponibilité de plusieurs bases de données différentes présentes sur le serveur impliquant une attention active;
- Importance capitale à lire le readme.txt ou le fichier détaillant l'organisation de la base de données;

Précautions sur le type de mesures effectuées :

- Nécessité de bien connaître les coordonnées GPS pour le choix adapté des stations de mesures dans la zone souhaitée;
- Dans certains cas, les mesures se font avec ou sans contact. Cela a un impact certain sur les données. Bien renseigner sur la manière dont les données ont été obtenues (exemple de la Télédétection comparée aux mesures réalisées au sol);
- Extension originale des fichiers téléchargés pouvant être contournée en les renommant simplement en .txt ou .csv;
- Fichiers téléchargés impossibles à décrypter parce que méconnaissant l'existence du logiciel dédié au décryptage, stocké parfois dans un dossier auxiliaire sur le serveur de téléchargement.

Environnement de travail optimisé suivant la taille des données :

- Ordinateur fixe ou portable avec un 1^{er} disque dur SSD M.2 (NVMe étant encore plus puissant) comme siège du système d'exploitation et des logiciels de traitement de données;
- Arrêts quotidiens de l'ordinateur de l'utilisateur avec la "veille prolongée" (MS Windows) pour les traitements de données s'étendant sur plusieurs semaines (option très utile pour des ordinateurs peu performants ou lents);
- 2^{ème} disque dur de stockage des données et dédié à l'enregistrement des résultats à posteriori (fichiers, images);
- Préférence accrue pour des systèmes d'exploitations "Light " comme Linux (Ubuntu KaliLinux), Tiny (Windows léger).

Compréhension préalable des données à étudier :

- Choix d'un bon logiciel d'exploration du type de données sélectionnées;
- Exploration longue et approfondie des données sélectionnées (indispensable);**
- Vérification du taux de disponibilité, puis sélection filtrage des données.

Précautions au niveau de l'analyse:

- Travailler sur une copie des données et non les données originales (suppression ou modification par inadvertance);**
- Moins il y a de données, plus la résolution est faible, surtout dans les analyses à base de transformée de Fourier (spectre, cepstre, analyse d'enveloppe, ...). Une analyse en Temps-Fréquence peut être une alternative plus sûre;**
- Manque de robustesse lorsque des données de taille trop différentes sont comparées dans des analyses graphiques (exemple de la dispersion autour des principaux pics fréquentiels.

Boîte à outils AdvDatAna (Advanced Data Analysis/Analyse Avancée des Données) :

- Traitement automatisé des données, avec prise en main facilitée par les scripts de démonstration;**
- En libre téléchargement sur les dépôts [Matlab File Exchange](#) et [Github](#) et adaptée à l'étude de tout type de données et le Big Data en particulier;
- Peu de programmation requise (**symbole %** devant 1 ligne permettant de la désactiver à l'exécution);

Données météorologiques du parc national de la Lopé :

- Facilement accessible pour les chercheurs, ingénieurs, stagiaires, ...;
- Gros effort de l'ANPN du Gabon dans la collecte de données depuis 1984 à aujourd'hui;**
- Problèmes liés aux données manquantes dans certains cas, réduisant des études plus fiables et pertinentes (si réalisées sur le long terme);**
- Perte de résolution si les données sont trop courtes (étude sur 1 ou 2 an);
- Nécessité de travailler sur des données plus complètes et/ou plus importantes (mesures par heure).

Caractérisation des températures du parc national de la Lopé :

- ❑ Conditions de travail : **Données (GARCIA), 1984-1992, Min et Max, MaxMin Thermometer manual, Saline;**
- ❑ Phénomènes cycliques :
 - environ 6/7 mois (alternance des périodes **longue saison de pluie + petite saison sèche VS petite saison de pluie + longue saison sèche**);
 - environ 3/4 mois (alternance des périodes saisonnières);
 - et 1 mois (peu visible);
- ❑ Rappels : longue saison de pluie (Sept. → Nov.), petite saison sèche (Déc. → Fév.),
petite saison de pluie (Mars → Mai), longue saison sèche (Juin → Août);
- ❑ Presque 2,5 fois plus de dispersion (Etendue) sur les Températures minimales (5°) que maximales (2°);
- ❑ Températures minimales : (Minimum) Δ de 4°, (Maximum) Δ de 1,5°;
- ❑ Températures maximales : (Minimum) Δ de 2°, (Maximum) Δ de 1°;
- ❑ Presque 3,5 fois plus de dispersion annuelle (Variance) sur les températures maximales que minimales.

**Changement climatique
entre 1984 et 1992**

ABERNETHY, K., BUSH, E.R., FORGET, P.M., MENDOZA, I. AND MORELLATO, L.P.C., 2018. *Current issues in tropical phenology: a synthesis*. Biotropica, 50(3), pp.477-482. <https://mnhn.hal.science/mnhn-02151136>

ADAMESCU, G. S., PLUMPTRE, A. J., ABERNETHY, K. A., POLANSKY, L., BUSH, E. R., SHOO, L. P., ... & BEALE, C. M. 2018. *Annual cycles are the most common reproductive strategy in African tropical tree communities*. Biotropica, 50(3), pp.418-430.
<https://www.researchgate.net/publication/324601006> Annual cycles are the most common reproductive strategy in African tropical tree communities

Djes-Fréys BILENGA MOUKODOUMA, Donald Romarick ROTIMBO MBOUROU, Christiane ATTEKE NKOULEMBENE, Christophe DENIS, *A temperatures variation favor human-elephant conflict in Gabon's Lékédi National Park*, International Journal of Advanced Engineering Research and Science, Volume 10, Issue 8, Pages 1-26, 3 august, 2023. <https://dx.doi.org/10.22161/ijaers.108.2>, <https://hal.science/hal-04180743v1>

BUSH, E.R., ABERNETHY, K.A., JEFFERY, K., TUTIN, C., WHITE, L., DIMOTO, E., DIKANGADISSI, J.T., JUMP, A.S. and BUNNEFELD, N., 2017. *Fourier analysis to detect phenological cycles using long-term tropical field data and simulations*. Methods in Ecology and Evolution, 8(5), pp.530-540.
<https://doi.org/10.1111/2041-210X.12704>

BUSH, E.R., BUNNEFELD, N., DIMOTO, E., DIKANGADISSI, J.T., JEFFERY, K., TUTIN, C., WHITE, L. and ABERNETHY, K.A., 2018. *Towards effective monitoring of tropical phenology: maximizing returns and reducing uncertainty in long-term studies*. Biotropica, 50(3), pp.455-464.
<https://doi.org/10.1111/btp.12543>

E. R., BUSH. Tropical Phenology In A time of Change, Thesis, Université de Stirling, December 2018. <https://core.ac.uk/download/pdf/199408393.pdf>

BUSH, E.R., MITCHARD, E.T., SILVA, T.S., DIMOTO, E., DIMBONDA, P., MAKAGA, L. and ABERNETHY, K. (2020) *Monitoring Mega-Crown Leaf Turnover from Space*. Remote Sens. 12, 429. <https://www.mdpi.com/2072-4292/12/3/429>

BUSH, E.R., JEFFERY, K., BUNNEFELD, N., TUTIN, C., MUSGRAVE, R., MOUSSAVOU, G., MIHINDOU, V., MALHI, Y., LEHMANN, D., NDONG, J.E., MAKAGA, L. and ABERNETHY, K. 2020. *Rare ground data confirm significant warming and drying in western equatorial Africa*. PeerJ, 8, p.e8732
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7164428/>

BUSH, E.R., WHYTOCK, R.C., BOURGEOIS, S., BUNNEFELD, N., CARDOSO, A.W., DIKANGADISSI, J.T., DIMBONDA, P., DIMOTO, E., NDONG, J.E., JEFFERY, K.J., LEHMANN, D., MAKAGA, L., MOMBOUA, B., MOMONT, L.R.W., TUTIN, C.E.G., WHITE, L.J.T., WHITTAKER, A. and ABERNETHY, K., 2020. *Long-term collapse in fruit availability threatens Central African forest megafauna*. *Science*, 370(6521), pp1219-1222, <https://doi.org/10.1126/science.abc7791>

CARDOSO, A.W., MALHI, Y., OLIVERAS, I., LEHMANN, D., NDONG, J.E., DIMOTO, E., BUSH, E., JEFFERY, K., LABRIERE, N., LEWIS, S.L., WHITE, L.T., BOND, W., and ABERNETHY, K. 2019. *The Role of Forest Elephants in Shaping Tropical Forest–Savanna Coexistence*. *Ecosystems*, 23(3), pp.602-616. <https://www.researchgate.net/publication/334816830> *The Role of Forest Elephants in Shaping Tropical Forest–Savanna Coexistence*

Damien GARCIA, Robust smoothing of gridded data in one and higher dimensions with missing values, *Computational Statistics & Data Analysis*, Volume 54, Issue 4, 2010, Pages 1167-1178, ISSN 0167-9473, <https://doi.org/10.1016/j.csda.2009.09.020>.

Damien GARCIA (2023). Inpaint over missing data in 1-D, 2-D, 3-D,... ND arrays <https://www.mathworks.com/matlabcentral/fileexchange/27994-inpaint-over-missing-data-in-1-d-2-d-3-d-nd-arrays>, MATLAB Central File Exchange. Retrieved October 24, 2023.

HACKET-PAIN, A., FOEST, J. J., PEARSE, I. S., LAMONTAGNE, J. M., KOENIG, W. D., VACCHIANO, G., BOGDZIEWICZ, M., CAIGNARD, T., CELEBIAS, P., VAN DORMOLEN, J., FERNÁNDEZ-MARTÍNEZ, M., MORIS, J. V., PALAGHIANU, C., PESENDORFER, M., SATAKE, A., SCHERMER, E., TANENTZAP, A. J., THOMAS, P. A., VECCHIO, D., ... BUSH, E.R., ASCOLI, D. (2022). *MASTREE+: Time-series of plant reproductive effort from six continents*. *Global Change Biology*, 28, pp3066–3082 <https://onlinelibrary.wiley.com/doi/full/10.1111/gcb.16130>

J. LEE, H. QIU, G. YU, J. LIN, and Rexnord Technical Services, IMS, University of Cincinnati, *Bearing Data Set*, NASA Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA, 2007. <https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/>

Donald Romarick ROTIMBO MBOUROU, *Contribution à la surveillance et au diagnostic de défauts dans les systèmes de transmission par courroie*, Mécanique [physics], thèse en cotutelle entre le Laboratoire d'Analyse des Signaux et des Processus Industriels LASPI (Université Jean Monnet de Saint-Etienne/Université de Lyon en FRANCE) et l'Ecole Polytechnique de Masuku EPM (Université des Sciences et Techniques de Masuku au Gabon), 31/10/2016. <https://hal.science/tel-02020540v1>

ROTIMBO MBOUROU Donald Romarick, *Boîte à outils Matlab d'Analyse Avancée des Données 1D AdvDatAna (Advanced Data Analysis,) pour la surveillance, le diagnostic et le pronostic*, Laboratoire de Recherche Multidisciplinaire en Environnement, LARME, USTM, 21 décembre 2020, Franceville, Haut-Ogooué, GABON.

https://www.mathworks.com/matlabcentral/fileexchange/79003-toolbox_advdatana_2020y12m21d19h08m

https://github.com/rod25361/Toolbox_AdvDatAna_2020y12m21d19h08m/releases/tag/1.0.0.1

TRAVERS, H., SELINSKE, M., NUNO, A., SERBAN, A., MANCINI, F., BARYCHKA, T., BUSH, E., RASOLOFOSON, R.A., WATSON, J.E. AND MILNER-GULLAND, E.J., 2019. *A manifesto for predictive conservation*. Biological Conservation, 237, pp.12-18.

<https://doi.org/10.1016/j.biocon.2019.05.059>

TUTIN, C. E. G., ABERNETHY, K., MUSGRAVE, R., JEFFERY, K. J., MOUSSAVOU, G., MIHINDOU, V., MALHI, Y., LEHMANN, D., NDONG J. E., MAKAGA, L., BUSH, E. R. (2019): *Lopé Weather Dataset. Version 1.1*. University of Stirling. Faculty of Natural Sciences. Dataset.

<http://hdl.handle.net/11667/133>.

Guojie WANG, Damien GARCIA, Yi Liu, Richard DE JEU, A. Johannes DOLMAN, A three-dimensional gap filling method for large geophysical datasets: Application to global satellite soil moisture observations, Environmental Modelling & Software, Volume 30, 2012, Pages 139-142, ISSN 1364-8152, <https://doi.org/10.1016/j.envsoft.2011.10.015>.

WILLIAMSON, J., TEH, E., JUCKER, T., BRINDLE, M., BUSH, E., CHUNG, A. Y., PARRETT, J., LEWIS, O. T., ROSSITER, S. J., SLADE, E. M. (2022) *Local-scale temperature gradients driven by human disturbance shape the physiological and morphological traits of dung beetle communities in a Bornean oil palm–forest mosaic*. Functional Ecology, 00, 1– 13. <https://doi.org/10.1111/1365-2435.14062>

Avez-vous des questions ou même

des remarques suite à cette présentation?