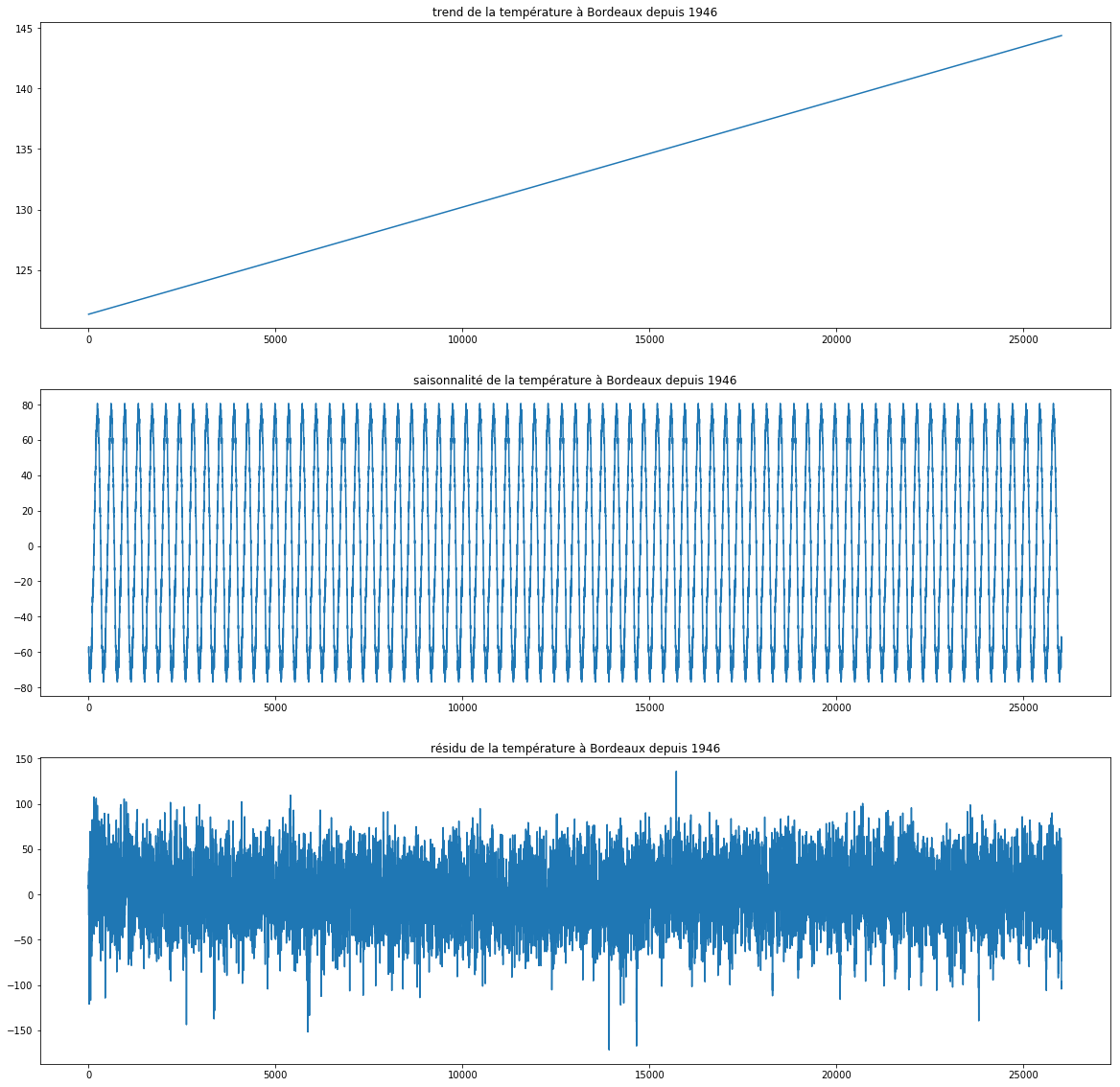
## Décomposition de la série

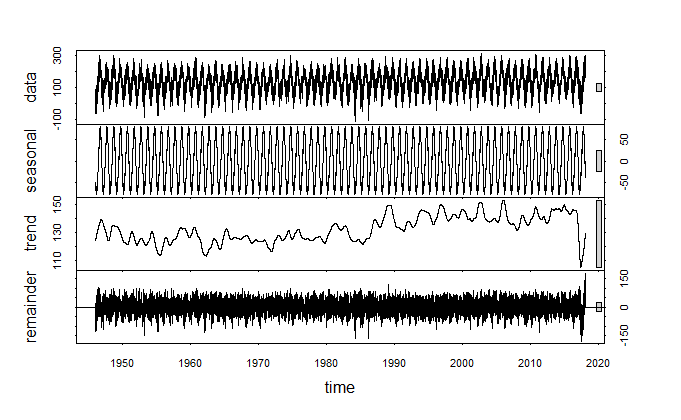
On commence par décomposer la série selon un trend et une saisonnalité. Pour cela, plusieurs méthodes s’offrent à nous. Nous choisirons dans un premier temps la méthode de Buys-Ballot généralisée, qui consiste à écrire chacune des composantes déterministes comme une combinaison linéaire de fonctions connues du temps.

On obtient alors la décomposition suivante :



La variance des résidus est élevée en comparaison de la différence du trend entre le début et la fin de la série temporelle ; les données sont fortement bruitées.

On aurait également pu envisager une décomposition de la forme Seasonal decomposition of Time series by Loess, ce qui est aisé avec R. On choisit une fenêtre de la largeur d’une période et l’on obtient la décomposition suivante :



On obtient pour cette décomposition un écart type de 33,7, à comparer avec le 33,5 de la méthode de Buys-Ballot.

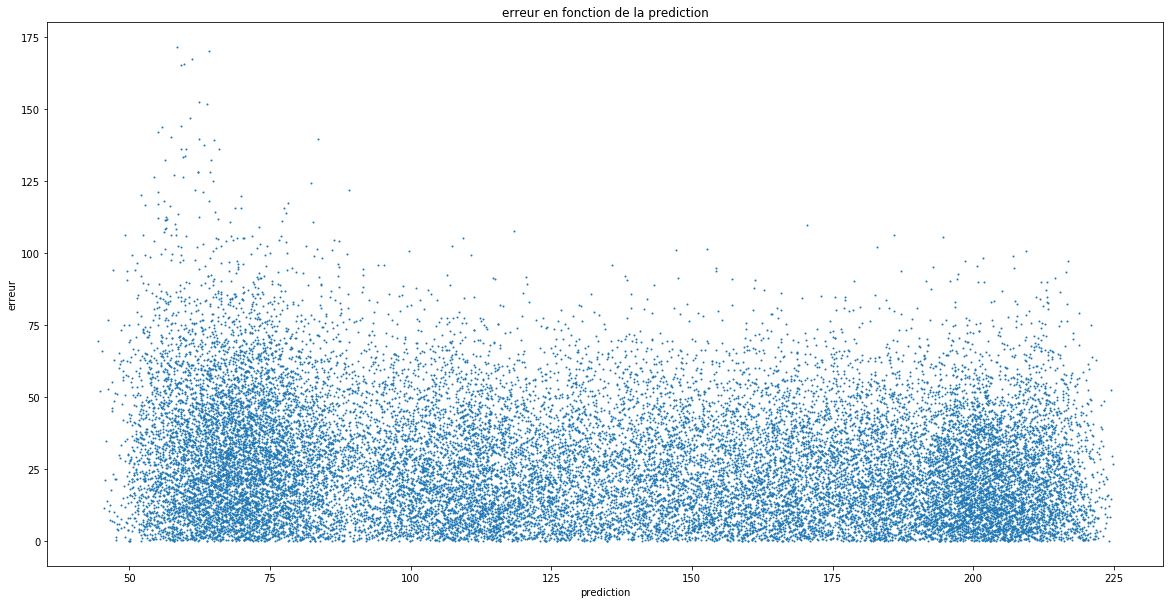
## Interrogation sur la période de la série

On s’interroge alors sur un point. On sait qu’une année dure 365,24 jours et non pas 365 comme on en fait souvent l’approximation. On a ici pris en compte le décalage dans la formule des Ŝ, mais qu’en serait-il si on l’omettait ?

En refaisant une simulation qui fait cette approximation, on obtient une variance de 33,8 sur les résidus, contre 33,5 dans le cas plus rigoureux. Par la suite on se permettra de désigner la période par 365 ou 365,24 compte tenu de la proximité des deux valeurs.

## Validité du modèle additif

On a jusque alors envisagé un modèle de type additif pour la série temporelle : la valeur temporelle s’écrirait comme la somme d’une tendance, d’une saisonnalité et d’une erreur que l’on suppose modélisable par un bruit gaussien faible. On va chercher à vérifier cette hypothèse d’additivité en traçant la courbe de l’erreur obtenue (en valeur absolue) en fonction de la valeur prédite par Buys-Ballot généralisé.



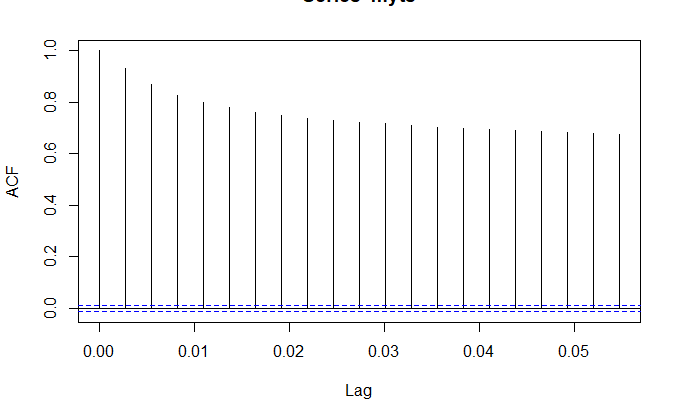
On ne distingue pas de relation nette entre la variance du bruit et la grandeur de la prédiction ; aussi le modèle additif sera-t-il considéré comme pertinent par la suite.

## Modélisation

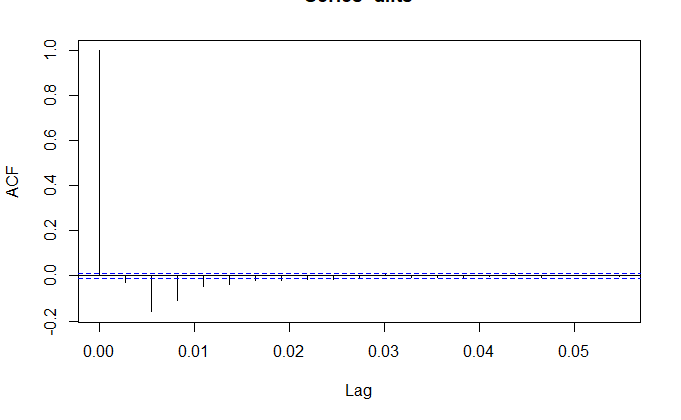
La présence d’une tendance et d’une saisonnalité nous invite à envisager la modélisation de la série temporelle par un processus de type SARIMA.

## Ordre de différentiation

Par acquis de conscience, examinons la fonction d’autocorrélation afin de déterminer l’ordre de différentiation requis.



Comme présumé, le processus est non-stationnaire. On va alors tracer l’autocorrélogramme de la série temporelle différentiée.



Cet autocorrélogramme nous invite à considérer qu’un ordre de différentiation 1 est suffisant pour le modèle SARIMA, ce qui concorde avec les décompositions que l’on a obtenu précédemment (les restes ne semblaient pas présenter de tendance quadratique.)

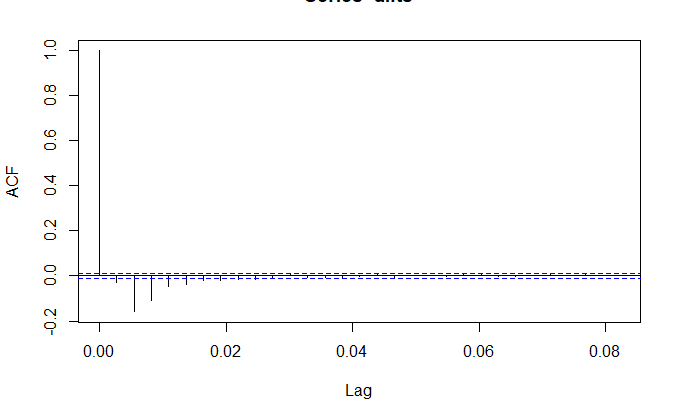
## Saisonnalité

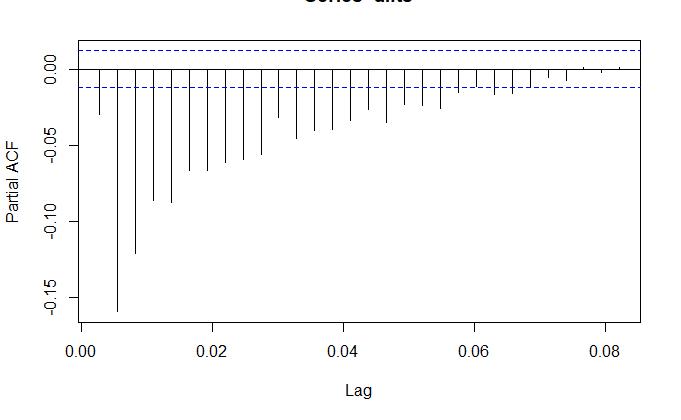
Les études précédentes et le bon sens climatique nous invitent à considérer une saisonnalité d’ordre 365.

## Étude des ordres p et q

Un processus AR d’ordre p se caractérise par sa fonction d’autocorrélation partielle qui s’annule à partir de l’ordre p + 1, et un processus MA d’ordre q se caractérise par sa fonction d’autocorrélation qui s’annule à partir de l’ordre q + 1.

Traçons donc les autocorrélogramme et autocorrélogramme partiel de la série différentiée.





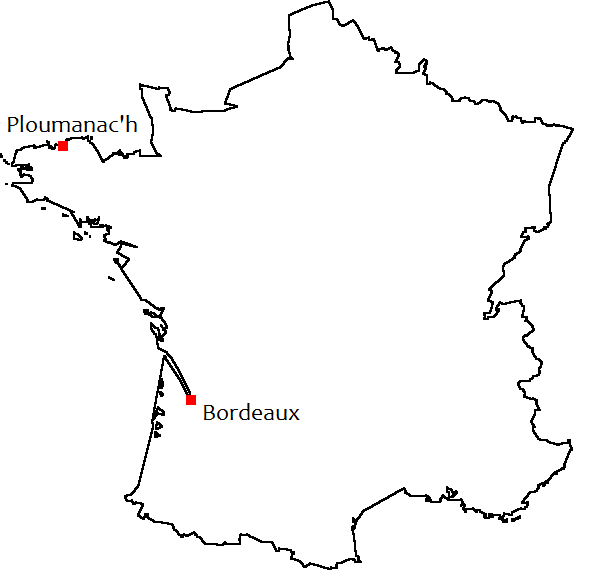
On peut dès lors envisager de choisir q = 6. En revanche le choix de p est problématique puisque la décroissance est très lente et les calculs sont trop lourds si l’on choisit p = 21 par exemple.

On peut ainsi considérer un modèle SARIMA365(26,1,6)(0,1,0). Néanmoins le temps de calcul est trop grand pour de telles valeurs, sans compter un risque important d’overfitting. On pressent que le tracé de autocorrélogramme partiel est pathologique, et un traitement plus important des données aurait été nécessaire afin d’obtenir une modélisation satisfaisante.

# ֤

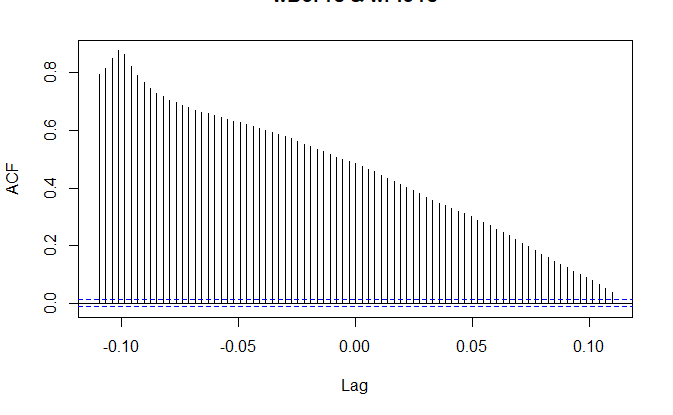
# Étude des corrélations

Dans une dernière partie, on va s’attacher à étudier les corrélations entre les séries temporelles de température à Bordeaux et à Ploumanac’h.



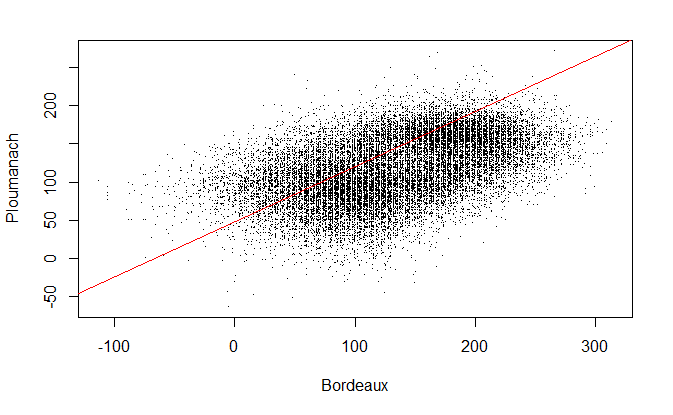
## Étude de la cross-correlation

Le tracé de la cross-correlation entre les deux séries temporelles révèle une corrélation forte (coefficient de 0.88) avec un maximum atteint pour un décalage de trois jours.



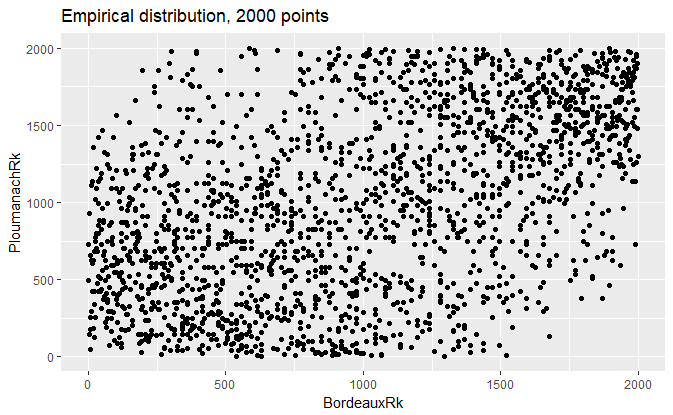
### Estimation d’une copule

On trace tout d’abord le graphe des températures obtenues à Ploumanac’h en fonction des températures relevées à Bordeaux.



La corrélation entre les deux séries est sensible et fait écho au sens physique. La corrélation de Spearman est de 0.51, celle de Kandall de 0,34, confirmant que les deux séries sont corrélées. On note à ce sujet des p-values très basses qui sont gages d’un résultat robuste : elles sont inférieures à 2e-16 dans les deux cas.

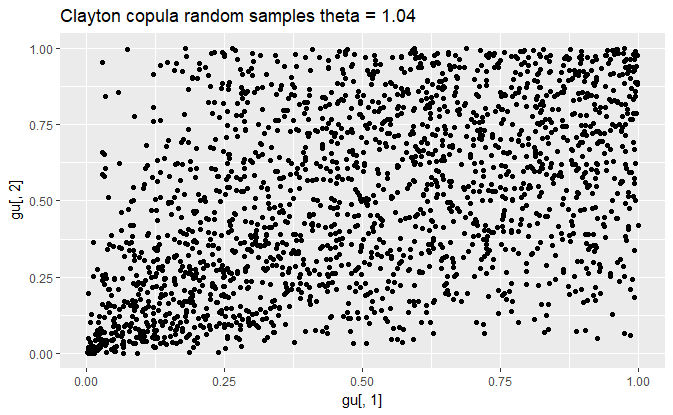
On peut tracer le graphe des rangs des températures à Ploumanac’h en fonction des rangs des températures à Bordeaux pour se figurer l’allure de la relation des deux séries temporelles. On ne considère ici que 2000 points par souci de lisibilité.



Nous allons à présent choisir une copule pour modéliser la corrélation entre les deux grandeurs. On a vu que les événements extrêmes ne sont pas fortement corrélés, aussi on privilégie la copule de Clayton qui découple plus les canicules dans un premier temps.

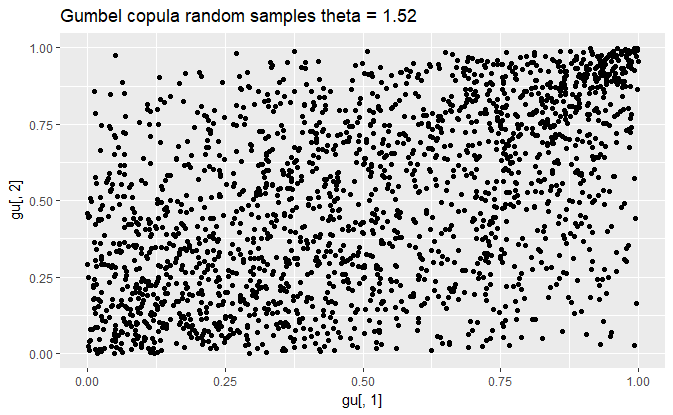
On sait que pour elle en dimension 2, le tau de Kandall vaut theta/(theta+2). On en déduit la valeur du paramètre : ici ce sera 1,04.

On en déduit une copule qui modélise la corrélation des deux séries temporelles :



Dans un second temps, on aborde la copule de Gumbel. On sait que pour elle en dimension 2, le tau de Kandall vaut 1 – 1/theta. On en déduit la valeur de theta dans notre cas : on obtient une valeur de 1,52.

On en déduit une copule qui modélise la corrélation des deux séries temporelles :



Ces tracés révèlent les limites des modélisations choisies, pour lesquelles les corrélations des événements extrêmes sont visiblement trop importantes.