



UNIVERSITÉ
CAEN
NORMANDIE

Rapport d'études

Université de Caen Normandie

Département de Mathématiques

Master 2 Professionnel

Statistiques Appliquées à l'Analyse Décisionnelle

Étude des températures des cours d'eau Haut-Normands

Étudiants :

Dagbegnon Herman DASSI

Mamadou Daya SYLLA

Enseignant Référent :

Mr. Bruno DARDAILLON

17 mars 2021

Résumé

Ce présent document constitue un rapport d'études sur les températures des cours d'eau Haut-Normands dans le cadre la validation de notre 2ème année de Master Professionnel spécialité Statistiques à l'Université de Caen Normandie.

Ce projet d'étude consiste à analyser les structure des températures des cours d'eau haut-normands sur la période 2013-2018. afin d'identifier le cycle saisonnier, journalier et bien sûr la nappe phréatique.

Ce document présente aussi les différentes étapes effectuées dans la réalisation de ce projet, en commençant par l'évolution de la couche présentation, puis l'analyse et la prédiction des défaillances pour le prochain trimestre suivant la période d'étude.

A titre informatif, les sorties de ce document ont été réalisés avec le logiciel RStudio et R Shiny.

Table des matières

Introduction	2
1 Présentations et Contexte de l'étude	3
1.1 Odon	3
1.2 Contexte de l'étude	3
2 Données et Méthodologie	5
2.1 Données	5
2.2 Méthodologie	6
3 Analyse des résultats	9
3.1 Tableaux descriptifs	9
3.2 Analyse des différentes courbes	13
3.3 Décomposition de la série	16
3.4 Corrélation des températures de l'Odon	21
3.5 Saisonnalité mensuelle des températures de l'Odon	25
3.6 Analyse en Composante Indépendante	29
3.6.1 Composante journalière	29
4 Application Shiny	34
Conclusion	37
Bibliographie	38
A Code	39

Introduction

L'évolution thermique d'un cours d'eau permet, entre autres, de mettre en évidence à long terme, l'effet du changement climatique. A cette fin, des données sont collectées par divers organismes, centralisées, et analysée avec différentes techniques.

Dans le cadre de notre projet de Master, nous avons à analyser des séries chronologiques constituées des données de températures de l'Odon, un cours d'eau situé en Normandie. L'objectif est de déterminer les différentes composantes qui expliquent la température de ce cours d'eau. Pour ce faire, nous allons utiliser l'Analyse en Composante Indépendante, exécutée avec le package `fastICA`.

Dans un premier chapitre, nous allons présenter le contexte de l'étude. Nous allons dans un second chapitre, vous présenter la data qui sera analysée. Dans un troisième chapitre, il sera question de la méthodologie utilisée pour analyser la data. Le quatrième chapitre sera consacré aux résultats de nos analyses et le cinquième à la construction d'une application pour visualiser la synthèse des données.

Chapitre 1

Présentations et Contexte de l'étude

1.1 Odon

D'une longueur de 47.3km, l'Odon est une rivière située en Normandie de type granitique c'est-à-dire aux eaux acides. Il prend sa source à Onde-fontaine, sur les hauteurs du Nord du Bocage Virois, et parcourt l'Ouest de la plaine de Caen jusqu'au confluent à Caen. Confluent de 12 affluents (11 ruisseaux et 1 rivière), le cours d'eau traverse 20 communes françaises situées tous dans le département du Calvados.

1.2 Contexte de l'étude

L'étude de la température de l'eau, du fait du changement climatique, suscite un intérêt de plus en plus croissant. En effet, partant simplement du principe que la température de l'eau est dépendante de celle de l'air, du sol, etc, il apparaît clair que la température de l'eau joue un rôle majeur dans le fonctionnement des écosystèmes aquatiques fluviaux notamment dans une perspective de réchauffement climatique.

L'étude de la température de l'eau d'un cours d'eau permet de déterminer la concentration en oxygène dissous dans l'eau. Oxygène et température sont deux paramètres essentiels à la vie aquatique. Certaines espèces aquatiques choisissent leurs habitats en tenant compte des combinaisons de facteurs dont la disponibilité en oxygène (celle-ci dépend principalement de la température de l'eau et des mouvements de l'eau).

Une légère augmentation de température va engendrer une diminution de l'oxygène et une accélération de la croissance des plantes, ce qui pourrait perturber les organismes les plus sensibles. La température de l'eau est aussi un paramètre important pour la production d'eau potable et la pêche (Hannah et Gardner, 2015).

Le Service d'Administration Nationale des données et des Référentiels sur l'Eau (SANDRE) a été mis en place afin de favoriser la mise en place d'un langage commun pour les données sur l'eau

compte tenu des acteurs qui interviennent dans le domaine de l'eau. Cette homogénéisation du langage est très utile au Système d'Information sur l'Eau (SIE), organisation créée depuis 1992. Il est composé d'un ensemble cohérent de dispositifs, processus et flux d'information, par lesquels les données relatives à l'eau sont acquises, collectées, conservées, organisées, traitées et publiées de façon systématique.

Les objectifs du SIE sont définis par le schéma national des données sur l'eau (SNDE) qui constitue par ailleurs le référentiel technique du système. La SANDRE dans le dictionnaire de données définit les notions fondamentales relatives à l'acquisition des données de température physique des cours d'eau.

Plusieurs études ont été réalisées sur l'étude des cours d'eau en général et sur l'Odon en particulier. Par le passé, la modélisation de la température des rivières a été réalisée suivant différents types de modèles. Les modèles à base physique demandent la connaissance de nombreux paramètres pour représenter les échanges de matière et d'énergie. Les modèles à base statistique se fondent sur des corrélations entre données. La plupart tente de mettre en relation la température de l'eau avec la température de l'air.

Chapitre 2

Données et Méthodologie

2.1 Données

Nos données ont été récupérées sur le site de la « NAIĀDE » qui est un portail de centralisation de données. Ces données sont produites par des organismes pour répondre à un besoin (étude, réglementation, recherche...) puis stockées dans des banques de référence qui peuvent stocker des données de plusieurs producteurs et répondant à plusieurs besoins.

La Direction Régionale de l'Environnement, de l'Aménagement et du Logement de Normandie (DREAL) est un intervenant du site naïade. Dans le cadre des réseaux de suivi du programme de surveillance des masses d'eau, et en partenariat avec d'autres organismes, elle produit des données pour l'évaluation de l'état écologique des eaux.

Depuis 2011, elle assure le suivi thermique des cours d'eaux normands grâce à des sondes disposées à des endroits spécifiques dans les différents cours d'eaux. Les sondes sont attachées aux racines des arbres, en position ombragée. Elles sont installées à une profondeur garantissant leur immersion complète quelle que soit la saison.

Les mesures de températures correspondent à des mesures horodatées de la température de l'eau exprimée en degrés Celsius. La mesure est effectuée toutes les deux heures. Les sondes utilisées sont des HOBO water Temp Pro v2 (U22-001). Leur précision est de $\pm 0,2^{\circ}\text{C}$. Les données peuvent être lues grâce à deux interfaces : optique et USB. La carte ci-dessous montre l'emplacement de quelques sondes dans les différents cours d'eaux dans lesquels elles ont été disposées.

Les données utilisées dans le cadre de cette étude sont des données journalières collectées toutes les 2 heures sur la période 2011 à 2017 à partir de 4 sondes (1, 2, 4, 5) installées dans la rivière Odon. Pour chacune des sondes, nous avons 26625 observations sur les 6 années d'étude.

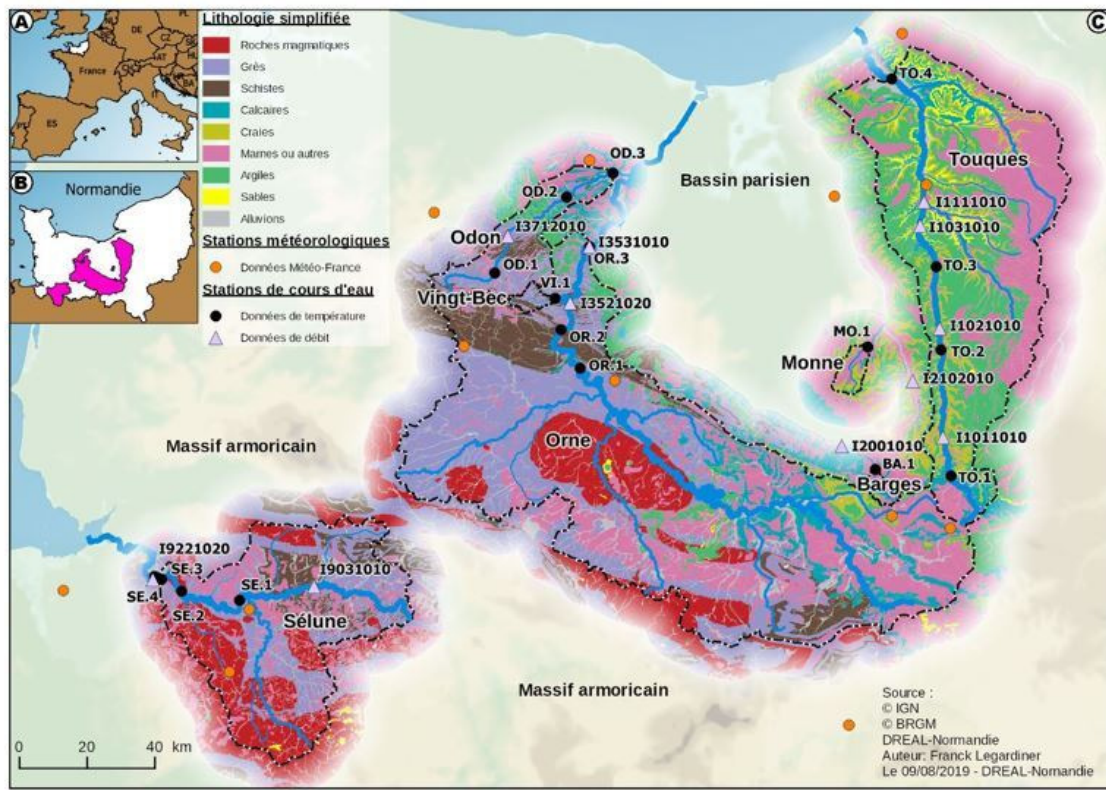


FIGURE 2.1 – Positions des sondes dans les cours d’eaux normands

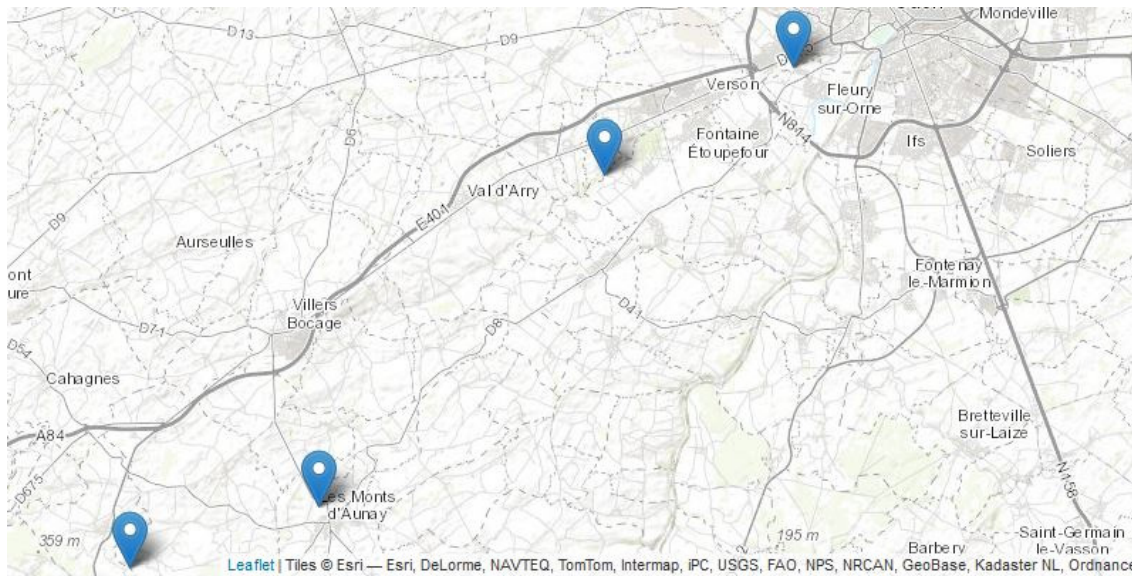


FIGURE 2.2 – Localisations des Sondes (812, 813, 815 et 816) positionnées de bas en haut

2.2 Méthodologie

Les modèles déterministes et les modèles statistiques sont les deux grandes familles de modèles utilisées pour l'étude des températures des cours d'eau. Selon **Loubna Benyahya and al** et N.

El-Jabi and al, les modèles déterministes de la température de l'eau sont plus nombreux et nécessitent souvent un grand nombre d'intrants qui sont parfois difficilement mesurables. Ces modèles utilisent la méthode du bilan énergétique d'un plan d'eau pour déterminer la température naturelle d'une rivière à partir des données hydrométéorologiques et caractéristiques du cours d'eau.

Les modèles statistiques (paramétriques et non paramétriques) quant à eux s'appuient sur les informations de la série temporelle de l'eau et/ ou de l'air mesurée. Ils offrent l'avantage de nécessiter peu de données et un temps de développement habituellement moins long que les modèles déterministes.

La synthèse suivante a été faite à partir du document de Loubna Benyaha and all.

Une petite synthèse sur les méthodes statistiques :

Les modèles statistiques non paramétriques sont des outils efficaces lorsque les systèmes à modéliser sont non-linéaires. La souplesse et le fait qu'il n'y ait pas de paramètres à estimer sont des avantages qu'offrent ces modèles. Cependant, il est difficile d'établir une prévision avec ces méthodes du fait de l'absence de fonctions mathématiques conventionnelles. Les réseaux de neurones artificiels et les k-voisins les plus proches sont des exemples de méthodes appartenant à cette famille.

Les modèles statistiques paramétriques regroupent les méthodes de régression linéaires, non linéaires, autorégressif (ARMA) et périodiques autorégressives (PAR, PARMA). Les modèles statistiques existants sont tous basés sur l'hypothèse de *stationnarité* des statistiques des séries chronologiques de la température de l'eau.

L'utilisation des méthodes autorégressif sur des séries chronologiques à petit pas de temps a permis de les séparer en composantes *saisonnnières* et en composantes *résiduelles*. La composante saisonnière représente les fluctuations qui se répètent plus ou moins régulièrement chaque année. Le reste des fluctuations est prises en compte par la composante résiduelle.

Pour notre étude, nous allons privilégier cette dernière approche et plus particulièrement l'Analyse en Composante indépendante (ACI).

Pourquoi l'ACI (Analyse de Composantes Indépendantes) ?

L'ACI est une méthode statistique d'analyse de données utilisée pour traiter des observations multivariées afin d'en extraire des composantes linéaires aussi indépendantes que possible. En série temporelles, l'ACI peut être utilisé pour séparer des composants indépendants d'une entrée de signaux mixtes.

La séparation avec l'ACI des signaux mixtes est basée sur deux hypothèses :

- *les signaux sources sont indépendants les uns des autres*
- *les valeurs de chaque signal source ont une distribution non gaussienne*

D'une manière générale, les travaux effectués sur l'ACI, peuvent se classer en deux catégories de méthodes : les méthodes basées sur les concepts algébriques en faisant appel aux statistiques d'ordre supérieur (Common, 1994 ; Cardoso, 1999 ; Achard, 2003) et les méthodes neuronales dans lesquelles l'ACI est reformulée sous forme de règles d'apprentissage non dirigées (Hyvärinen et al.,

2001).

Dans le cadre de notre étude, les sondes utilisées pour enregistrer les données de températures reçoivent des signaux mixtes (températures de l'eau). Les sondes disposées à différents endroits du cours d'eau, l'enregistrement d'une sonde n'a aucun effet sur celui des autres. Chacune des sondes enregistrent donc des données de températures, qualifiées de mélange de sources, et leurs enregistrements sont dénotés OD1i, OD2i, OD4i et OD5i, $i=1, \dots, n$ où n représente le nombre d'observations.

Soit X la température mesurée.

$$X = A \% * \% S$$

S : Vecteur des signaux sources originaux à composantes indépendantes inconnues car on ne les mesure pas directement.

A : Matrice de mélange à coefficients non aléatoires mais inconnus puisqu'on ne connaît pas le modèle physique de mélange des sources.

Notre objectif sera de faire ressortir des signaux sources aussi indépendants que possibles.

$$S = W.S$$

Avec la matrice de séparation W qui vaut :

$$W = A^{-1}$$

Ce qui pose le problème de séparation aveugle des sources car pour retrouver la matrice S , nous devons partir de la matrice de mélange de sources X .

Sous certaines conditions, l'ACI, selon Comon (Comon 2004) peut être une solution au problème de séparation aveugle des sources. Lorsqu'il y a autant de sources S que de signaux mixtes X , l'ACI va consister à trouver une matrice W telle que le vecteur obtenu par le produit de W et X aient les composantes les plus indépendantes que possibles (Cardoso, 1999).

Les analyses statistiques effectuées dans le cadre de cette étude porteront sur les moyennes journalières des données collectées au niveau de chaque sonde.

Chapitre 3

Analyse des résultats

3.1 Tableaux descriptifs

TABLE 3.1 – Variation de la température station 812

Température minimale	Moyenne	Médiane	Température maximale	Ecart type
1.566	10.660	10.493	22.902	3.524396

TABLE 3.2 – Variation de la température station 813

Température minimale	Moyenne	Médiane	Température maximale	Ecart type
0.6393	11.5103	11.0298	22.9546	4.024876

TABLE 3.3 – Variation de la température station 815

Température minimale	Moyenne	Médiane	Température maximale	Ecart type
0.5813	11.4882	11.1338	20.1229	3.853359

TABLE 3.4 – Variation de la température station 816

Température minimale	Moyenne	Médiane	Température maximale	Ecart type
1.114	12.088	11.75	22.902	4.163977

Nous observons une augmentation de la température moyenne d'amont en aval. Des 4 stations objets de notre analyse, la température la plus élevée et la plus faible sont observées respectivement au station 813 et 815. Cependant, en considérant la température moyenne de chaque station, c'est la station 816 qui a la température moyenne la plus élevée. On note qu'au niveau de cette station, les relevés de température sont plus hétérogènes (écart-type=4.163977). Tandis qu'au niveau de la station 1, les données sont plus homogènes, ce qui s'explique par un écart-type faible 3.52 (par rapport aux autres).

Nous remarquons que la moyenne des températures est supérieure à la médiane au niveau des 4 stations. Cela pourrait traduire la présence de températures extrêmes dans notre base de données. Compte tenu des différences obtenues entre ces 2 mesures (**2** : 0.48 ; **4** : 0.35 ; **5** : 0.33 ; **1** : 0.167), nous pouvons déduire qu'il y a plus de valeurs extrêmes au niveau de la station 813.

Ce qui est confirmé par le graphe ci-après :

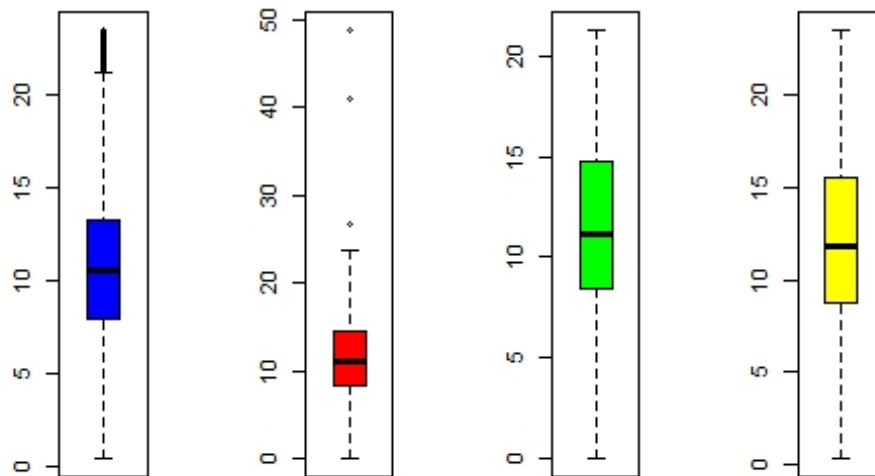


FIGURE 3.1 – Représentation de la température des 4 sites

Afin d'étudier les températures extrêmes, observons les graphes ci-dessous.

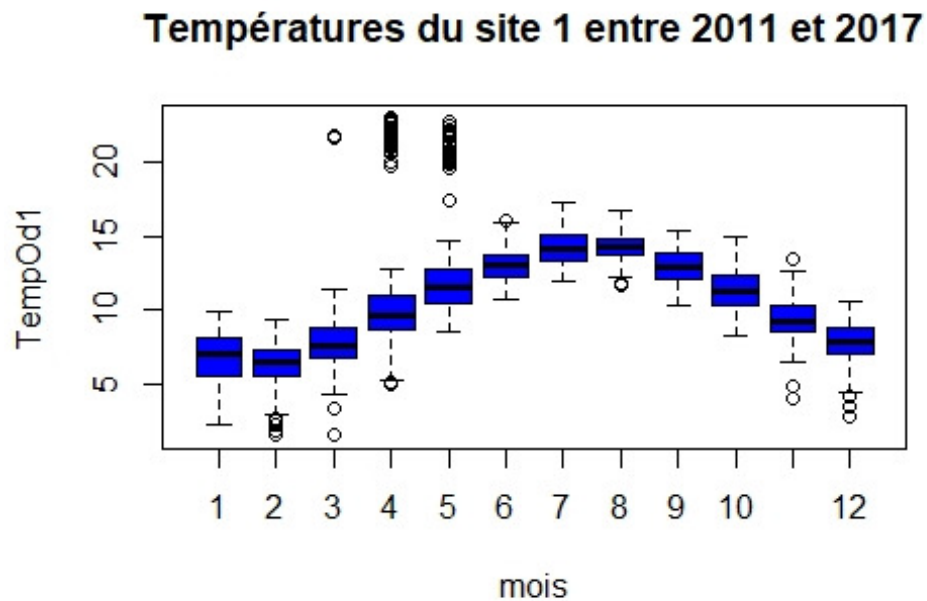


FIGURE 3.2 – Évolution mensuelle de la température au niveau de la station 812

Températures du site 2 entre 2011 et 2017

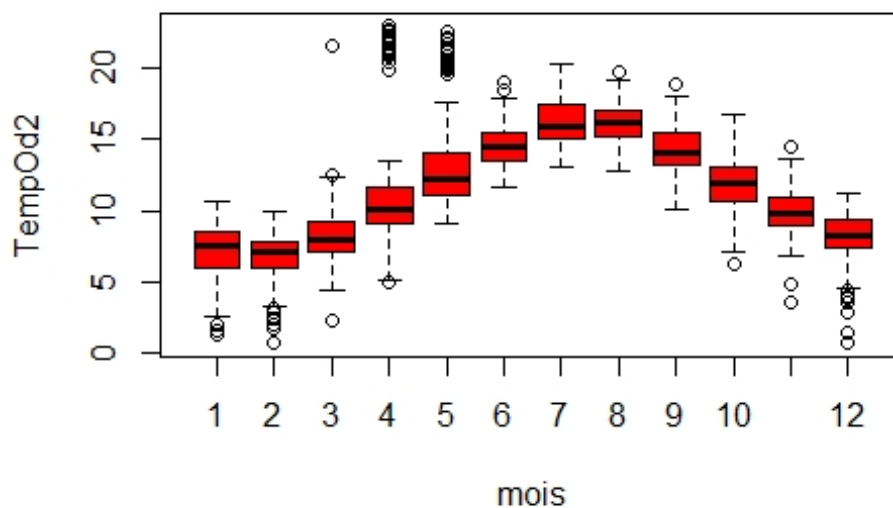


FIGURE 3.3 – Évolution mensuelle de la température au niveau de la station 813

Températures du site 4 entre 2011 et 2017

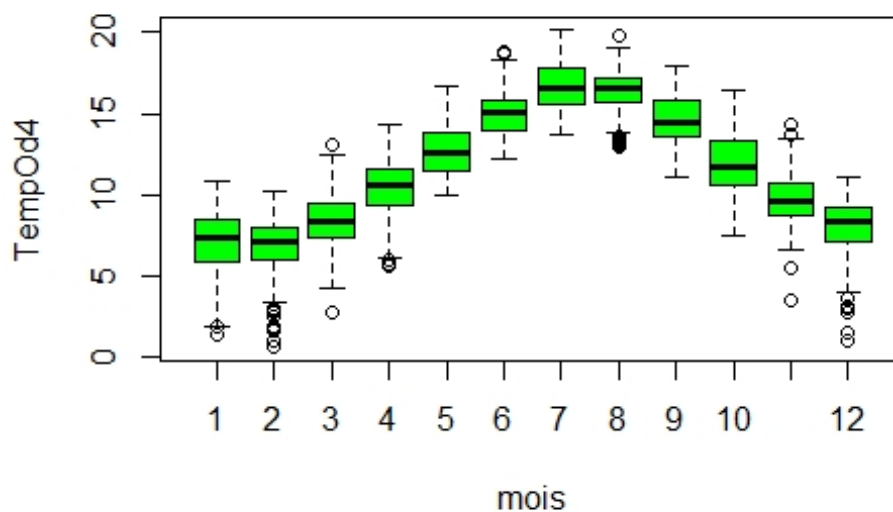


FIGURE 3.4 – Évolution mensuelle de la température au niveau de la station 815

Dans l'ensemble, l'évolution de la température est la même au niveau des 4 stations. Les températures chaudes maximales sont atteintes en juillet et août, correspondant à l'été. Cependant, il faut noter que les valeurs extrêmes sont principalement observées aux mois d'avril et mai (stations

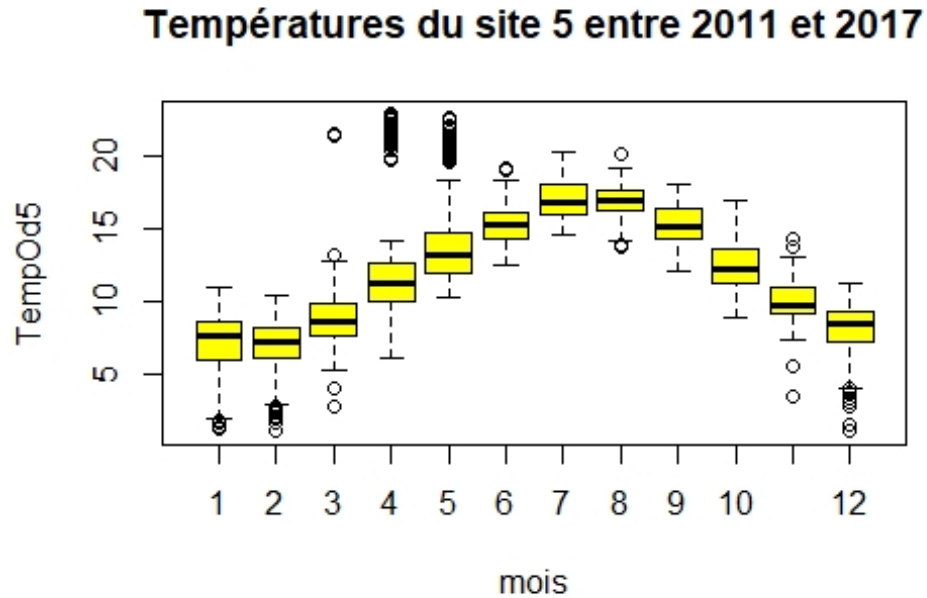


FIGURE 3.5 – Évolution mensuelle de la température au niveau de la station 816

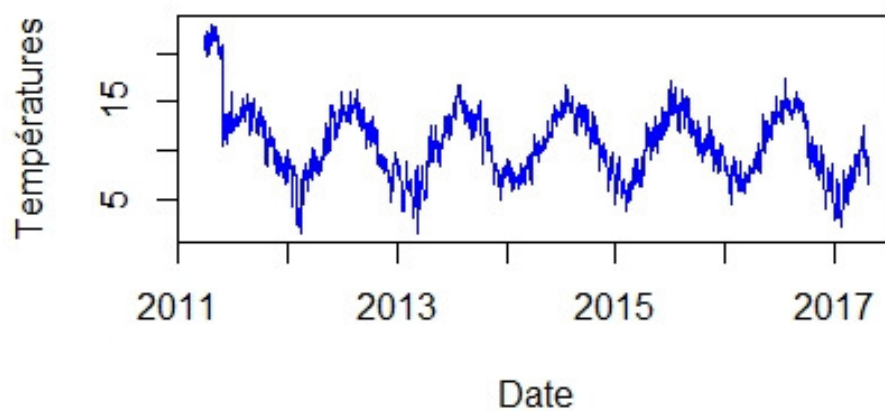
812, 813 et 816). A contrario, les mois de décembre, janvier et février sont les plus froids avec des extrêmes atteints en février et décembre.

3.2 Analyse des différentes courbes

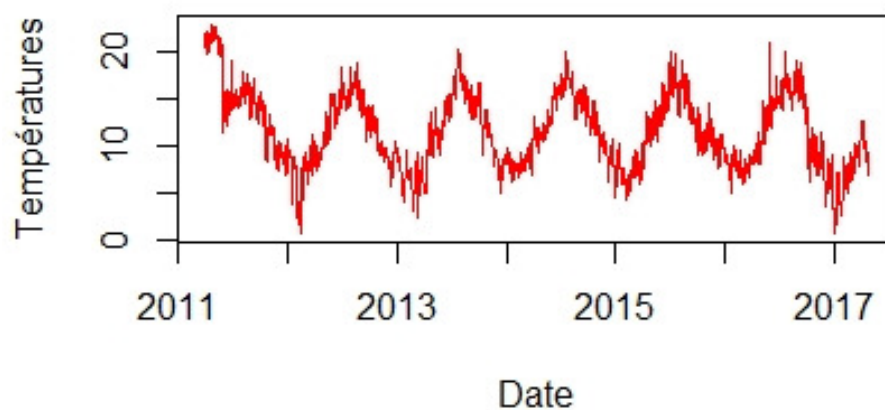
Les graphes ci-dessous traduisent l'évolution moyenne journalière de la température de l'Odon au niveau des différentes stations entre 2011 et 2017. L'évolution est en dent de scie au niveau des 4 stations.

Nos séries sont "*additives*" avec une saisonnalité d'1 an.

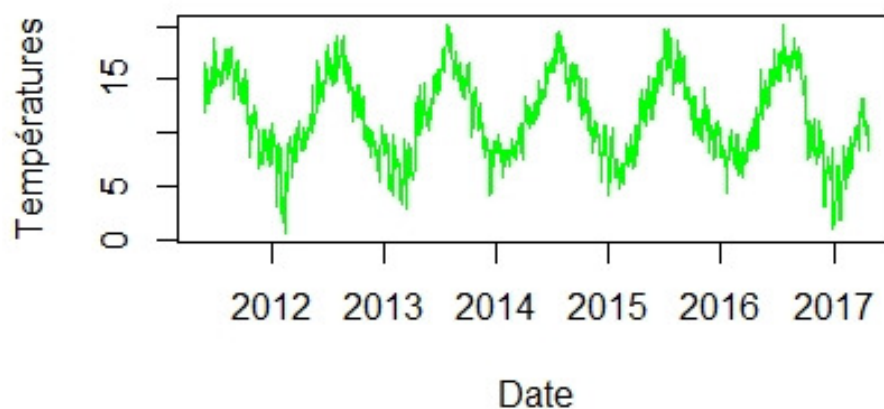
Températures du site 812 entre 2011 et 2017



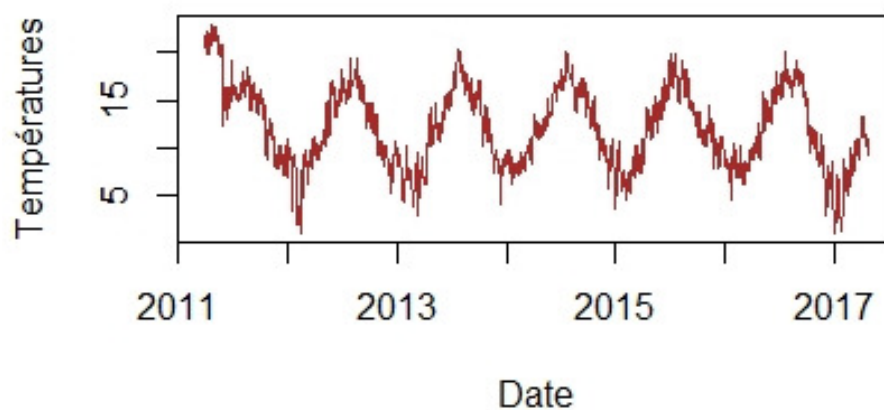
Températures du site 813 entre 2011 et 2017



Températures du site 815 entre 2011 et 2017



Températures du site 816 entre 2011 et 2017



Pour mieux apprécier les variations de la température de l'Odon, nous avons zoomer sur la période allant de janvier 2016 à Avril 2017. Ce graphe nous permet aussi d'apprécier l'augmentation d'amont en aval énoncée plus haut.

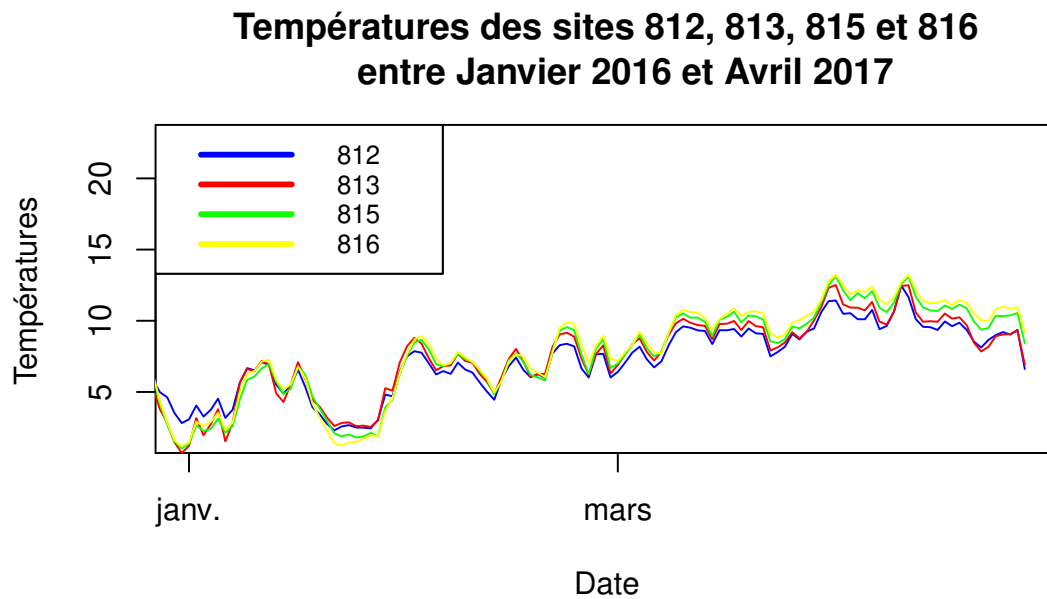


FIGURE 3.6 – Température des 4 sites

3.3 Décomposition de la série

Dans cette section nous allons décomposer notre série afin d'étudier la tendance, la saisonnalité et les résidus. Pour ce faire, nous allons utiliser la fonction `decompose` de R qui permet d'estimer la tendance, la saisonnalité et les résidus.

Decomposition of additive time serie

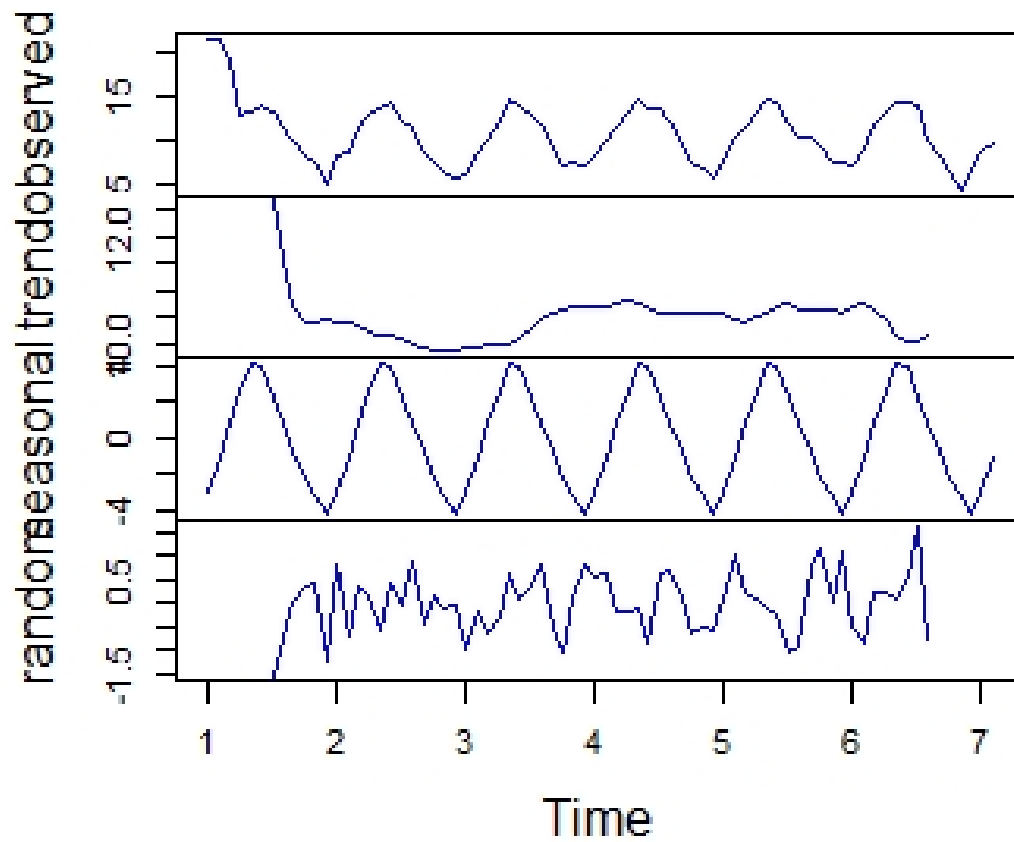


FIGURE 3.7 – Décomposition de la série Od1

Decomposition of additive time serie

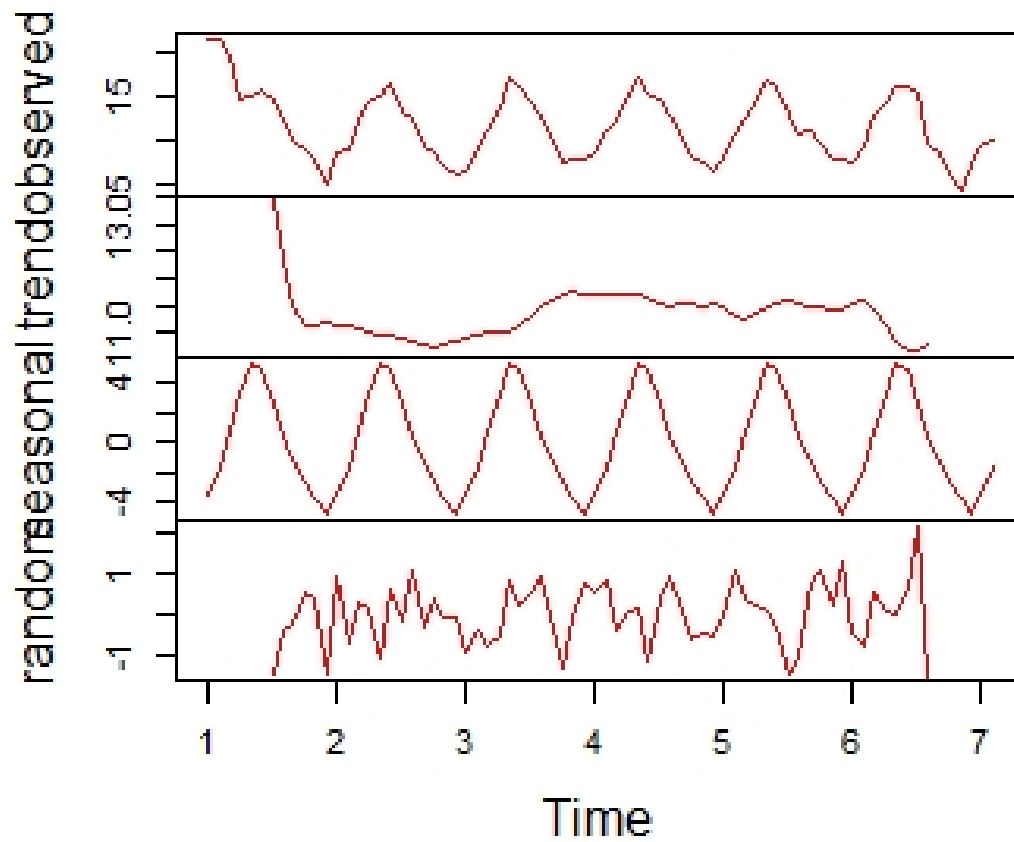


FIGURE 3.8 – Décomposition de la série Od2

Decomposition of additive time serie

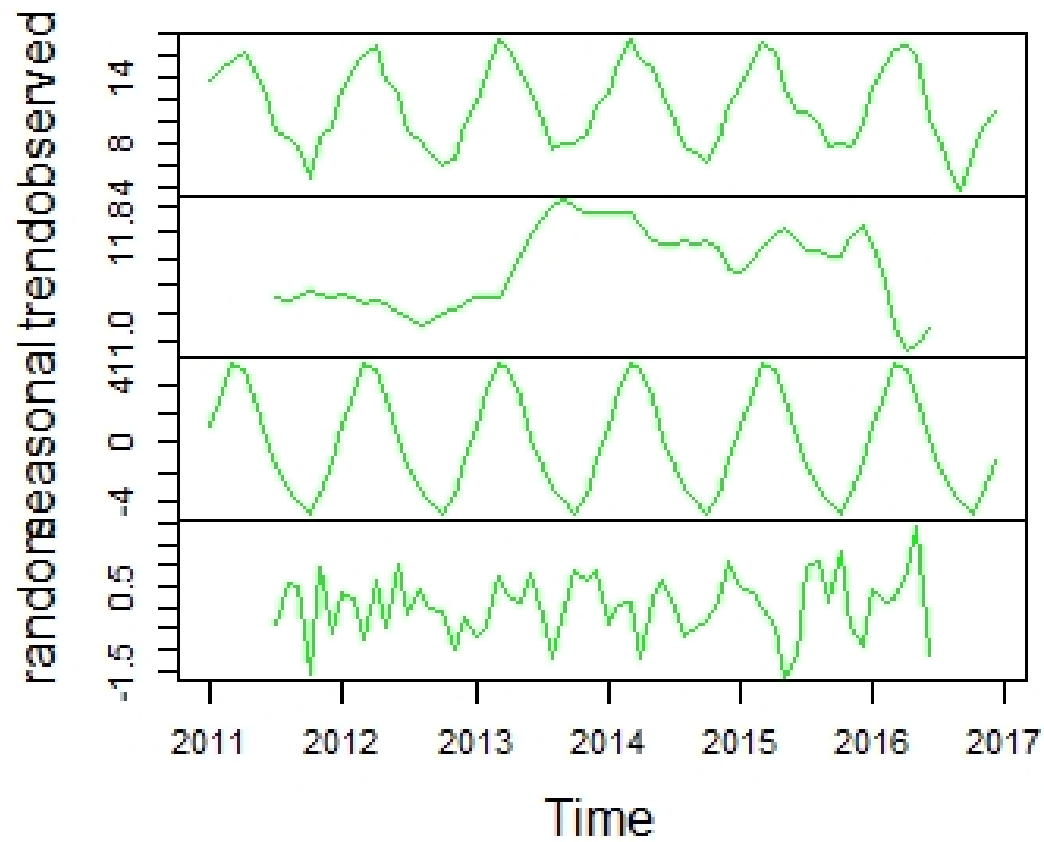


FIGURE 3.9 – Décomposition de la série Od4

Decomposition of additive time serie

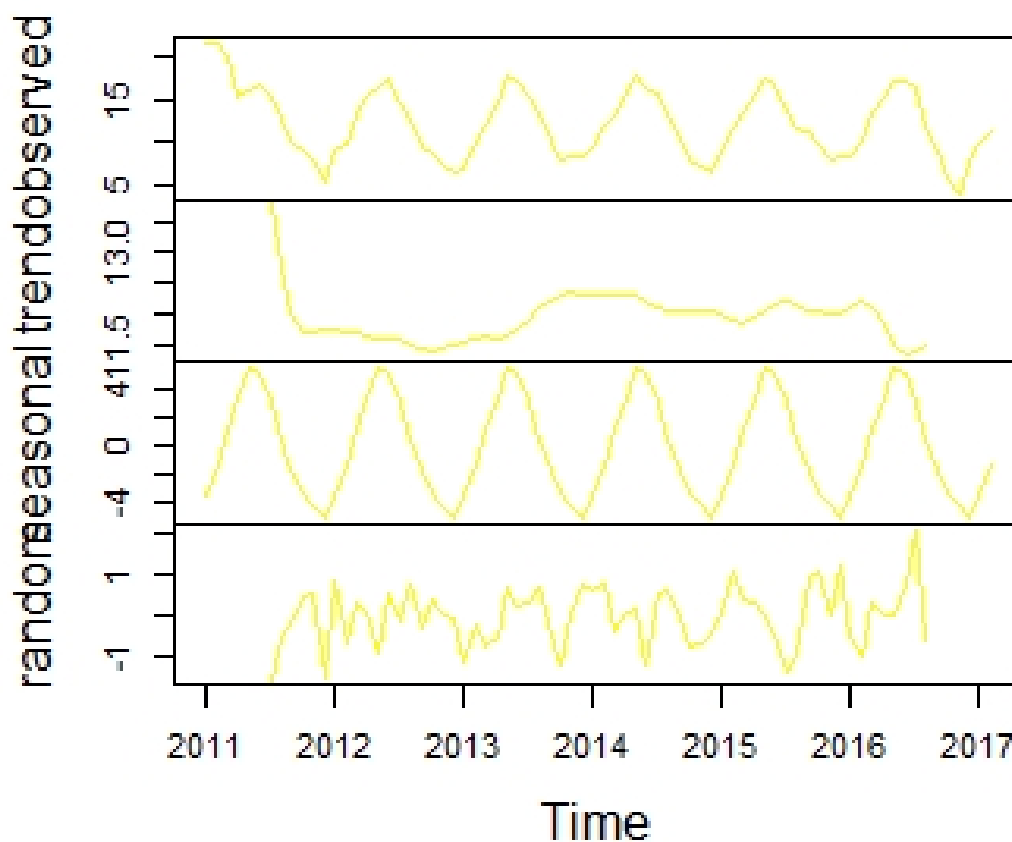


FIGURE 3.10 – Décomposition de la série Od5

L'allure décroissante de la courbe de la tendance "**trend**" pour les stations **812**, **813** et **816** indique une baisse de température sur l'ensemble de la période tandis que l'allure croissante de la courbe de la station **815** confirme l'augmentation des températures sur l'ensemble de la période. Quand à la saisonnalité, il subsiste ici une composante périodique de période 1 an. Et l'observation du graphe des résidus indique une variation croissante de ces derniers.

3.4 Corrélation des températures de l'Odon

Dans cette section, nous allons identifier les éventuelles corrélations entre nos séries.

Station 812 :

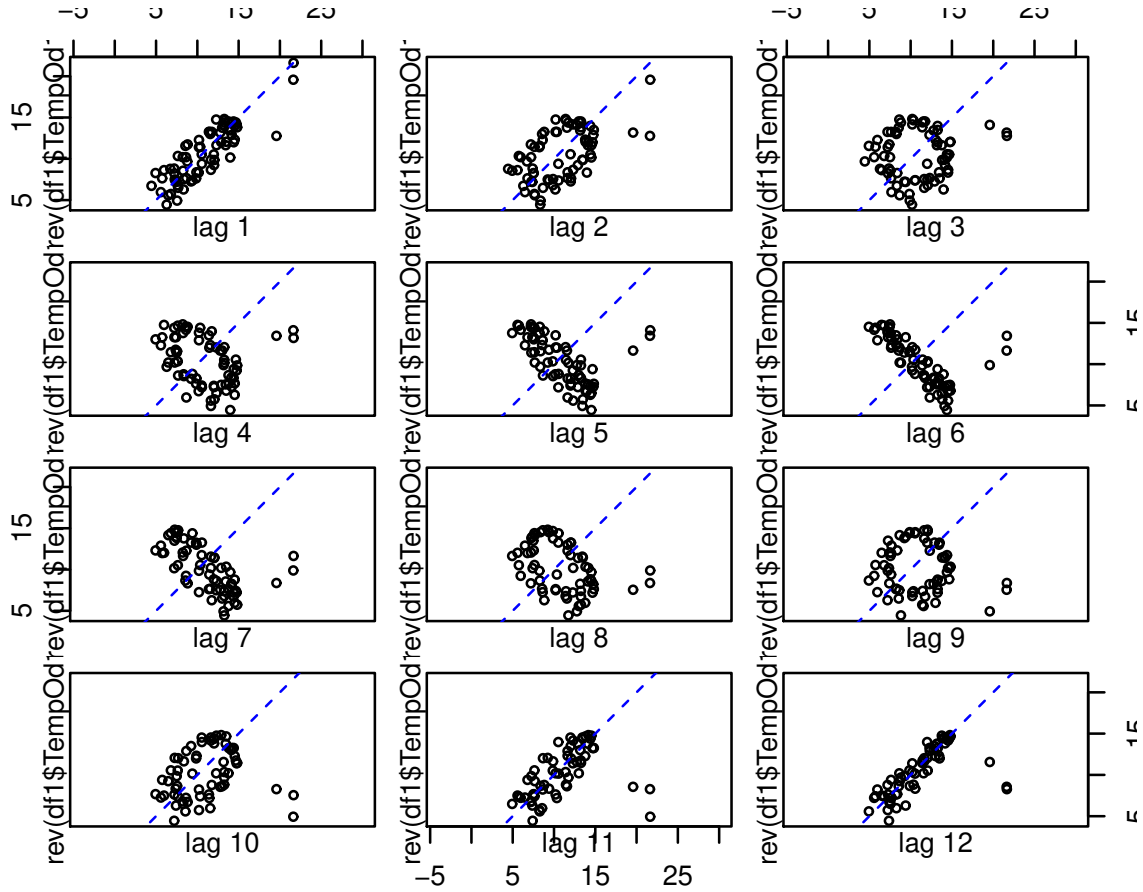


FIGURE 3.11 – Lag Plot Od1

Station 813 :

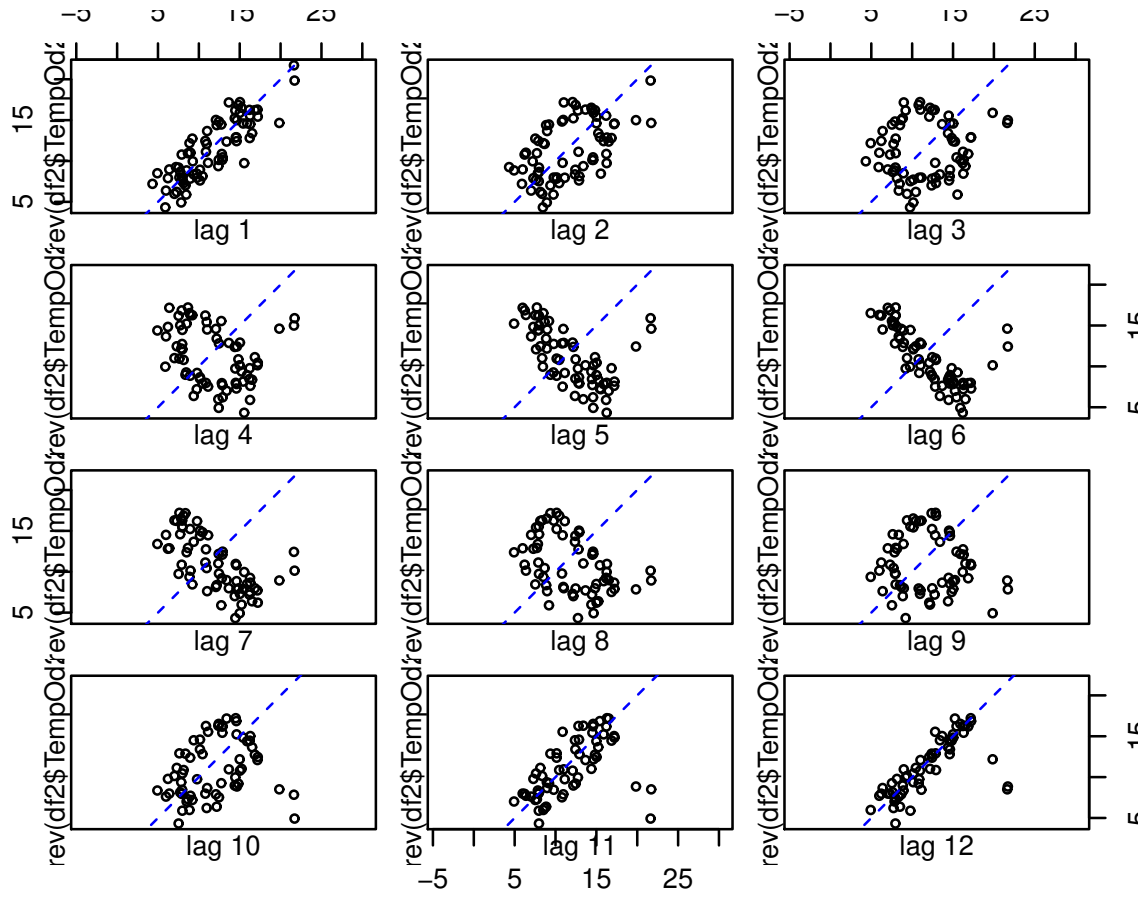
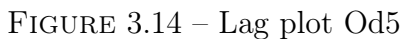


FIGURE 3.12 – Lag plot Od2



24

3.5 Saisonnalité mensuelle des températures de l'Odon

Dans cette section, nous allons visualiser graphiquement la saisonnalité mensuelle sur la période d'étude au niveau de chaque station.

Station 812 :

Saisonnalité de la température de l'Odon

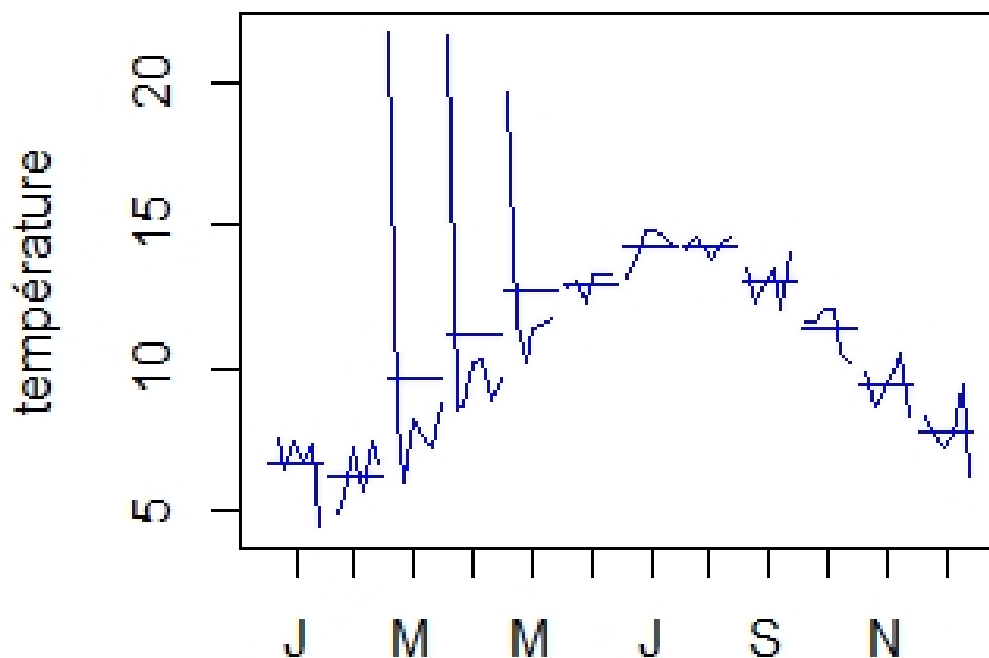


FIGURE 3.15 – Month Plot Od1

Station 813 :

Saisonnalité de la température de l'Odon

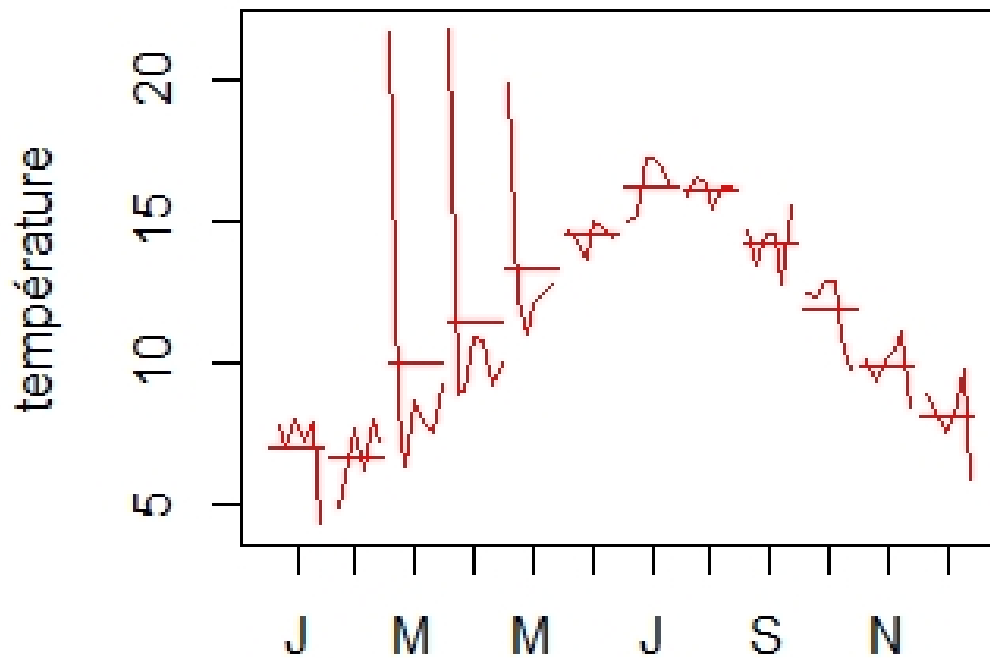


FIGURE 3.16 – Month Plot Od2

Station 815 :

Saisonnalité de la température de l'Odon

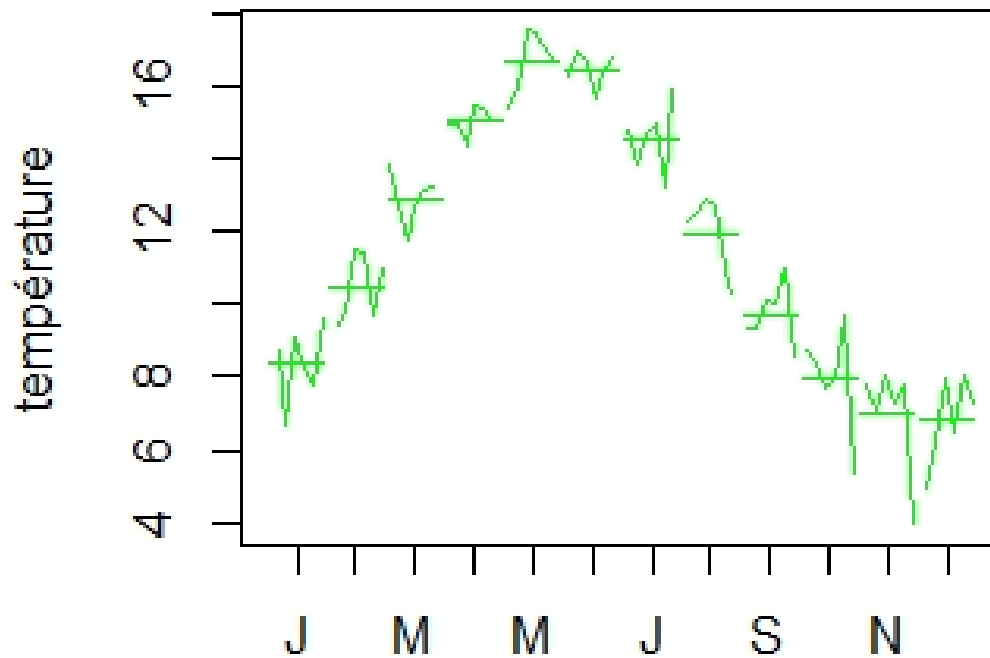


FIGURE 3.17 – Month Plot Od4

Station 816 :

Saisonnalité de la température de l'Odon

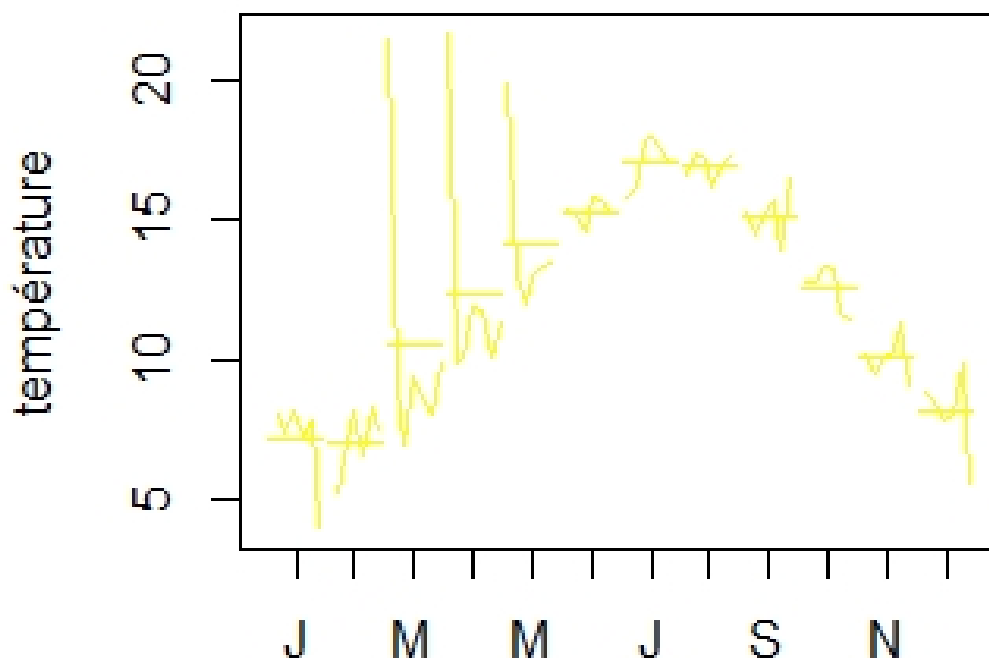


FIGURE 3.18 – Month Plot Od5

En observant le graphe des stations 812, 813 et 816, la température de l'Odon atteint son maximum en juillet et son minimum en décembre – janvier. Quant à la station 815, c'est au mois de mai et de novembre - décembre que les températures maximums et minimums sont respectivement observées.

Pour le mois de janvier par exemple, le trait horizontal représente la moyenne des températures du dit mois sur la période 2011 à 2017. Les autres traits pour janvier représentent la moyenne des années 2011 à 2017.

3.6 Analyse en Composante Indépendante

Dans cette section, nous allons étudier la température de l'eau des 4 sites de l'Odon situées en Normandie. Pour notre analyse, Nous avons choisi d'extraire 2 composantes, journalier et saisonnier. En effet, ce choix s'explique par le fait que la température d'un cour d'eau dépend de la température initiale de l'eau, et des échanges thermiques avec l'environnement. Autrement dit, nous considérons dans le cas de notre étude le moment t de la journée ou la collecte a été faite et la saison (hiver ou été) car ce dernier aspect détermine la température ambiante dans l'environnement.

3.6.1 Composante journalière

Nous allons créer une nouvelle base de données (2162 observations et 4 variables) dans laquelle on combine les quatre variables qui représentent les températures des quatre sites sans la colonne date. Rappelons que le package **fastICA** utilisé dans **r** pour nos analyses, ne fonctionne qu'avec des données numériques raison pour laquelle, dans notre nouvelle base, il n'y aura pas la date. L'objectif est d'extraire les composantes communes à nos 4 variables. Le code ci-dessous a été exécuté.

```
a <- fastICA(base_temp, 2, alg.typ = "parallel", fun = "logcosh", alpha = 1,
             method = "R", row.norm = FALSE, maxit = 200,
             tol = 0.0001, verbose = TRUE)
```

TABLE 3.5 – Matrice de passage

	[,1]	[,2]	[,4]	[,5]
[1,]	3.0697518	3.7006218	3.84969533	3.92495932
[2,]	-0.3459229	-0.4301612	0.01377717	0.07452123

Commentaires :

Les lignes du tableau ci-dessus correspondent aux différentes composantes et les colonnes aux variables. Par exemple, on a rentré dans la matrice tout d'abord la température journalière de la station 812, puis les températures journalières des stations 813 et 815 et enfin celle de la station 816. Ainsi le coefficient $[1,1]$ correspond à la première composante de la première variable.

Ci-dessous, la représentation graphique des composantes journalières et saisonnières des 4 sites.

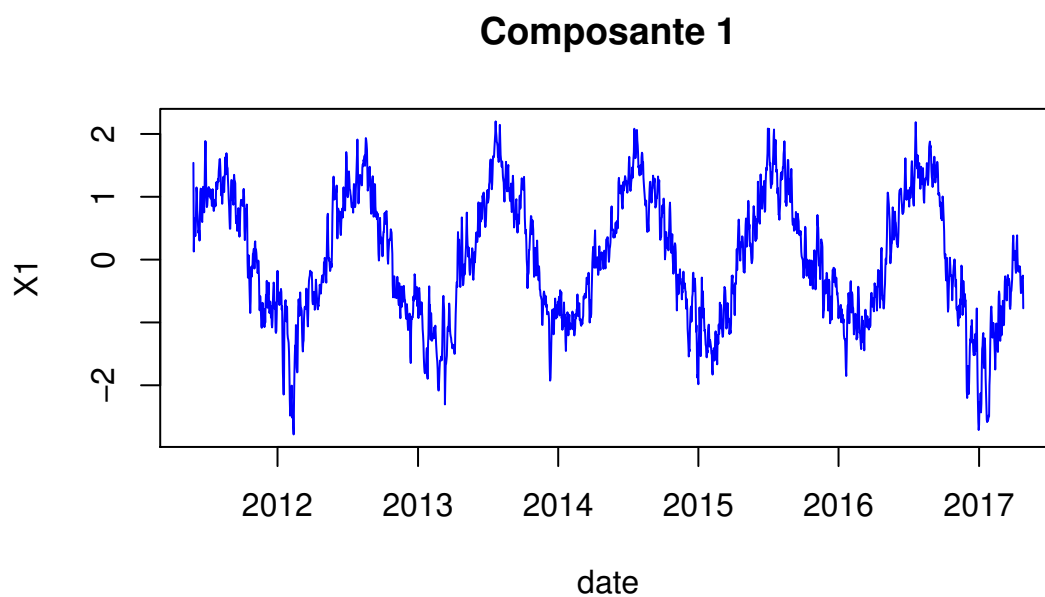


FIGURE 3.19 – Composantes Saisonnières des 4 sites

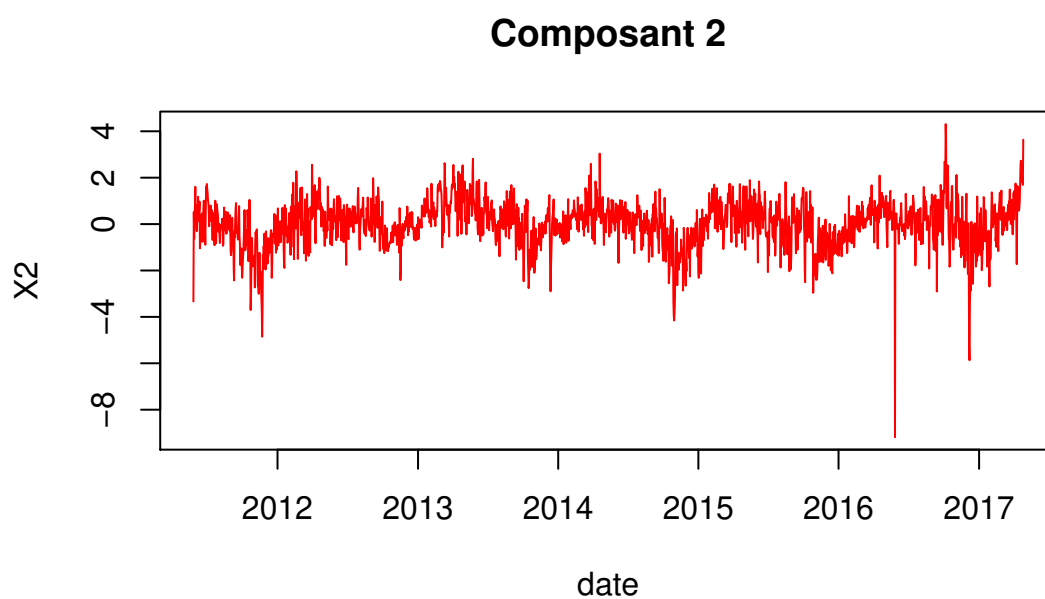


FIGURE 3.20 – Composantes Journalières des 4 sites

À partir des 2 graphes ci-dessus, nous observons que la température de l'Odon varie en fonction du jour et de la saison.

Par la suite, nous avons ci-dessous, la représentation des composantes C1 et C2 de la station 812.

Composante saisonnière station 812

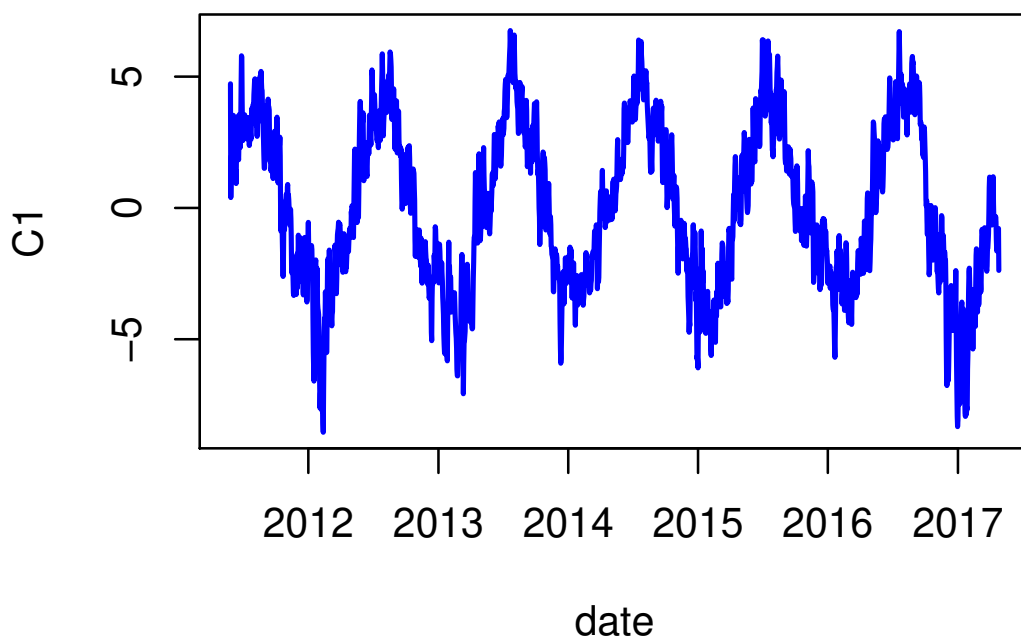
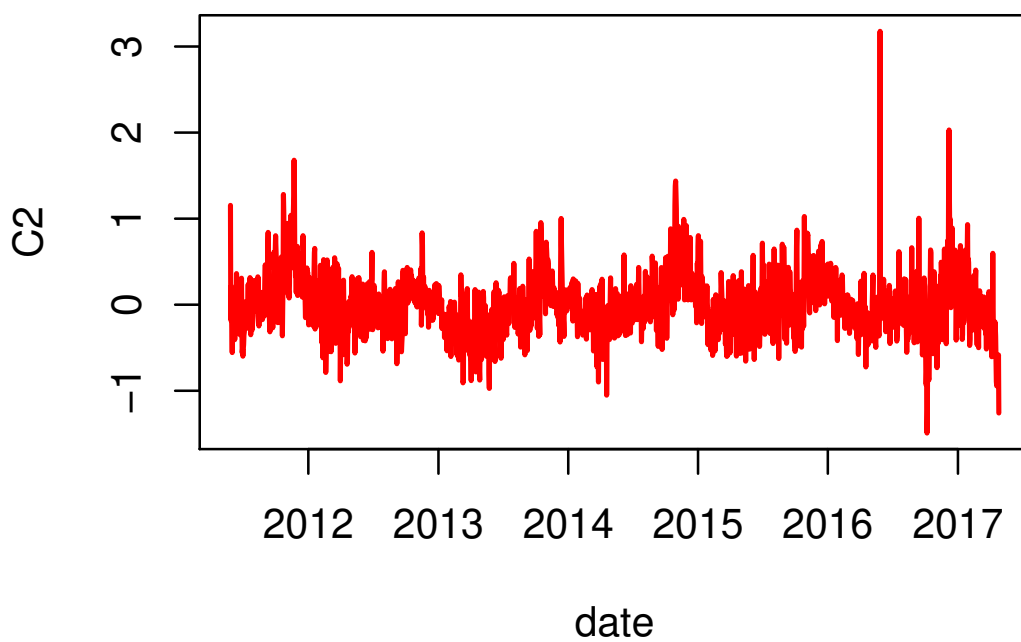


FIGURE 3.21 – Composante Saisonnière station 812

Composante journalière station 812



32
FIGURE 3.22 – Composante journalière station 812

À partir de ces données, nous pouvons reconstituer la données de départ en procédant comme suit.

```
B$Z=B$C1+B$C2
```

```
B$Z
```

```
B$A=mean(OdonT1ag$TempOd1)+B$Z
```

```
B$A
```

Chapitre 4

Application Shiny

Dans cette section, nous allons présenter l'application "*Web_Odon Analytics*" construite pour analyser les données de températures mesurées par les 4 différentes sondes (812, 813, 815 et 816).

Les images ci-dessous représentent les différents onglets de l'application.

- Import Data
- Import Map Data
- Interactive Analysis
- General Analysis Charts

L'Application *Web_Odon Analytics* est un tableau de bord (*Dashboard*).

Deux bases de données sont nécessaires pour faire fonctionner l'application correctement. La première, chargée depuis l'onglet **Import Data** comporte notre série chronologique et la seconde, les coordonnées géographiques des stations à afficher sur une carte.

L'utilisateur a la possibilité de voir s'afficher les 5 (10, 25,..., ou 100) premières lignes du jeu de données ainsi que le résumé statistique de la base de données chargée.

Dans **Import Map Data**, il s'affiche ainsi la localisation 4 différentes sondes et les données de location de chacune des sondes dans une table de données. Vous pouvez également sélectionner une sonde et zoomer jusqu'à obtenir une localisation précise du site que vous voulez visualiser.

L'onglet Interactif offre à l'utilisateur la possibilité de choisir la période d'analyse qu'il souhaite visualiser selon la station. Il s'affiche aussi le résumé statistique des températures recueillies sur les 4 stations et la table donnant une idée sur le jeu de données.

Ensuite, dans **l'onglet Interactif** onglet Interactive Analysis Charts, vous pouvez agir sur les données, notamment sur les dates pour obtenir les graphes qui vous permettront de faire des analyses très pertinentes et très poussées dans les détails du jeu de données.

Enfin **l'onglet General Analysis Charts** permet d'afficher les différents graphiques issus de l'analyse et que l'utilisateur a la possibilité de télécharger.

Ci-dessous, nous avons la présentation de l'application *Web_Odon Analytics* avec ses différents onglets :

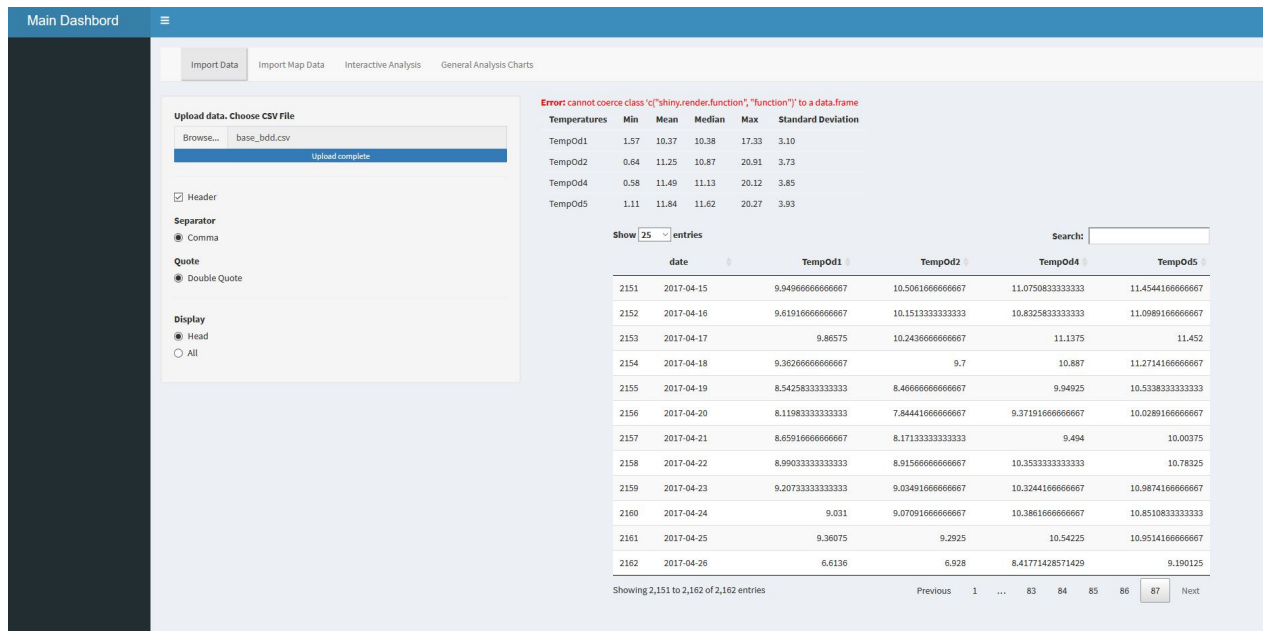


FIGURE 4.1 – **Import Data** : Onglet de chargement des données

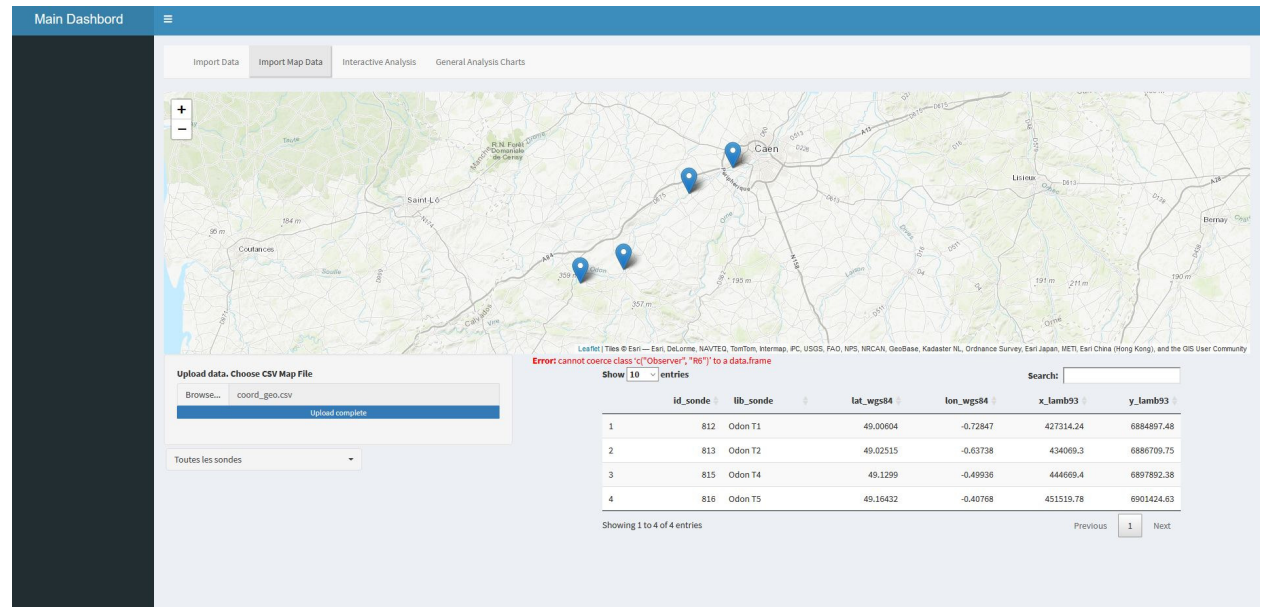


FIGURE 4.2 – **Import Map Data** : Onglet de chargement des coordonnées géographiques

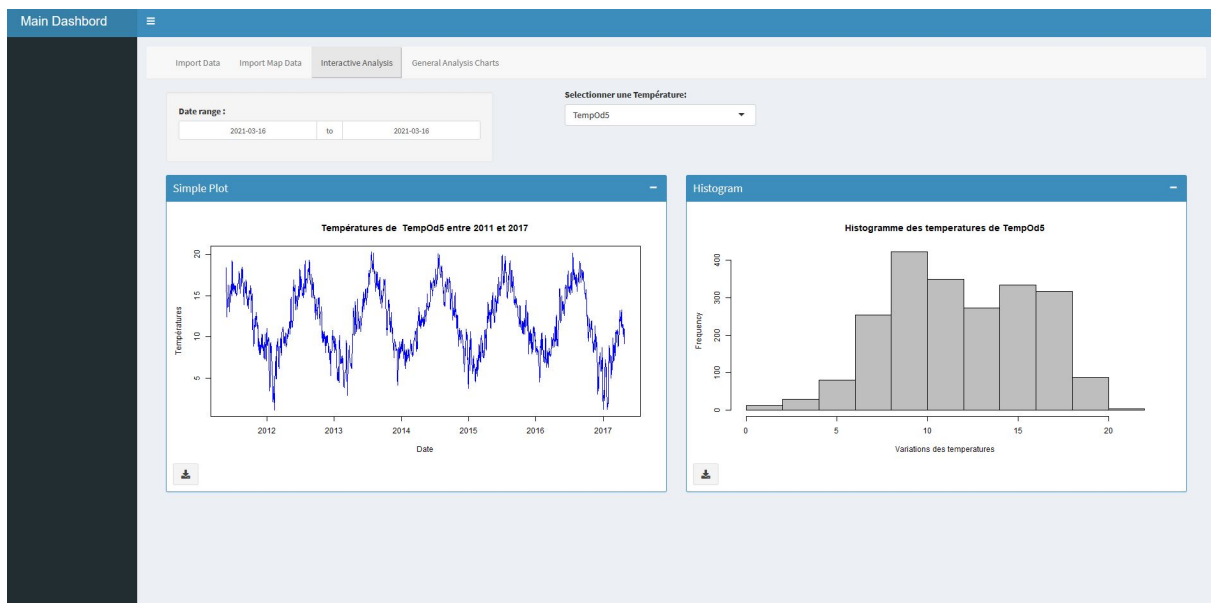


FIGURE 4.3 – **Interactive Analysis** : Onglet d'analyse Interactive

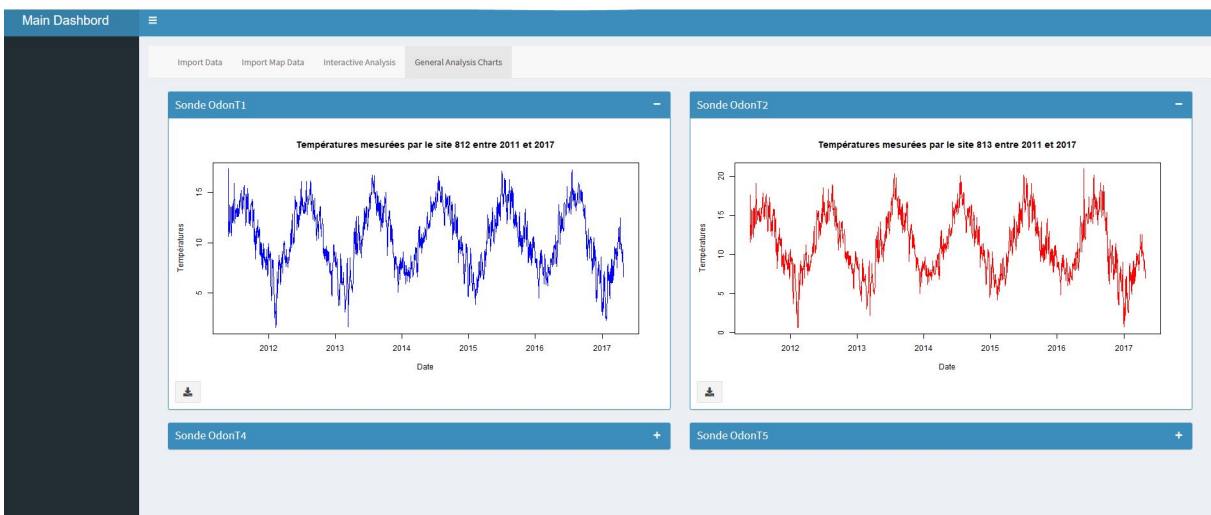


FIGURE 4.4 – **General Analysis Charts** : Onglet de résumé de graphes

Conclusion

Les analyses statistiques effectuées dans le cadre de cette étude ont portées sur les moyennes journalières des données collectées au niveau de chaque station. La température de l'Odon évolue en dent de scie sur la période 2011 à 2017. En général, les mois de juillet et août sont les plus chauds tandis que les mois de décembre, janvier et février sont les plus froids. Des valeurs extrêmes sont notées.

L'analyse en composante indépendantes effectuée à partir de l'algorithme FastICA nous a permis de mettre en évidence une composante journalière et une composante saisonnière. Cependant, il serait possible de mettre en évidence d'autres composantes comme par exemple l'effet d'un barrage, de l'air ou de la nappe phréatique. Au terme de l'analyse, une application shiny dénommée, Web_Odon Analytics, mise en oeuvre à partir du logiciel r a été construite. Son but est de fournir une analyse synthétique des données ainsi qu'une carte montrant la position des sondes. Cette application pourrait être améliorée.

Bibliographie

Liens Utilisés dans la réalisation du **rapport**

- https://professionnels.ofb.fr/sites/default/files/pdf/RE_Explore2070_Hydrologie_thermie.pdf
- http://www.sandre.eaufrance.fr/ftp/documents/fr/pre/atc/1/sandre_presentation_ATC_1.pdf
- <http://www.normandie.developpement-durable.gouv.fr/temperature-des-cours-d-eau-r1150.html>
- https://geosciences.univ-tours.fr/images/media/20150901113402-these_beaufort.pdf
- https://www.unb.ca/research/institutes/cri/_resources/pdfs/Benyahya2007.pdf
- <https://www.erudit.org/fr/revues/rseau/1995-v8-n1-rseau3278/705214ar.pdf>

Liens Utilisés dans l'aide de la création de l'application **Shiny (Web_Odon Analytics)**

- <https://shiny.rstudio.com/gallery/>
- <https://lmi.cnrs.fr/r/graphical-interfaces-with-shiny.html>
- <https://www.youtube.com/watch?v=LSnWGmVkB6A>
- https://www.htmlwidgets.org/showcase_leaflet.html
- <https://bookdown.org/weicheng/shinyTutorial/ui.html>
- <https://shiny.rstudio.com/gallery/sliders.html>
- <https://rstudio.github.io/dygraphs/>
- <https://www.javaer101.com/en/article/12782482.html>
- <https://stackoverflow.com/questions/49215000/synchronise-dygraph-and-daterangeinput->

Annexe A

Code

```
[language=rstudio]
#####
#      Sujet: Etude des températures des cours d'eau
#####

# Chargement des libraries
library(lubridate)
library(fastICA)
library(dplyr)

# Recuperation des Coordonnées géographiques -----
# library(dplyr)
#
# file_coord="C:/Users/mamad/Dropbox/UE_Prorjet_Partag/base/l-reseautemprivieres-d-r28.csv"
# cordonnes <- read.csv2(file_coord,header = T,encoding = 'UTF-8', sep = ",",stringsAsFactors = F)
# attach(cordonnes)
#
# #Recuperations des coordonnées des 4 stations (T1,T2,T4 et T5)
# code<-cordonnes %>%
#   select(id_sonde,lib_sonde,lat_wgs84,lon_wgs84,x_lamb93,y_lamb93) %>%
#   filter(id_sonde %in% c("812","813","815","816"))
#
# #Exportations de la table de coordonnees
# write.csv(code,file= 'coord_geo.csv',append = FALSE, quote = TRUE, sep = ",")

#-----
# Sources de donnees
file1="C:/Users/mamad/Dropbox/UE_Prorjet_Partag/base/OdonT1 812_2011-2017.csv"
file2="C:/Users/mamad/Dropbox/UE_Prorjet_Partag/base/OdonT2 813_2011-2017.csv"
file4="C:/Users/mamad/Dropbox/UE_Prorjet_Partag/base/OdonT4 815_2011-2017.csv"
file5="C:/Users/mamad/Dropbox/UE_Prorjet_Partag/base/OdonT5 816_2011-2017.csv"
#-----

#I) Importations, Nettoyages et Validations des donnees
```

```

#####
#
#                               Station OdonT1
#####
OdonT1 <- read.csv2(file1,header = FALSE,encoding = 'UTF-8', sep = ",",stringsAsFactors = F)
attach(OdonT1)

OdonT1<-OdonT1[-1:-2,,]  #Suppression de lignes 1 et 2 dans les données
colnames(OdonT1)<-c("Indice","date","Heure","TempOd1")  #Renommer l'entête de la table
OdonT1
head(OdonT1)

#Recherche de Valeurs manquantes (NA) dans OdonT1
library(dplyr)
apply(OdonT1, MARGIN = 2, FUN = function(x){x%>%is.na%>%sum}) #Rech par colonne
#Il n'ya pas de valeurs manquantes dans la table OdonT1

#-- Mise en format des données
OdonT1$Indice=as.integer(OdonT1$Indice)
OdonT1$TempOd1=as.double(OdonT1$TempOd1)

#Transformation de la date et l'heure
library(lubridate)
OdonT1$t <- paste(OdonT1$date, OdonT1$Heure)
OdonT1$t<-dmy_hms(OdonT1$t)
OdonT1$date=date(OdonT1$t)

#Statistique descriptives
summary(OdonT1)

#####
#
#                               Station OdonT2
#####
OdonT2 <- read.csv2(file2,header = FALSE,encoding = 'UTF-8', sep = ",",stringsAsFactors = F)
attach(OdonT2)
OdonT2<-OdonT2[-1:-2,,]  #Suppression de lignes 1 dans les données
colnames(OdonT2)<-c("Indice","date","Heure","TempOd2")  #Renommer l'entête de la table
OdonT2

#Recherche de Valeurs manquantes (NA) dans OdonT2
library(dplyr)
apply(OdonT2, MARGIN = 2, FUN = function(x){x%>%is.na%>%sum})
#Il n'ya pas de valeurs manquantes dans la table OdonT2

#-- Mise en format des données
OdonT2$Indice=as.integer(OdonT2$Indice)
OdonT2$TempOd2=as.double(OdonT2$TempOd2)

```

```

#Transformation de la date et l'heure
library(lubridate)
OdonT2$t <- paste(OdonT2$date, OdonT2$Heure)
OdonT2$t<-dmy_hms(OdonT2$t)
OdonT2$date=date(OdonT2$t)

#Statistique descriptives
summary(OdonT2)

#####
#                               Station OdonT4
#####
OdonT4 <- read.csv2(file4,header = FALSE,encoding = 'UTF-8', sep = ",",stringsAsFactors = F)
attach(OdonT4)
OdonT4<-OdonT4[-1:-2,,]  #Suppression de lignes 1 dans les données
colnames(OdonT4)<-c("Indice","date","Heure","TempOd4")  #Renommer l'entête de la table
OdonT4

#Recherche de Valeurs manquantes (NA) dans OdonT4
library(dplyr)
apply(OdonT4, MARGIN = 2, FUN = function(x){x%>%is.na%>%sum})
#Il n'ya pas de valeurs manquantes dans las table OdonT4

#-- Mise en format des données
OdonT4$Indice=as.integer(OdonT4$Indice)
OdonT4$TempOd4=as.double(OdonT4$TempOd4)

#Transformation de la date et l'heure
library(lubridate)
OdonT4$t <- paste(OdonT4$date, OdonT4$Heure)
OdonT4$t<-dmy_hms(OdonT4$t)
OdonT4$date=date(OdonT4$t)

#Statistique descriptives
summary(OdonT4)

#####
#                               Station OdonT5
#####
OdonT5 <- read.csv2(file5,header = FALSE,encoding = 'UTF-8', sep = ",",stringsAsFactors = F)
attach(OdonT5)
OdonT5<-OdonT5[-1:-2,,]  #Suppression de lignes 1 et dans les données
colnames(OdonT5)<-c("Indice","date","Heure","TempOd5")  #Renommer l'entête de la table
OdonT5

```

```

#Recherche de Valeurs manquantes (NA) dans OdonT5
library(dplyr)
apply(OdonT5, MARGIN = 2, FUN = function(x){x%>%is.na%>%sum})
#Il n'ya pas de valeurs manquantes dans las table OdonT5

#-- Mise en format des données
OdonT5$Indice=as.integer(OdonT5$Indice)
OdonT5$TempOd5=as.double(OdonT5$TempOd5)

#Transformation de la date et l'heure
library(lubridate)
OdonT5$t <- paste(OdonT5$date, OdonT5$Heure)
OdonT5$t<-dmy_hms(OdonT5$t)
OdonT5$date=date(OdonT5$t)

#Statistique descriptives
summary(OdonT5)

*****
# Merge des tables -----
*****
library(rlang)
library(tidyverse)
library(lubridate)
library(dplyr)

# Moyenne aggregate par jour
OdonT1ag<-aggregate(TempOd1~date, data=OdonT1, FUN = mean)
OdonT2ag<-aggregate(TempOd2~date, data=OdonT2, FUN = mean, na.rm=TRUE)
OdonT4ag<-aggregate(TempOd4~date, data=OdonT4, FUN = mean, na.rm=TRUE)
OdonT5ag<-aggregate(TempOd5~date, data=OdonT5, FUN = mean, na.rm=TRUE)

#Merge des différentes bases de données (avec date et l'heure)
#pour obtenir une seule base de données
fusion1<-merge(OdonT1ag,OdonT2ag, by="date")
fusion2<-merge(fusion1,OdonT4ag, by="date")
base<-merge(fusion2,OdonT5ag, by="date")    #base de données principale

#base: Ce fichier contiendra les variables suivantes:
#date, TempOd1, TempOd2, TempOd4 et TempOd5

#----- Exporter la base de donnees-----
#Celle-ci constituera le fichier (CSV data.file ) qu'on utilisera dans l'application
# write.csv(base,file= 'base_bdd.csv',append = FALSE, quote = TRUE, sep = ",")

*****
#          1°) Analyses descriptives des sites(Stations)-----

```

```

#####
par(mfrow=c(2,2))
# Traçage de le courbe des températures -----
plot(Temp0d1~date, type="l", data=OdonT1ag, col="red", xlab="Date", ylab = "Températures",m
plot(Temp0d2~date, type="l", data=OdonT2ag, col="blue", xlab="Date", ylab = "Températures"
plot(Temp0d4~date, type="l", data=OdonT4ag, col="green", xlab="Date", ylab = "Températures"
plot(Temp0d5~date, type="l", data=OdonT5ag, col="yellow", xlab="Date", ylab = "Températures"

#Summary des données
summary(OdonT1ag$Temp0d1)
summary(OdonT2ag$Temp0d2)
summary(OdonT4ag$Temp0d4)
summary(OdonT5ag$Temp0d5)
summary(base) #résumé stats sur la base principale

#----- Autre methode de calcul du summary -----
# library(dplyr)
# base %>%
#   summarise_each(funs(min,median,mean,max, sd), Temp0d1,
#   Temp0d2, Temp0d4,Temp0d5)

#----- Histogrammes -----
par(mfrow=c(2,2))
#Histogramme OdonT1
hist(OdonT1ag$Temp0d1,col='blue',xlab = 'Température', main="Histogramme des températures d
abline(v=mean(OdonT1ag$Temp0d1),col='black',lty=3 , lwd = 3)
text(mean(OdonT1ag$Temp0d1),y=6.5,paste("La temp moyenne est égale à 10.66 "))

#Histogramme OdonT2
hist(OdonT2ag$Temp0d2,col='red',xlab = 'Température', main="Histogramme des températures d
abline(v=mean(OdonT2ag$Temp0d2),col='black',lty=3 , lwd = 3)
text(mean(OdonT2ag$Temp0d2),y=6.5,paste("La temp moyenne est égale à 11.51 "))

#Histogramme OdonT4
hist(OdonT4ag$Temp0d4,col='red',xlab = 'Température', main="Histogramme des températures d
abline(v=mean(OdonT4ag$Temp0d4),col='black',lty=3 , lwd = 3)
text(mean(OdonT4ag$Temp0d4),y=6.5,paste("La temp moyenne est égale à 11.48 "))

#Histogramme OdonT5
hist(OdonT5ag$Temp0d5,col='blue',xlab = 'Température', main="Histogramme des températures d
abline(v=mean(OdonT5ag$Temp0d5),col='black',lty=3 , lwd = 3)
text(mean(OdonT5ag$Temp0d5),y=6.5,paste("La temp moyenne est égale à 12.08 "))

#####
#           2°) Décomposition en séries temporelles
#####
library(lubridate)

```

```

OdonT1ag$year=year(OdonT1ag$date)
OdonT2ag$year=year(OdonT2ag$date)
OdonT4ag$year=year(OdonT4ag$date)
OdonT5ag$year=year(OdonT5ag$date)

#Transformation de la date en format journalier

#####
#                               3°) ACI: Analyse en composantes indépendantes
#####
#Composantes journalières
library(lubridate)
OdonT1$date=ymd(OdonT1$date)
OdonT2$date=ymd(OdonT2$date)
OdonT4$date=ymd(OdonT4$date)
OdonT5$date=ymd(OdonT5$date)

# Moyenne aggregate par jour (Tendance par jour)
OdonT1ag<-aggregate(TempOd1~date, data=OdonT1, FUN = mean)
OdonT2ag<-aggregate(TempOd2~date, data=OdonT2, FUN = mean, na.rm=TRUE)
OdonT4ag<-aggregate(TempOd4~date, data=OdonT4, FUN = mean, na.rm=TRUE)
OdonT5ag<-aggregate(TempOd5~date, data=OdonT5, FUN = mean, na.rm=TRUE)

#Merge des différentes bases de données (avec date et l'heure)
#pour obtenir une seule base de données
fusion1<-merge(OdonT1ag,OdonT2ag, by="date")
fusion2<-merge(fusion1,OdonT4ag, by="date")
base<-merge(fusion2,OdonT5ag, by="date")    #base de données principale

base2<-select(base,-date) #base avec temperature et sans la date
base3<-select(base,date)  #table contenant uniquement les dates

summary(base2)

#-----
#                               Applications de la FastICA
#-----
set.seed(1)
#Suivant journalière et saisonnière
a <- fastICA(base2, 2, alg.typ = "parallel", fun = "logcosh", alpha = 1,
             method = "R", row.norm = FALSE, maxit = 200,
             tol = 0.0001, verbose = TRUE)

#2 Matrices passage A et des sources S
a$A
#Cette matrice nous permet d'obtenir les resultats respectivement suivant chaque ligne
#la composantes saisonniere et la composante journaliere

```

```
#Et les 4 colonnes representent les 4 stations

a$S
matA<-data.frame(a$S) #Création d'une dataframe
matB<-cbind(base3, matA) #création d'une matrice B avec les dates

#3 représentation des 2 signaux
par(mfrow = c(1,1))
plot(X1~date, type="l", col="blue", data=matB)
plot(X2~date, type="l", col="red", data=matB)

matB$comp1=a$A[1,1]*a$S[,1]
matB$comp2=a$A[2,1]*a$S[,2]

#Representation des 2 composantes de la première variable (OdonT1)
par(mfrow = c(1,2))
plot(comp1~date, type="l", data=matB, col="blue")
plot(comp2~date, type="l", data=matB, col="red")

matB$Z=matB$comp1+matB$comp2+mean(OdonT1ag$TempOd1)
matB$Z
```