

UNIVERSITÉ DE CAEN

MASTER 1 SAAD - INFORMATIQUE

2021-2022

ANALYSE DE DONNÉES

Rapport d'analyse de données

ANALYSE DE LA TOUQUES

Antoine BOITEAU

Deann ANTONOV

Paul GOSSE

Valentin LEROUGE



UNIVERSITÉ
CAEN
NORMANDIE

Table des matières

1	Introduction	2
2	Représentation simple des données	2
3	Analyse en composantes indépendantes (ACI)	3
3.1	Les influences annuelles sur la température de l'eau	3
3.2	L'influence journalière sur la température de l'eau	3
4	Analyse en composantes principales (ACP)	5
5	Regression linéaire	6
6	Limites de la base étudiée	7
7	Conclusion	8

1 Introduction

La Touques est un fleuve de 108km qui prend sa source dans l'Orne, traverse le Calvados du Sud au Nord et se jette dans la Manche. Avec un nombre de Strahler de 5, on peut considérer ce fleuve comme de moindre envergure.

Située dans une zone géographique dite "océanique", la Touques traverse principalement des plateaux argileux Normands et des terres possédant des nappes phréatiques assez profondes.

Dans le cadre d'une analyse des données de relevés de 4 sondes immergées dans ce fleuve, du 29 Mai 2013 au 05 Octobre 2018, nous essaierons de répondre à la problématique suivante avec les méthodes statistiques vues en cours :

Quelles sont les facteurs qui influent sur la température de l'eau dans la Touques ?

2 Représentation simple des données

En prémisses des différentes analyses effectuées, il nous a fallu nous familiariser avec le jeu de données utilisé. Pour cela nous avons réalisé différents graphiques afin d'avoir une première vue des variables et leurs intérêts au travers de l'étude. Voici quelques exemples de graphiques que nous avons pu effectuer dans ce but.

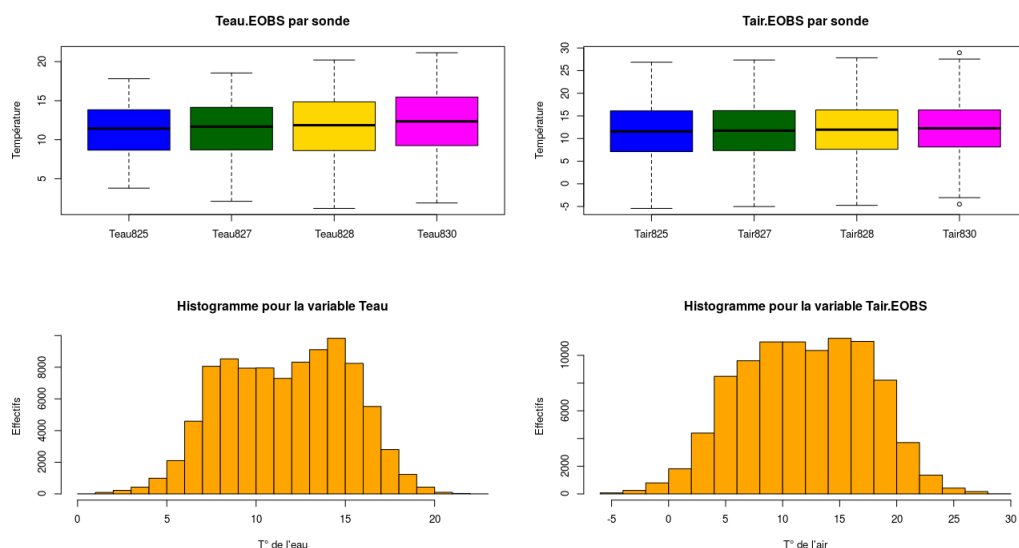


FIGURE 1 – Différentes représentations de la donnée

Une première analyse en visualisant les boxplots nous a permis d'estimer que la température de l'eau est plus chaude en aval du fleuve qu'en amont. Rappelons en effet que la sonde 825 est placée en amont du fleuve tandis que la sonde 830 est placée en aval de ce dernier. La température montant légèrement de sonde en sonde, nous pouvons en déduire que la température de l'eau a tendance à se réchauffer en s'éloignant de l'amont. Il semble de plus y avoir une augmentation de la dispersion des températures à mesure que la sonde est en aval du fleuve.

On peut s'attendre à une forte relation linéaire entre la température de l'eau et la température de l'air, du fait de la ressemblance des histogrammes. Malgré une distance relativement importante entre les deux sondes 825 et 830 (de l'ordre de 100 km), la moyenne de température ne semble connaître qu'une faible augmentation.

3 Analyse en composantes indépendantes (ACI)

Dans le but de trouver des signaux sources cachés dans les données des températures de l'eau relevées par les 4 sondes 825, 827, 828 et 830, nous avons réalisé plusieurs ACI.

3.1 Les influences annuelles sur la température de l'eau

Le jeu de données étant très grand, avec 12 relevés par jour et par sonde, nous avons tout d'abord réduit ce dernier en travaillant sur des moyennes de *Teau* journalières par sonde (réduisant ainsi par 12 le nombre de lignes de notre base de travail).

En séparant en 4 séries temporelles les relevés par sonde nous avons pu lancer une ACI sur ces dernières grâce au package *fastICA* de R.

La première ACI effectuée avait pour but de chercher 4 signaux distincts, néanmoins les résultats n'étaient pas concluants avec seulement deux composantes qui paraissaient sinusoïdales périodiques et 2 autres qui ne semblaient être que du bruit. Afin d'affiner notre analyse nous avons donc réduit le nombre de signaux recherchés à 3 pour obtenir les composantes de la figure 2.

On retrouve encore une fois une première composante qui paraît peu exploitable. En effet elle ne semble indiquer une forte variation qu'au début de l'année 2015 mais nous n'avons pas trouvé dans nos recherches une raison permettant de l'expliquer.

Néanmoins les deux autres signaux des composantes 2 et 3 sont bien plus intéressants. Ces deux signaux sinusoïdaux périodiques annuels démontrent deux influences déphasées :

- **la composante 3** propose des variations en phase avec le cycle naturel des saisons, tous les ans un pic est atteint au coeur de l'été pendant les mois de Juillet et Août et la température diminue vers un minimum annuel au plus froid de l'hiver en Janvier/Février. Cette composante peut donc décrire l'impact des saisons sur la température du cours d'eau de la Touques.
- **la composante 2** quand à elle montre une courbe annuelle assez similaire quoique décalée, elle atteint ses maximums en début/milieu d'automne et ses minimums locaux au printemps. Or la Touques se trouve près de nappes phréatiques profondes qui peuvent influencer sur la température des "moyens et petits cours d'eau" selon un *rapport de l'Université de Tours de 2015* par A. Beaufort et F. Moatar. Ainsi il semble probable que ce signal proviennent de ces nappes phréatiques à proximité de la Touques.

3.2 L'influence journalière sur la température de l'eau

Lors de cette première ACI nous utilisons des moyennes de *Teau* journalières, nous avons pu grâce à cela déterminer des phénomènes d'ampleur annuelle. Nous cherchons maintenant à déterminer si des phénomènes à l'échelle des journées s'appliquent aussi.

Pour cela nous avons préparé de nouvelles colonnes de la base de la manière suivante : Pour chacune des 4 sondes, nous avons calculé la différence entre la température enregistrée à l'instant t (un relevé toutes les 2h) et la température moyenne de la journée pour cette sonde.

Suite à cela, et avec nos 4 séries temporelles préparées nous avons de nouveau utilisé la méthode de l'ACI. En cherchant uniquement 2 signaux nous obtenons le résultat de la figure 3.

- **la composante 1** nous montre des variations importantes début 2014 et 2018, mais ne nous permet pas de déceler un effet de cycle ou de conclure clairement sur sa source, nous ne parvenons pas à l'utiliser.
- **la composante 2** quand à elle propose bien plus clairement un cycle annuelle de variations, ainsi on remarque que les différences de *Teau* à sa moyenne journalière sont bien plus élevées pendant les périodes estivales, à l'inverse on a peu de variations de températures dans la journée lors des hivers.

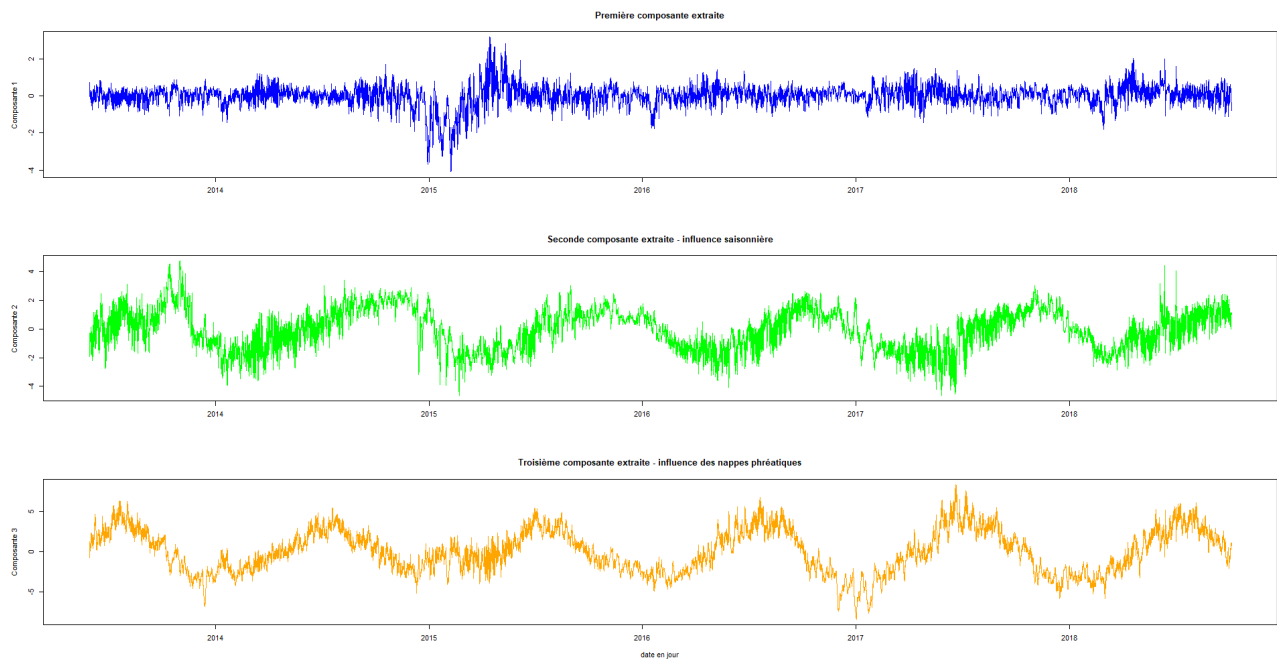


FIGURE 2 – Les 3 composantes identifiées lors de la première ACI sur les *Teau*

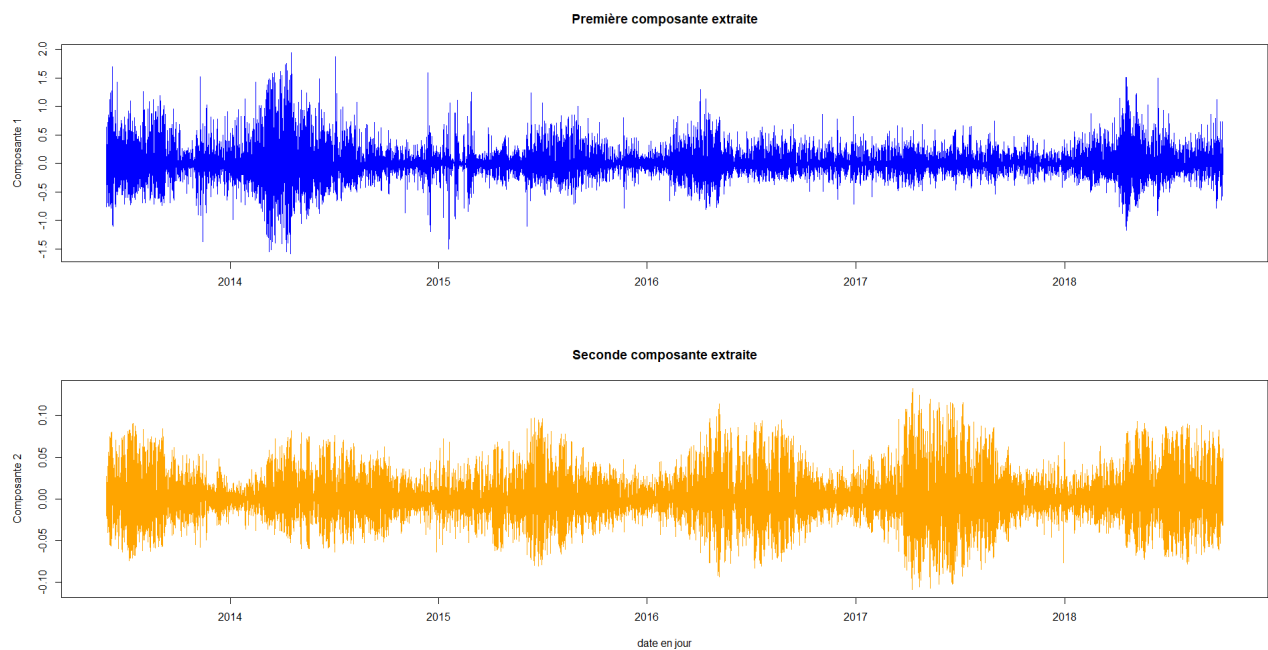


FIGURE 3 – Les 2 composantes résultantes d'une ACI cherchant des effets journaliers sur *Teau*

On peut conclure de ces deux ACI que les périodes estivales ont un double effet, à la fois elles réchauffent la température de l'eau de la Touques, mais augmentent aussi les variations de températures dans la journée. En outre les nappes phréatiques environnant ce cours d'eau ont aussi un impact sur la température de ce petit fleuve.

Néanmoins il reste des signaux que nous n'avons pas su interpréter ainsi que de nombreuses variables que nous n'avons pas encore étudiées. Voyons donc ensemble ce que le croisement de ces variables peut nous apporter dans la suite de ce rapport.

4 Analyse en composantes principales (ACP)

Afin de pouvoir visualiser plus facilement les corrélations entre variables, on procède à une analyse en composantes principales. On effectuera une ACP pour chaque sonde. Un vecteur *temps* a été créé pour pouvoir mettre en évidence une éventuelle augmentation de la température au cours du temps. On note *Teau* la température de l'eau, *Tair* celle de l'air et *Rainf* la pluviométrie.

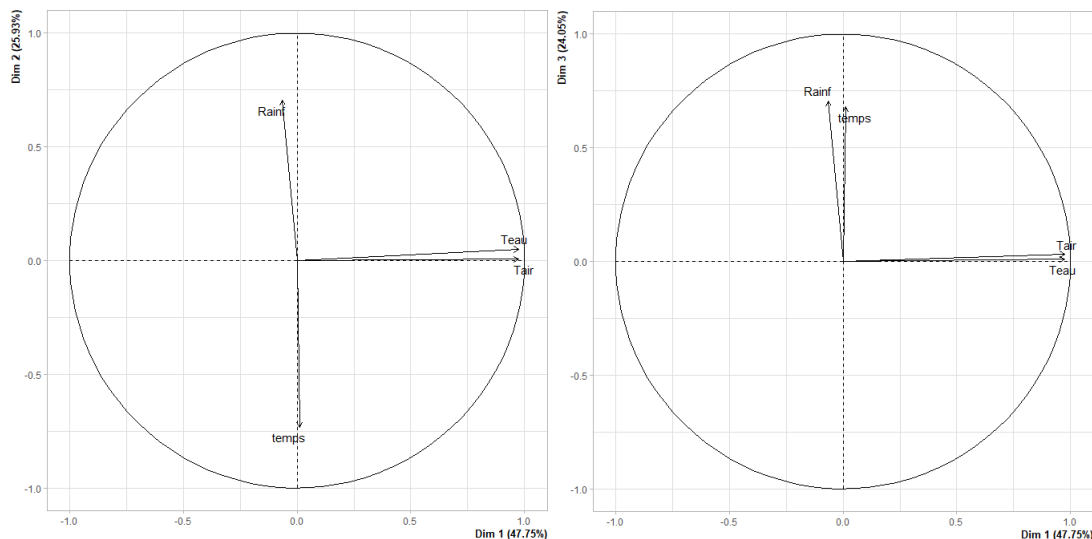


FIGURE 4 – Cercle des corrélations pour la sonde 825

Les variables *Teau* et *Tair* sont fortement corrélées de façon positive et contribuent grandement au premier axe. Le temps et la pluviométrie semblent indépendants des températures. La faible qualité de représentation des variables *temps* et *Rainf* les fait apparaître corrélées tantôt positivement, tantôt négativement. Ces deux variables seraient mieux représentées par les axes 2 et 3 mais on perdrait alors de l'information sur *Teau*, *Tair* et on perdrait une grande partie de la variabilité totale : 47.7% de la variance totale est en effet représentée par la dimension 1. L'ACP sur les mesures réalisées par les autres sondes conduit aux mêmes conclusions.

Il peut être intéressant de réaliser une autre ACP avec les composantes indépendantes extraites de l'ACI sur les valeurs de *Teau*. On présente une nouvelle fois les résultats de la sonde 825, mais les conclusions sont les mêmes pour les autres sondes.

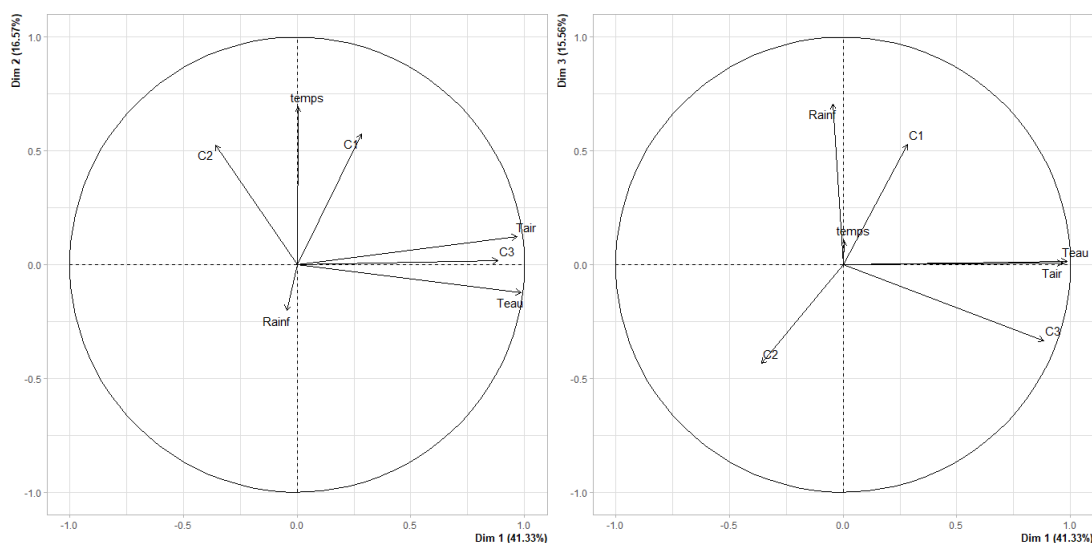


FIGURE 5 – Cercle des corrélations avec composantes de l'ACI

On observe une forte corrélation positive entre *Teau*, *Tair* et la troisième composante indépendante

C3. Cela n'est pas surprenant puisque cette dernière prend des valeurs élevées en été et basses en hiver. On s'attend au même genre de variation pour les températures de l'eau et de l'air, d'où une corrélation élevée.

Précisons également que si certaines composantes de l'ACI n'apparaissent pas comme indépendantes, cela n'est dû qu'à la faible qualité de leur représentation par l'ACP.

5 Régression linéaire

Suite aux résultats observés précédemment lors de l'ACP, nous avons décidé d'effectuer une régression linéaire pour visualiser plus précisément la relation entre Teau et Tair.EOBS. Pour cela, nous avons tout d'abord calculé le coefficient de Pearson, qui permet d'exprimer la force et la nature (croissante ou décroissante) de la relation linéaire entre deux variables quantitatives. Nous avons calculé ce coefficient pour les trois relations suivantes :

- Teau - Tair.EOBS
- Teau - Rainf.EOBS
- Tair.EOBS - Rainf.EOBS

Avec un résultat à 0.922 pour la relation Teau - Tair.EOBS, nous avons décidé de réaliser une régression linéaire simple pour cette relation. Les deux autres relations, ayant un coefficient de Pearson proche de 0, nous indiquent qu'il n'existe pas de corrélations directes entre la pluviométrie et les températures de l'eau et de l'air. Cependant il faut être vigilant, car il peut quand même exister une influence entre ces variables.

Pour s'assurer de la qualité du modèle de régression, nous avons décidé de calculer le coefficient de détermination R^2 .

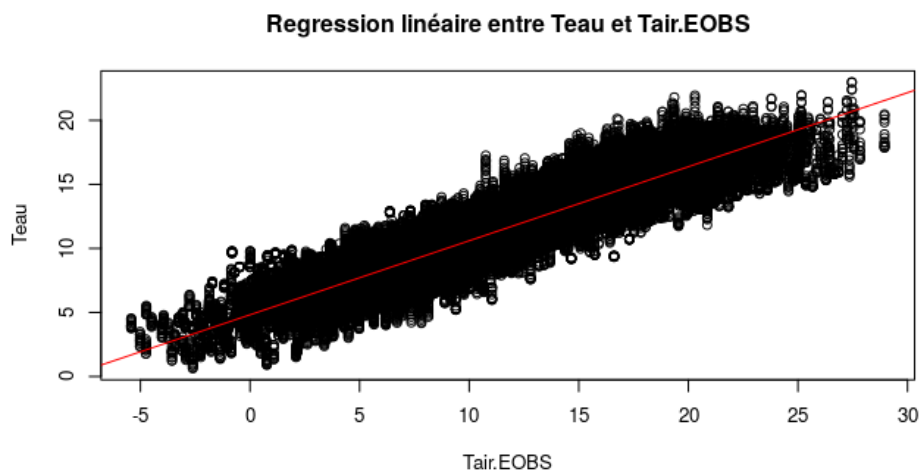


FIGURE 6 – Graphique de regression

Après étude de la relation grâce à la commande suivante :

```
mylm=summary(lm(Teau Tair.EOBS, data=base))
```

Nous pouvons dire que :

- La p-value du test de Pearson (unilatéral à droite) est très proche de 0. Le lien linéaire croissant entre Teau et Tair.EOBS est donc hautement significatif.
- La valeur de pente est estimée à 0.57, la température de l'air a donc un effet positif non négligeable sur la température de l'eau.
- Le R^2 est égal à 0.85, la qualité d'ajustement est donc très forte. En conséquence il devient possible de prédire de façon satisfaisante la température de l'eau à partir de celle de l'air.

Nous avons donc réalisé deux prédictions : Une première se basant sur une température de l'air fixée à 30 degrés et une seconde se basant sur la température maximale mesurée par les sondes. Sur un intervalle de confiance à 99%, on estime que si Tair.EOBS est à 30 degrés, Teau sera entre 18.716 et

25.60 degrés avec une moyenne avoisinant les 22.15 degrés. Pour vérifier cette estimation, nous avons récupéré les relevés des températures maximales maritimes et aériennes. Ainsi pour une température maximale à 28.95 degrés pour l'air, la température de l'eau mesurée par les sondes se situe entre 17.8 et 20.4 degrés. Rappelons cependant que la température présente dans le jeu de données n'est qu'une moyenne journalière. Ainsi la température à 2h du matin n'est pas forcément la même à 14h. De plus, pour une température de l'eau maximale mesurée à 22.698 degrés, on relève une température à 27.46 degrés. Ces deux résultats nous confortent dans la prédiction réalisée. Afin d'appuyer notre analyse, nous avons donc réalisé une seconde prédiction en prenant comme valeur de référence cette fois-ci la valeur maximale de température de l'air relevé. Avec un intervalle de confiance à 99% encore une fois, on estime que la température de l'eau sera entre 18.10 et 24.99 degrés pour une température de l'air à 28.95 degrés.

Au vu des résultats obtenus, nos prédictions semblent être valides

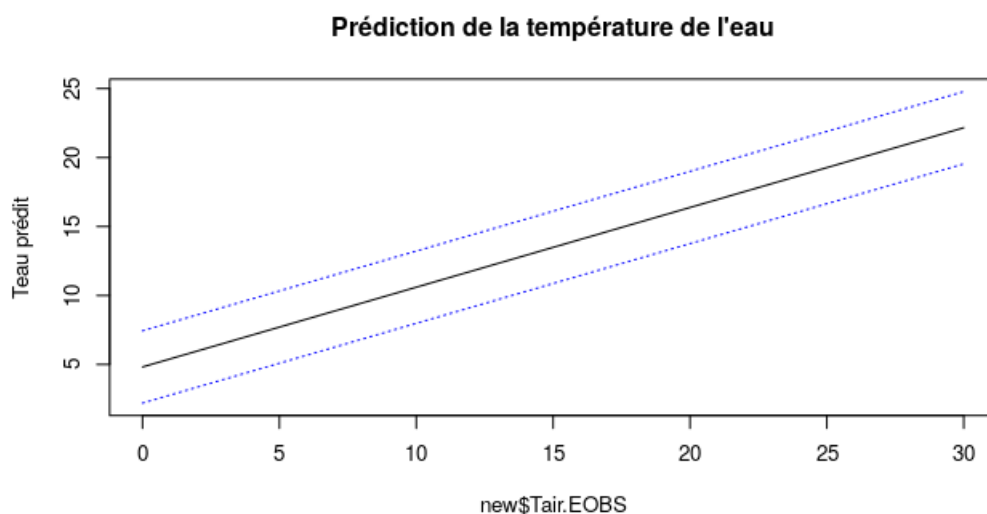


FIGURE 7 – Graphique de prédiction de la température de l'eau en fonction de la température de l'air

6 Limites de la base étudiée

Bien que le jeu de données utilisé soit relativement complet, nous avons relevés des limites qui ne sont pas négligeables pour les analyses que nous avons effectuées.

Tout d'abord, les données étudiées recouvrent la période 2013-2018. Cependant, les années 2013 et 2018 n'étant pas complètes, il faut être vigilant. En effet, les relevés ne commencent qu'à partir du 29 mai 2013. Ainsi l'hiver 2013 est omis. De même pour l'année 2018, dont les relevés se terminent le 05 octobre. La différence est moins impactante pour 2018 mais elle est à prendre en compte.

De plus, notre analyse pourrait être plus précise et plus représentative de la réalité concernant la relation entre Teau et Tair.EOBS. En effet, Teau est actualisé toutes les 2 heures alors que Tair.EOBS ne varie jamais la journée. Ceci s'explique parce que Tair.EOBS provient de l'ECAD, qui a fait une moyenne de la température journalière. De ce fait, nous n'avons pas de contraste entre les températures matinales et l'habituelle douceur de l'après-midi.

Autre limite de la base, l'indicateur de pluie "RainF.EOBS" est aussi donné en moyenne sur la journée. Nous avons dans un premier temps fait la moyenne des températures de l'eau pour chaque sonde entre les jours pluvieux (à RainF non nulle) et non-pluvieux (RainF égale à 0) et nous avons remarqué que les moyennes pour les jours ensoleillés étaient toujours plus élevées qu'en jours de pluie, ce qui nous avaient mené à penser qu'il existait une corrélation entre les variables Teau et RainF.EOBS.

Néanmoins d'après les résultats de l'ACP il semblerait que la quantité de pluie n'influe pas clairement sur la température de l'eau, et que d'autres facteurs entreraient en compte lors des jours pluvieux. On peut penser tout d'abord à la température de l'air qui est généralement moins haute pendant les temps de pluie, mais on aurait aussi apprécié avoir des informations sur la pression atmosphérique ainsi que le taux d'ensoleillement qui pourraient tous deux avoir un impact certain sur la température de l'eau et qui jouent un rôle sur la pluviométrie.

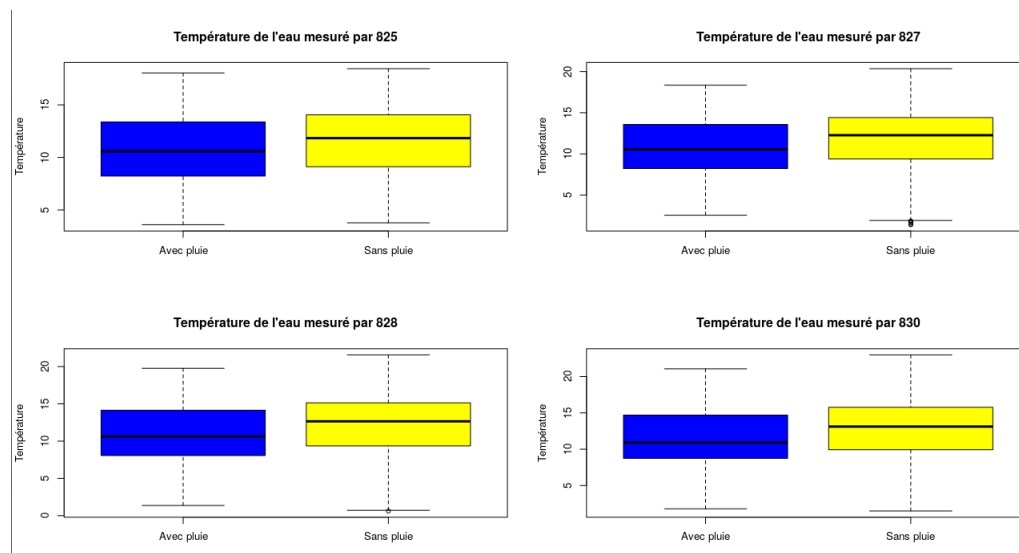


FIGURE 8 – Boxplots des températures de l'eau par sonde en fonction de la présence ou non de pluie

7 Conclusion

Pour terminer ce rapport nous pouvons répondre à notre problématique de départ en citant les facteurs qui, d'après nos études statistiques, influent sur les variations de températures du cours d'eau de la Touques.

Nous avons pu identifier dans nos recherches que la géographie du fleuve pouvait influencer de manière substantielle sur la température de l'eau, en effet les relevés indiquent un faible échauffement de l'eau le long du cours d'eau. De plus la présence de nappes phréatiques de tailles importantes peut aussi faire varier la température par des transferts thermiques entre les abords de la Touques et les eaux souterraines.

La température de l'air est aussi fortement corrélée avec celle de l'eau, au moins en surface, c'est ce que révèlent les relevés, à noter que la Touques reste peu profonde. Mais à défaut de plus d'informations sur l'évolution de la profondeur du fleuve nous ne pouvons indiquer clairement quels sont les effets de la profondeur sur la corrélation entre température de l'air et de l'eau.

La Touques subit aussi des fortes variations annuelles de sa température, et ce au gré des saisons, comme nous le démontrent les ACI réalisées.

Enfin nous n'avons pas pu trouver de corrélation ou d'impact de la pluviométrie sur la température de la Touques avec les informations à disposition. Néanmoins nous soupçonnons qu'une légère influence puisse être décelée au regard d'un complément de données et de nouvelles variables pour compléter notre étude. De même, nous n'avons pas relevé de phénomène pouvant mettre en évidence le réchauffement climatique. Un jeu de données s'étendant sur une période plus longue aurait cependant pu nous permettre de le visualiser.