

UNIVERSITÉ DE CAEN

ANALYSE DE DONNÉES

M1 INFORMATIQUE

Les températures de la Touques

Auteurs :

Robin GALLIS

Pierre MAURAND

Nicolas AUBRY

Lucie PETIT

Professeur :

M Bruno DARDAILLON

13 mars 2021



UNIVERSITÉ
CAEN
NORMANDIE

Table des matières

1	Introduction	2
2	L'analyse descriptive	2
3	L'analyse en composantes indépendantes	3
4	L'analyse en composantes principales	4
5	Conclusion	5

La Touques, fleuve prenant sa source dans le Perche, traversant le Calvados et l'Orne pour se jeter dans la Manche entre Trouville-sur-Mer et Deauville, est sujet à des modifications de températures importantes au cours d'une année. C'est de 2013 à 2018 que des chercheurs de L'ECAD se sont penchés sur les modifications de température de ce fleuve.

Pour ce faire, 4 sondes furent placées à différents endroits afin de capter la température de l'eau, la température de l'air et la pluviométrie. Les relevés furent effectués toutes les 2 heures sur les 4 sondes durant un peu plus de 5 ans (allant du 29 mai 2013 à 14h au 5 octobre 2018 à 8h). Ce qui nous donne donc plus de 93 000 observations !

Toute cette enquête nous amène alors à nous poser la question : La température de l'air, la pluviométrie et la date permettent-elles de comprendre les variations de températures de la Touques entre 2013 et 2018 ?

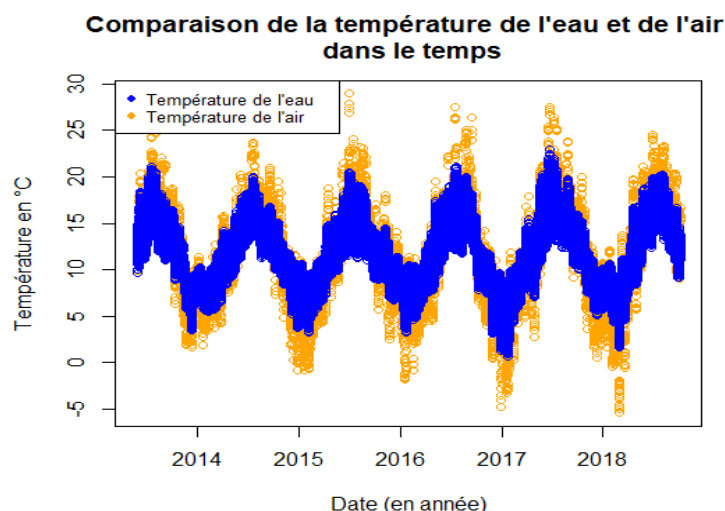
Pour répondre à cette problématique, nous allons vous présenter une analyse descriptive des résultats de cette enquête, ensuite nous tenterons de prouver qu'il y a des corrélations grâce à une analyse de composantes indépendantes et enfin nous vérifierons ces corrélations à l'aide d'une analyse en composantes principales.

L'analyse Descriptive

La Base que nous allons étudier lors de cette analyse Descriptive contient 6 informations (variables) que nous allons détailler par la suite :

- La première variable, est la variable "*t*", c'est une variable **qualitative ordonnée**, représentant le temps toutes les deux heures.
- La seconde variable, est la variable "*date*", cette variable est-elle aussi **qualitative ordonnée**, représentant la date en jour, mois et année.
- La troisième variable est la variable "*Teau*", c'est une variable **quantitative continue**, qui représente la température de l'eau.
- La quatrième variable est la variable "*id_sonde*", qui est une variable **qualitative nominale**, elle représente le numéro de la sonde qui a relevé la température.
- La cinquième variable est la variable "*Rainf.EOBS*", qui représente le volume de pluie issu du site ECAD. C'est une variable **quantitative continue**.
- La sixième variable "*Tair.EOBS*" est la variable représentant la température de l'air selon le site ECAD. C'est une variable **quantitative continue**.

Notre objectif étant de déterminer s'il y a un rapport entre la température de l'air, pluviométrie et température de la Touques, nous allons présenter sur un graphique les variations de température de l'eau et de l'air tout au long de l'étude. Par cela nous pourrions déjà savoir si la température de l'air semble suivre la température de l'eau (sans pour autant être corrélées).



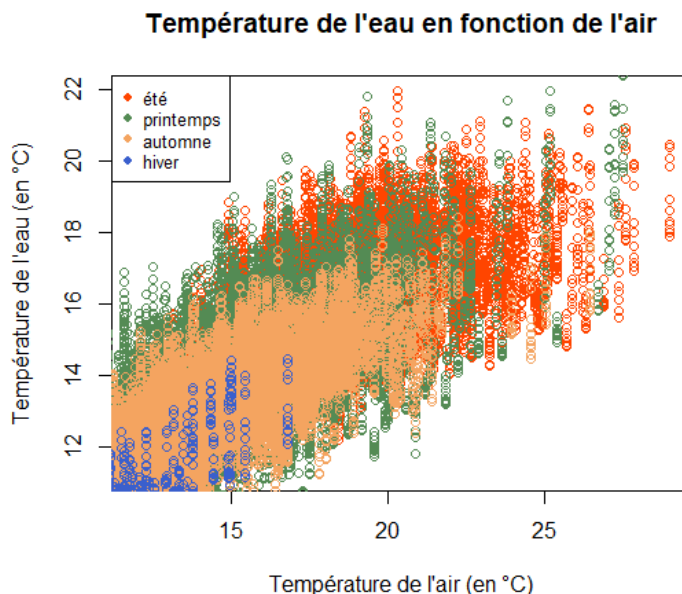
La première chose marquante est l'étendue des températures de l'eau qui est nettement moins grande que celle de l'air. Cela est visible sur le graphique ci-contre : l'air va plus fluctuer que l'eau, et ce, toute l'année. Néanmoins, ces deux variables semblent suivre une logique au cours de l'année :

En hiver les températures sont plus basses qu'en été ! De ce fait nous allons donc avoir une courbe similaire à une sinusoïde entre le début et la fin de cette étude.

Concernant les moyennes, sur toute l'étude, la moyenne des températures de l'eau est de 11.26°C alors que celle de l'air est de 11.58°C.

Après interprétation, il nous est possible de prendre en compte ces moyennes dans notre analyse, car nous remarquons qu'il y a peu de valeurs extrêmes aussi bien pour l'eau que pour l'air ce qui ne biaiserait pas nos moyennes. D'ailleurs, lorsque nous calculons les médianes de ces deux variables, nous obtenons des chiffres assez proches (11.46°C pour l'eau et 11.57°C pour l'air). Les moyennes des températures de l'air et de l'eau sont assez proches pour nous poser la question de leur corrélation. Pour finir, les températures semblent présenter une saisonnalité avec des pics froids en hiver et des pics de chaleur en été. C'est pourquoi nous allons par la suite analyser la température de l'eau et de l'air durant les saisons.

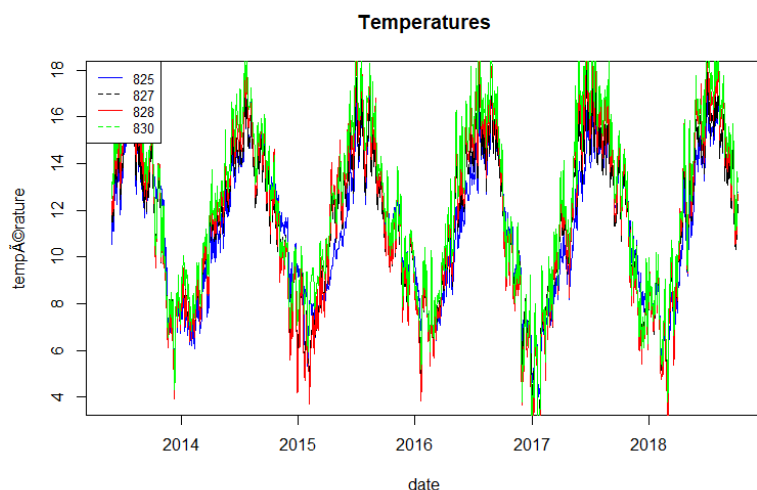
Le graphique ci-contre représente la température de l'eau en fonction de la température de l'air groupé par saison. Nous voulions être sûr qu'il y avait un lien entre saisons et chaleur globale. Après analyses, en hiver, la température est toujours plus faible qu'en été, mais surtout les températures d'hiver ont une étendue beaucoup moins grande que les températures de printemps. En outre, il semble y avoir donc un lien entre date et température, en hiver il fait plus froid, donc la température de l'air et celle de l'eau baisse, c'est le contraire en été. Ce qui est aussi remarquable c'est que la différence de température entre l'eau et l'air est d'autant plus visible quand on rentre dans les jours les plus chauds. Nous pouvons donc ressortir que plus la température de l'eau augmente plus la température de l'air s'éloigne de celle de l'eau en augmentant plus et plus la température de l'eau est froide plus la température de l'air a tendance à baisser en dessous de celle de l'eau.



Une analyse en composantes indépendantes nous permettra de repérer la probable relation entre ces deux variables qui nous semblent liées depuis le début.

L'analyse en composantes indépendantes

Afin de commencer cette analyse qui nous permettra de repérer d'éventuelles valeurs aberrantes et de détecter l'indépendance de nos valeurs, nous allons tester pour chaque sonde (qui sont au nombre de 4) l'évolution des températures de l'eau en fonction de la date.



Les températures des 4 sondes sont cohérentes entre-elles, et il ne semble pas y avoir d'aberrations probantes lorsque nous séparons les résultats de ses sondes. De plus elles semblent toutes, comme décrit ci-dessus, suivre une saisonnalité avec des températures plus basses l'hiver et des températures plus hautes l'été. Le fait d'afficher la sonde 830 en premier n'est pas anodin car cette sonde semble tout de même aller à des extrêmes de températures avec dans la quasi-totalité

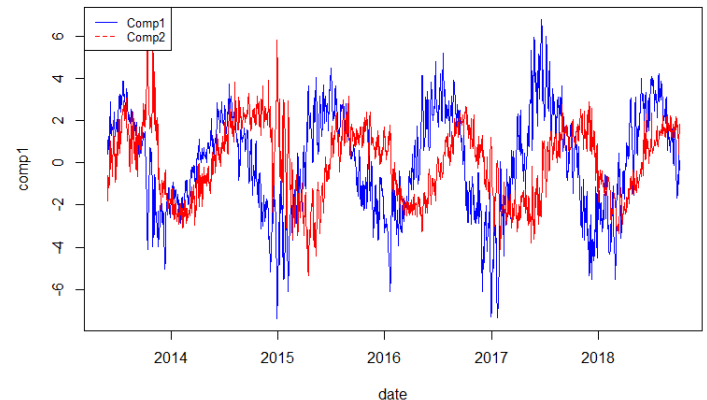
des années des températures un peu plus basses que les autres en hiver et un peu plus hautes en été. Cela

Teau825	Teau827	Teau828	Teau830
Min. : 3.799	Min. : 2.115	Min. : 1.207	Min. : 1.897
1st Qu.: 8.658	1st Qu.: 8.699	1st Qu.: 8.606	1st Qu.: 9.247
Median :11.428	Median :11.673	Median :11.861	Median :12.342
Mean :11.262	Mean :11.419	Mean :11.702	Mean :12.302
3rd Qu.:13.832	3rd Qu.:14.138	3rd Qu.:14.847	3rd Qu.:15.461
Max. :17.809	Max. :18.545	Max. :20.208	Max. :21.135

semble se confirmer avec l'analyse mathématique de la sonde 830 (ici Teau 830), mis à part ses températures minimales, sa température moyenne, sa température médiane et son maximum est plus élevé

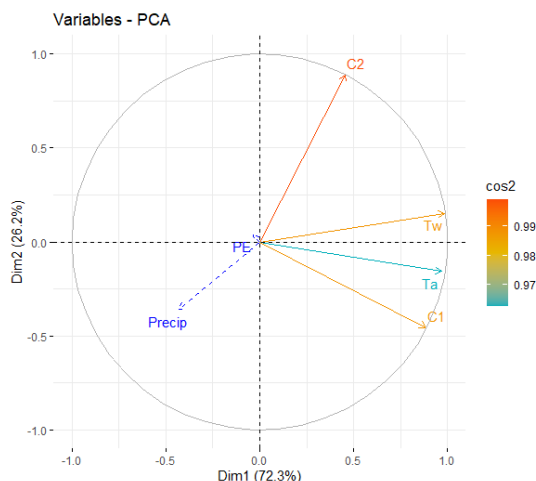
que les températures des autres sondes. Nous pouvons d'ailleurs observer que la sonde 825 (étant la plus proche de la source) aura les températures globales les plus faibles, et plus la sonde se rapproche de la mer (avec la sonde 830 qui est la plus proche de la mer), y aurait-il un rapport, même étroit entre source et température ? Nos données présentes ne permettent pas de pouvoir répondre à cela, néanmoins nous allons tenter de détecter la présence d'éventuelles nappes phréatiques avec la résolution de l'ACI.

Nous avons extrait les 2 composantes qui présentent une saisonnalité. La composante 1 semble suivre le principe évoqué précédemment : La température baisse en hiver et remonte en été, avec cela quelques difformités liées à des années plus chaudes ou plus froides que la normale. Par contre, la composante 2 qui est déphasée avec la première semble nous montrer le contraire : Les températures augmentent pendant l'hiver et rebaissent pendant l'été. Cela est typique de la présence d'une nappe phréatique ! De plus la composante 2 est beaucoup plus stable que la composante 1, signifiant que les variations sont moins influencées par la température de l'air pour les sondes de cette composante que pour les autres sondes. Encore une fois cela est typique de la présence d'une nappe qui va influencer les températures car étant sous terre, une nappe va se réchauffer plus tard qu'une rivière en surface et donc grâce à cette chaleur emmagasinée, elle va réchauffer la rivière en surface produisant donc ce déphasage. Il y a donc quelque chose (probablement une nappe) qui va influencer les températures à un endroit. Il nous reste désormais à trouver des corrélations entre nos variables afin de détecter les éventuels liens faisant varier la température de l'eau.



L'analyse en composantes principales

Afin de déterminer d'éventuelles corrélations entre nos variables nous allons effectuer une analyse en composantes principales sur la sonde 825, nous aurions pu effectuer cette même analyse pour les autres sondes mais prendre la sonde 825 nous permet d'avoir les résultats au plus proche de la source de ce fleuve. En plus de nos données, nous avons rajouté de nouvelles variables : il s'agit de la piézométrie sur cette même sonde, cela nous donnera d'avantages de composantes à analyser et à interpréter afin de trouver des corrélations. Le cercle des corrélations nous donnera donc une première idée des corrélations.



Sur ce cercle des corrélations nous remarquons une très faible corrélation entre les précipitations et le reste des variables : la pluviométrie n'influence donc pas ou très peu la température de l'eau, et encore moins la température de l'air ! Nonobstant cette variable qui n'explique pas ses températures, nous remarquons une très forte corrélation entre la température de l'air (Ta) et la température de l'eau (Tw). C'est désormais certain : l'un influence l'autre. De plus de grandes corrélations sont aussi à dénoter entre ces deux

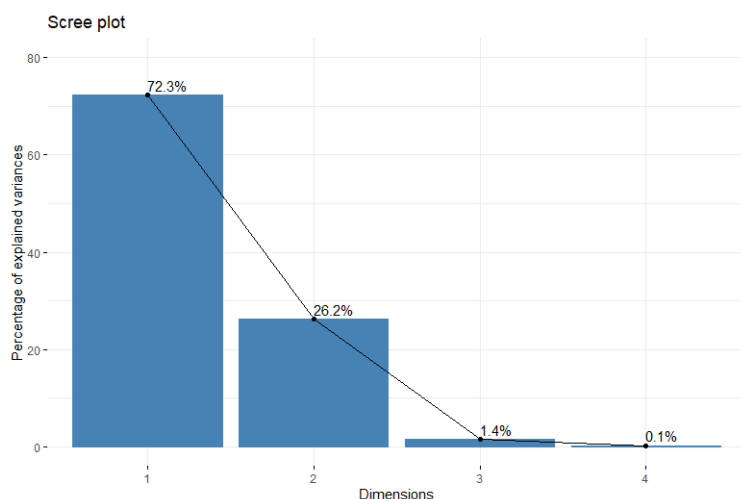
variables et les composantes utilisées précédemment dans l'ACI. Aussi il est important de noter que les deux composantes issues de l'ACI ne sont pas corrélées car elles sont parfaitement orthogonales. Par le calcul des corrélations, nous allons donc vérifier nos dires :

Première variable	Deuxième Variable	Coefficient de corrélation
Température de l'eau	Composante 1	0.81
Température de l'eau	Composante 2	0.58
Température de l'eau	Température de l'air	0.90
Pluviométrie	Température de l'eau	-0.02
Pluviométrie	Piézométrie	0.10

Cette figure confirme nos dires précédents, la température de l'eau est bien corrélée à la première composante (donc aux saisons), une

corrélations non négligeable est à noter entre la température de l'eau et la composante 2 (qui semblait indiquer la présence d'une nappe) et surtout la température de l'eau est extrêmement corrélée à la température de l'air, contrairement à la pluviométrie. La pluviométrie et la Piézométrie ne semblent pas (ou peu) corrélées. Enfin il est important de noter que plus une variable est proche du cercle, plus elle est bien représentée, de ce fait la température de l'eau, la température de l'air ainsi que les deux composantes sont bien représentées, par contre la piézométrie et la pluviométrie beaucoup moins, ce qui fait que nous ne pourrions qu'analyser leurs corrélations qu'avec l'incertitude de la mauvaise représentation.

Pour finir nous allons pouvoir construire un diagramme des éboulis afin de savoir combien de dimensions sont réellement nécessaires pour représenter la majorité des données.



Seules les deux premières dimensions sont réellement nécessaires pour représenter la majorité des données car à elles deux elles représentent environ 98 % des données. Ce qui est une écrasante majorité. S'appuyer uniquement sur ces deux axes serait donc suffisant pour cette étude.

Conclusion

La température de l'eau est liée à la température de l'air dans la Touques de 2013 à 2018. C'est ce qui est le plus probant lors de l'analyse des variables associées à l'étude des chercheurs de l'ECAD. Ceci étant dit, ce n'est pas la seule raison de l'évolution de la température de ce fleuve durant ces années. Tout d'abord nous savons qu'une source indépendante (probablement une nappe) influence aussi la température de l'eau sur certaines sondes. De plus la saison influence sur la température de l'eau et sur celle de l'air, donc il est possible de prévoir à peu près les températures sur une année. Malgré cela, la pluviométrie ne semble que très peu influencer sur la température de l'eau, et encore moins sur celle de l'air. Toutes ses données nous permettent donc d'arriver à la conclusion que la plus grande source d'influence pour la température de l'eau est la température de l'air, si la température de l'air monte, celle de l'eau augmentera aussi et réciproquement si la température de l'air baisse. Néanmoins il est naturel de penser que cette température n'est pas la seule influence, en effet l'activité humaine, la végétation, les cours d'eau affluents de cette rivière ou encore l'altitude pourraient aussi influencer sur la température de l'eau mais nous manquons de données afin de l'affirmer, cependant nous connaissons du moins la principale influence qui encore une fois est la température de l'air.