

UNIVERSITÉ LILLE1

---

# TRAVAIL ENCADRÉ DE RECHERCHE

MASTER 1 MAS PARCOURS ISN

---

*Auteur :*  
Cédric DJAMEGO

*Superviseur :*  
Azzouz DERMOUNE

**Analyse Statistique de la température en France entre 1900 et 2015**

26 mai 2017

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Analyse globale sur le tableau des températures moyenne annuelle</b>	<b>3</b>
2.1	Données . . . . .	3
2.1.1	Statistiques descriptives . . . . .	4
2.1.2	Boîtes à moustaches . . . . .	4
2.1.3	Histogrammes . . . . .	9
2.1.4	Histogrammes des dérivées discrètes . . . . .	10
2.1.5	Transformation de Box-Cox . . . . .	11
2.2	Distribution des couples de la température des années consécutives . . . . .	13
2.3	Analyses sur les décennies . . . . .	15
2.3.1	Analyse des corrélations significatives entre variables . . . . .	16
2.3.2	Analyse des moyennes et des variances . . . . .	17
<b>3</b>	<b>Analyse en composantes principales</b>	<b>19</b>
3.1	Valeurs propres . . . . .	19
3.2	Axes factoriels . . . . .	21
3.2.1	Description des axes factoriels par les variables . . . . .	21
<b>4</b>	<b>Régression linéaire</b>	<b>24</b>
4.1	Régression linéaire sur la période de 1900 à 1980 . . . . .	25
4.2	Régression linéaire sur la période de 1981 à 2015 . . . . .	25
<b>5</b>	<b>Splines cubiques</b>	<b>28</b>
5.1	Définition . . . . .	28
5.1.1	Paramétrisation et base d'un espace vectoriel . . . . .	28
5.1.2	Modèle pour la température : Dermoune, Preda . . . . .	29
5.1.3	Construction des Bases $\mathbf{B}_{0123}$ , $\mathbf{B}_{012}$ , $\mathbf{B}_{002}$ . . . . .	29
5.1.4	Construction de la Base $\mathbf{B}_{202}$ . . . . .	31
<b>6</b>	<b>Conclusion</b>	<b>34</b>
<b>7</b>	<b>Annexes : code R</b>	<b>37</b>

# 1 Introduction

Dans le cadre du projet TER (Travail Encadré de Recherche), nous avons choisi de nous intéresser à l'évolution de la température en France de 1900 à 2015. Nous allons utiliser différentes méthodes statistiques afin d'évaluer la tendance de la température durant cette période de 116 années. Dans un premier temps nous nous intéressons à l'ensemble des données sur lesquelles on effectue une statistique descriptive (moyennes, médianes, variances, histogrammes, boîtes à moustaches). Ensuite, nous nous intéressons uniquement aux températures moyennes annuelles sur cette même période, on se demande alors si derrière ces données se cache une loi usuelle (dans ce cas la loi normale) pour cela on va utiliser des histogrammes. On effectue ensuite une régression linéaire, afin d'évaluer la tendance linéaire ; puis on fait une analyse sur les résidus pour voir si ceux ci sont gaussiens (si c'est le cas il est possible alors de faire des prédictions à partir par exemple des intervalles de confiances). Nous nous intéressons ensuite aux couples consécutifs sur les températures moyennes afin de faire des prévisions sur une année.

Après avoir effectué cette analyse, on s'intéresse ensuite aux températures moyennes sur plusieurs décennies notamment ici les onze premières décennies soit de 1900 à 2009, on effectue une statistique descriptive sur chaque décennie, ensuite on étudie les corrélations qui existent entre celles-ci ; le but ici étant toujours de savoir comment évolue la température c'est-à-dire de voir en quelque sorte si on a une hausse de la température d'une décennie à l'autre. Pour corroborer les résultats obtenus, on effectue une analyse en composantes principales (ACP) sur le tableau dont les variables sont les onze premières décennies. A partir de l'ACP, on regarde comment a évolué la température sur les différentes décennies et ceci permettra d'avoir une idée sur les décennies plus chaudes et les moins chaudes durant ces 116 années ; en effet on va regarder les corrélations entre les différentes variables, mais aussi regarder les contributions des variables sur les différents axes factoriels retenus et à partir du cercle de corrélation on va faire une distinction des variables bien représentées sur les axes principaux et conclure.

Enfin, après avoir remarqué la difficulté liée à la modélisation de la température à partir d'un modèle linéaire (résultats de la régression), on introduit la notion de spline cubique aléatoire, à forte variabilité dans le but de modéliser au mieux l'évolution de la température.

## 2 Analyse globale sur le tableau des températures moyenne annuelle

### 2.1 Données

Date	T min	T max	T moy	Date	T min	T max	T moy	Date	T min	T max	T moy
1900	7,4	16,9	12,2	1947	7,7	17,6	12,6	1994	9,1	17,6	13,3
1901	6,5	15,8	11,1	1948	7,2	17,1	12,1	1995	8,4	17,3	12,9
1902	6,6	15,8	11,2	1949	7,5	17,8	12,6	1996	7,5	16,3	11,9
1903	6,8	16,2	11,5	1950	7,4	16,7	12,1	1997	8,4	17,9	13,1
1904	7,1	16,8	12	1951	7,3	16,2	11,7	1998	8	17,1	12,6
1905	6,6	15,9	11,2	1952	7,2	16,6	11,9	1999	8,6	17,4	13
1906	6,8	16,7	11,8	1953	7	16,7	11,9	2000	8,8	17,5	13,1
1907	6,7	16,2	11,5	1954	6,9	15,8	11,4	2001	8,4	17,2	12,8
1908	6,6	16,2	11,4	1955	7,1	16,7	11,9	2002	8,8	17,6	13,2
1909	6,1	15,6	10,8	1956	6,1	15,1	10,6	2003	8,6	18,4	13,5
1910	7	15,8	11,4	1957	7,3	16,5	11,9	2004	8,2	17,1	12,6
1911	7,4	17,3	12,3	1958	7,5	16,1	11,8	2005	8	17,3	12,6
1912	6,8	16	11,4	1959	7,7	17,5	12,6	2006	8,7	17,8	13,3
1913	7,4	16,5	11,9	1960	7,5	16,2	11,8	2007	8,4	17,5	12,9
1914	6,9	16,1	11,5	1961	7,9	17,3	12,6	2008	8,1	17	12,6
1915	7	16	11,5	1962	6,5	15,6	11,1	2009	8,2	17,7	13
1916	7,1	16,2	11,6	1963	6,4	15	10,7	2010	7,4	16,4	11,9
1917	6	15,2	10,6	1964	7,3	16,2	11,7	2011	8,8	18,5	13,6
1918	6,8	16,6	11,7	1965	7	15,6	11,3	2012	8,2	17,4	12,8
1919	6,5	15,7	11,1	1966	7,8	16,3	12	2013	8,1	16,8	12,4
1920	7	16,7	11,9	1967	7,4	16,5	12	2014	9,1	18,4	13,8
1921	7,3	17,6	12,4	1968	7,3	16	11,6	2015	8,7	18,3	13,5
1922	6,6	15,7	11,1	1969	7,2	16	11,6				
1923	7,1	16,4	11,7	1970	7,3	16,1	11,7				
1924	6,9	16	11,5	1971	6,9	16,3	11,6				
1925	6,6	15,8	11,2	1972	6,9	15,8	11,4				
1926	7,5	16,8	12,1	1973	7	16,3	11,6				
1927	7,3	16,1	11,7	1974	7,6	16,3	11,9				
1928	7,5	17,1	12,3	1975	7,4	16,2	11,8				
1929	6,5	16,6	11,6	1976	7,3	16,9	12,1				
1930	7,9	16,5	12,2	1977	7,7	16,2	11,9				
1931	6,8	15,7	11,2	1978	7	15,8	11,4				
1932	7,2	16	11,6	1979	7,3	15,9	11,6				
1933	6,9	16,3	11,6	1980	7	15,4	11,2				
1934	7,6	16,9	12,2	1981	7,7	16,2	11,9				
1935	7,3	16,2	11,7	1982	8,2	17,1	12,7				
1936	7,5	16,1	11,8	1983	7,9	16,9	12,4				
1937	7,8	16,8	12,3	1984	7,2	16,1	11,6				
1938	6,9	16,6	11,8	1985	6,7	16	11,4				
1939	7,2	15,8	11,5	1986	7,3	16,1	11,7				
1940	6,3	15,4	10,9	1987	7,5	15,8	11,7				
1941	6,5	15,3	10,9	1988	8,2	16,7	12,5				
1942	6,8	16,1	11,5	1989	8,1	17,9	13				
1943	7,8	17,3	12,6	1990	8,3	17,7	13				
1944	6,9	15,9	11,4	1991	7,5	16,5	12				
1945	7,3	17,5	12,4	1992	8	16,7	12,3				
1946	6,9	16,1	11,5	1993	7,7	16,4	12,1				

### 2.1.1 Statistiques descriptives

On s'intéresse dans un premier temps aux Températures minimales, moyennes et maximales représentées par les variables  $x$ ,  $y$ ,  $z$  respectivement et sur lesquelles on effectue une analyse uni-variée, on obtient les résultats suivants :

-  $\text{Var}(x)=0.46$  ,  $\text{Var}(y)=0.47$ ,  $\text{Var}(z)=0.57$

```
> summary(x)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  6.0    6.9    7.3    7.4    7.8    9.1
> summary(y)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
10.60 11.50 11.90 11.97 12.40 13.80
> summary(z)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
15.00 16.00 16.40 16.55 17.10 18.50
```

FIGURE 1 –

### 2.1.2 Boîtes à moustaches

Les boîtes à moustaches permettent de visualiser plusieurs paramètres de distribution d'une variable à partir de la médiane ( $m$ ) qui la valeur centrale de la boîte, l'intervalle inter-quartile ( $IC = q_3 - q_1$ ), 1<sup>er</sup> quantile( $q_1$ , bord supérieure du rectangle) et le 3<sup>e</sup> quantile ( $q_3$ , bord inférieure du rectangle), l'arrête supérieure est égale à  $q_3 + 1.5IC$  tandis que l'arrête inférieure vaut  $q_1 - 1.5IC$ .

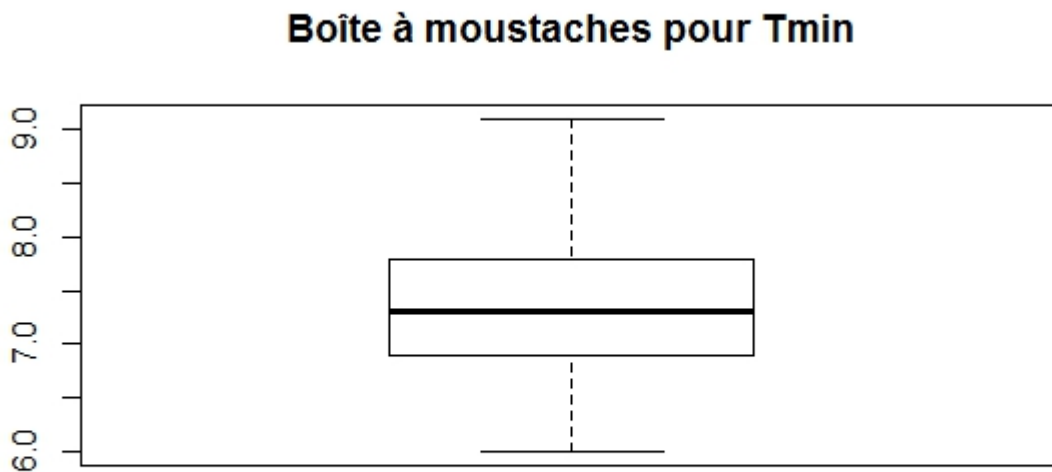


FIGURE 2 – Températures minimales

### Boîte à moustaches pour Tmoy

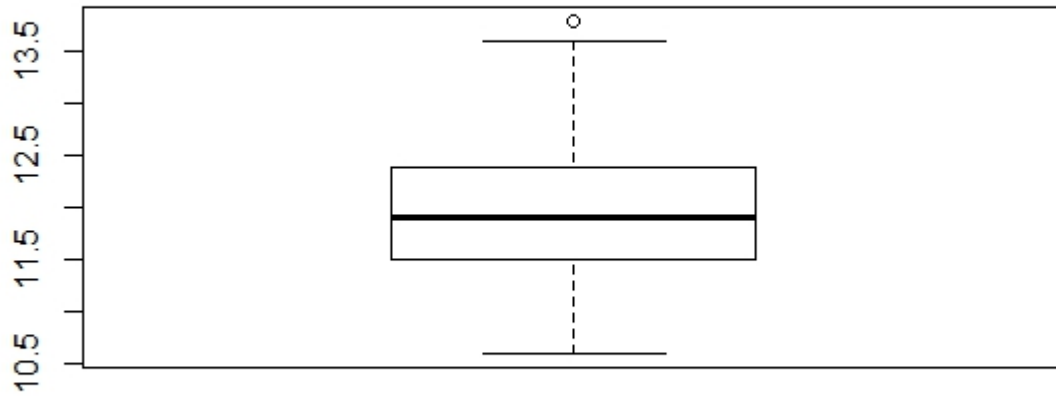


FIGURE 3 – Températures moyennes

### Boîte à moustaches pour Tmax

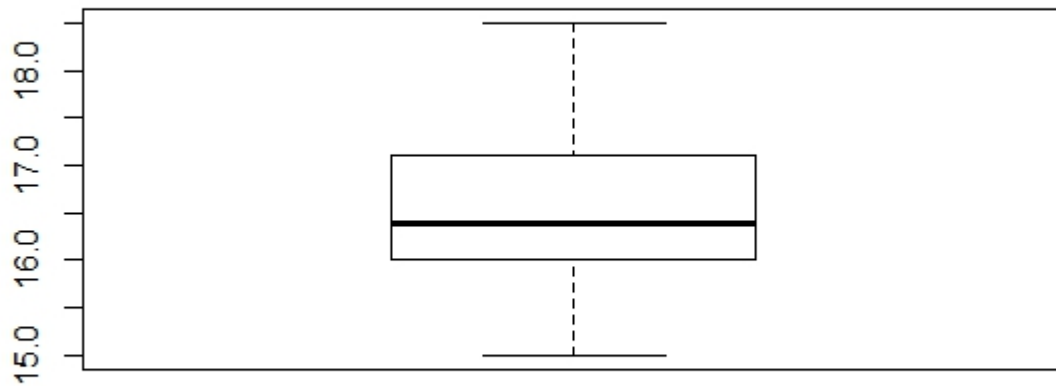


FIGURE 4 – Températures maximales

### Boîtes à moustaches

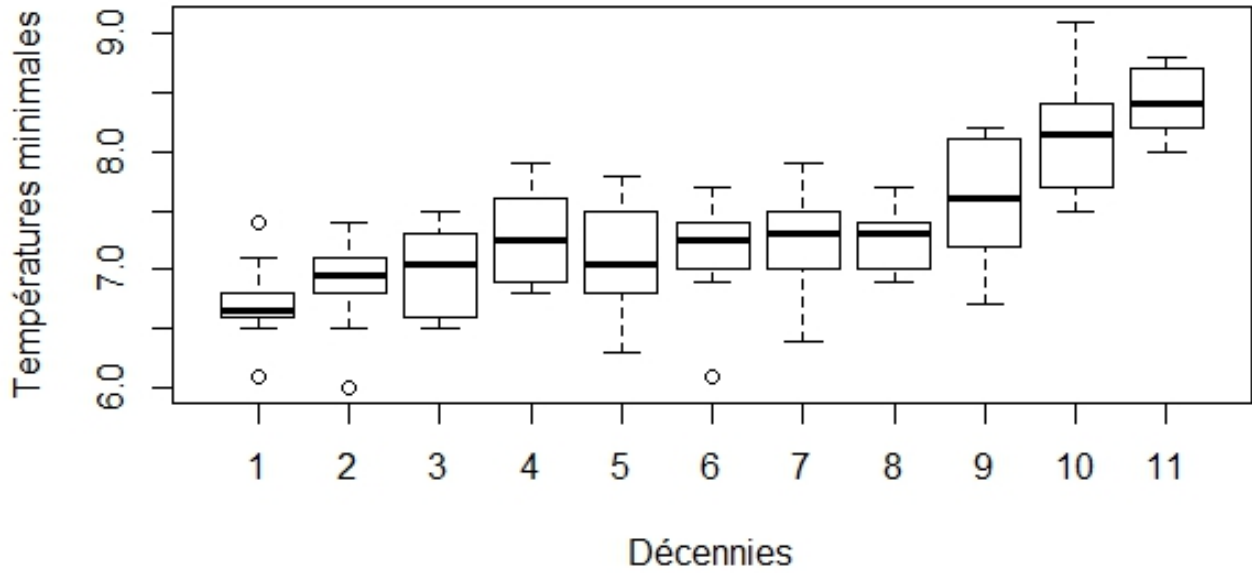


FIGURE 5 – Températures minimales

### Boîtes à moustaches

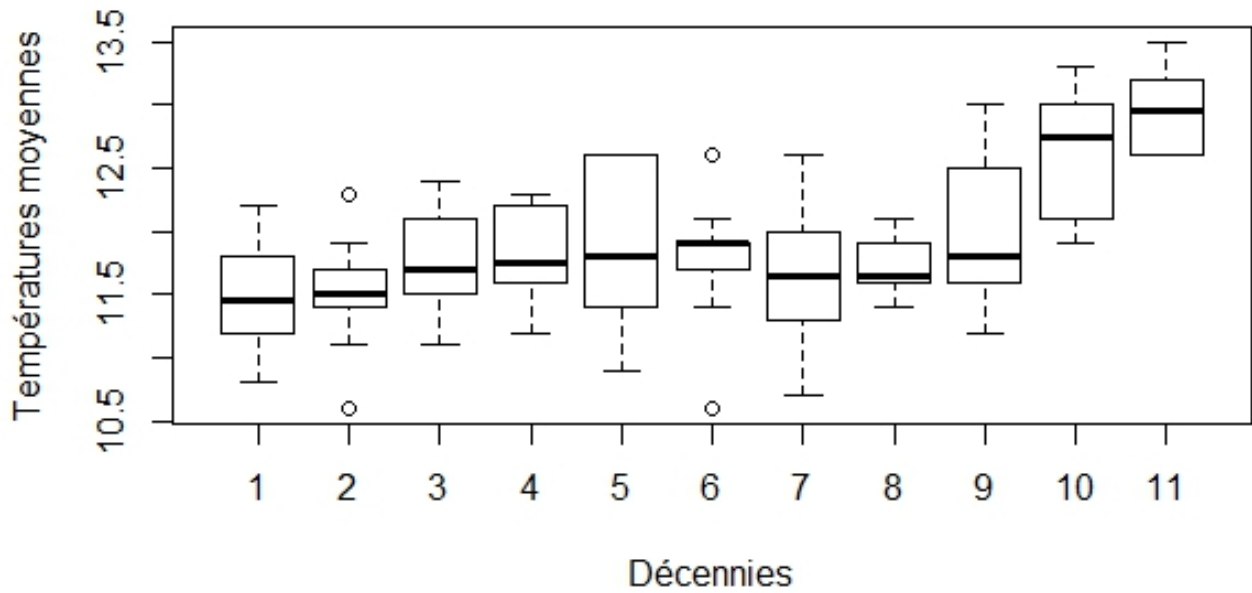


FIGURE 6 – Températures moyennes

## Boîtes à moustaches

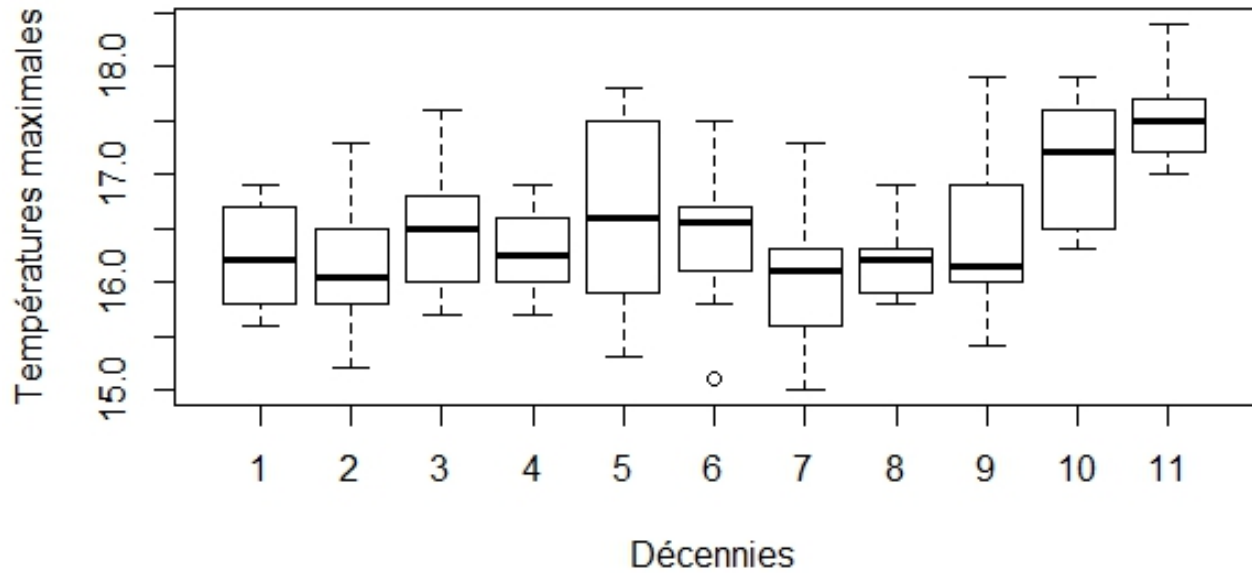


FIGURE 7 – Températures maximales

## Temperature moyenne annuelle de 1900 à 2015 en France

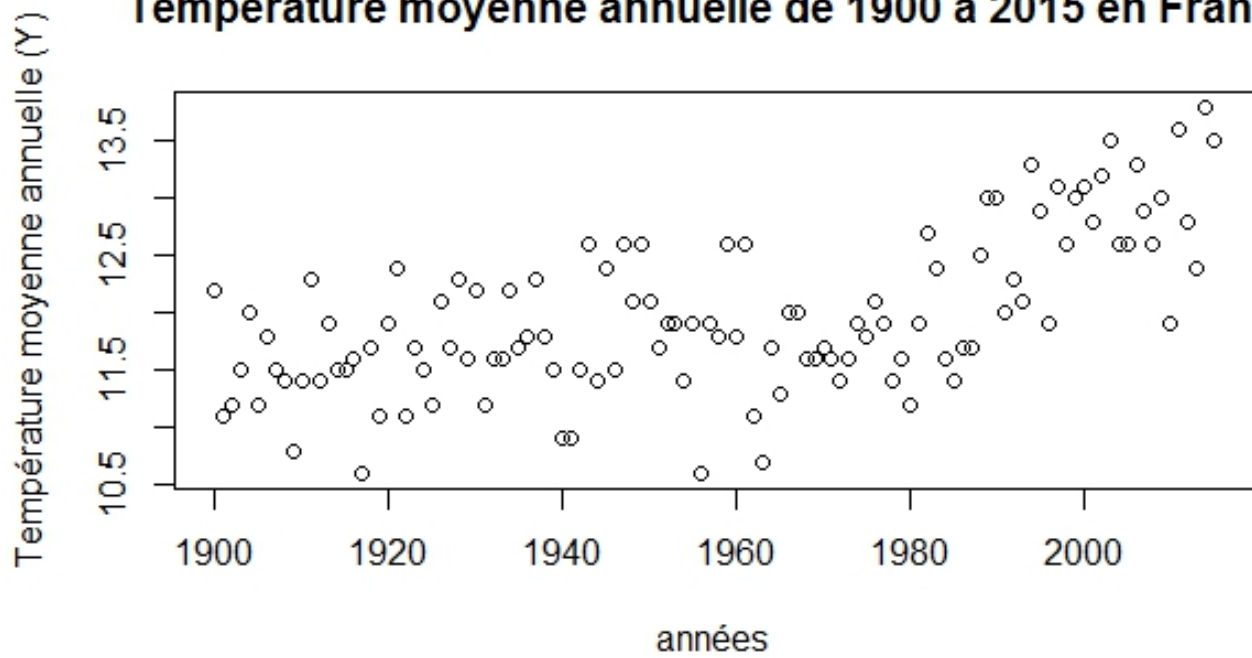


FIGURE 8 – Tracé des températures moyennes annuelles



Bilan :

- Pour une bonne lisibilité, nous avons fait le tracé des boîtes à moustaches sur les décennies pour les températures minimales, moyennes et maximales. On réalise pour les températures minimales que les médianes évoluent en hausse de la première à la onzième décennies. Les valeurs en dehors de la boîte pour certaines décennies sont considérés comme des valeurs aberrantes. On fait la même remarque sur les tracés de boîtes à moustaches pour les températures moyennes et maximales. L'évolution grandissante des médianes dans chaque cas montre clairement que nous sommes en présence d'un réchauffement au fil des décennies.
- Le tracé des températures annuelles moyennes (figure 8) confirme le réchauffement, en effet on peut remarquer ici deux périodes une période entre 1900 et 1980 où la hausse n'est pas visible, et une autre période entre 1981 et 2015 où on remarque une hausse importante de la température.

### 2.1.3 Histogrammes

Dans la suite, on se demande alors qu'elle loi se cache derrière ces données. Dans notre cas, on va vérifier s'il s'agit ici d'une loi normale. Plusieurs techniques nous permettent de faire cette vérification, mais nous allons nous intéresser aux histogrammes.

D'après les histogrammes ci-dessous, on constate que pour les températures minimales et moyennes, 34% de valeurs sont comprises entre 7 et 7.5 degrés pour  $T_{min}$  et entre 11.5 et 12 degré pour les  $T_{moy}$ . Et pour les températures maximales, 30% de valeurs sont comprises entre 16 et 16.5 degré. De plus, les trois histogrammes montrent à chaque fois un mélange de lois normales.

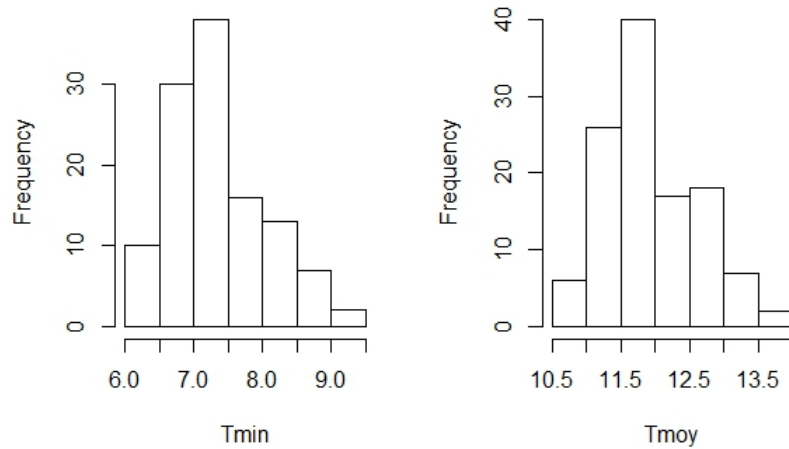


FIGURE 9 – Histogrammes des températures minimales et moyennes

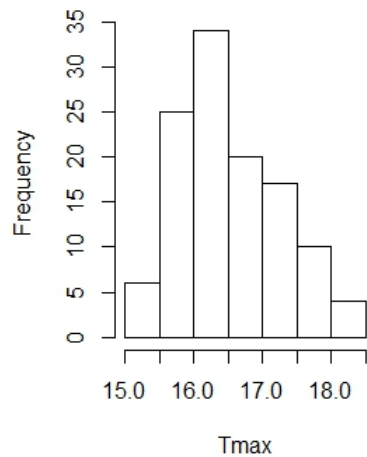


FIGURE 10 – Histogrammes des températures maximales

### 2.1.4 Histogrammes des dérivées discrètes

On rappelle que les dérivées discrètes jusqu'à l'ordre 3 sont définies respectivement par  $DT(i) = T_{i+1} - T_i$ ,  $D^2T(i) = DT(i+1) - DT(i)$ ,  $D^3T(i) = D^2T(i+1) - D^2T(i)$ . Une des méthodes pour transformer une série temporelle en une série normale est de considérer ses dérivées.

Les histogrammes obtenus ici ont un profil qui se rapproche d'une distribution gaussienne.

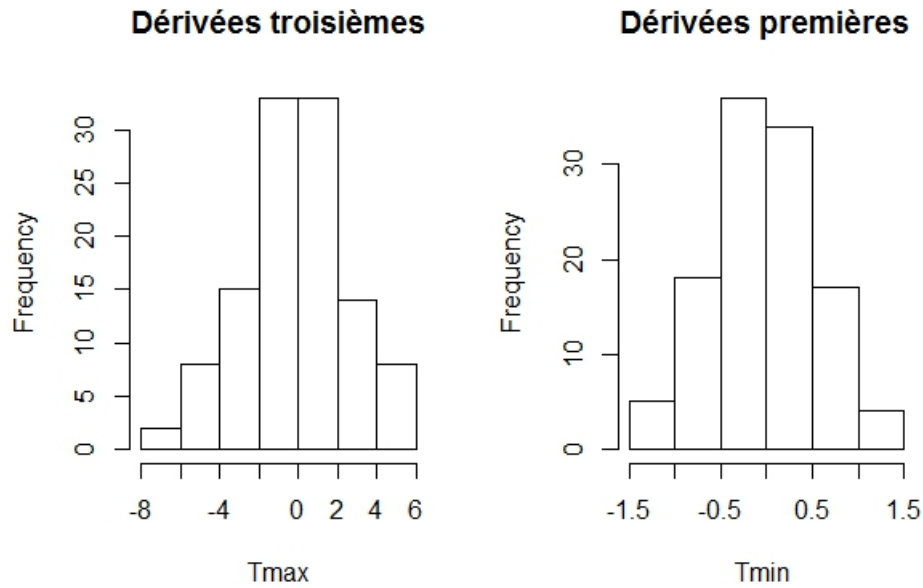


FIGURE 11 – Histogrammes des dérivées discrètes

### 2.1.5 Transformation de Box-Cox

Une autre méthode pour transformer une série temporelle en une série normale est de considérer la trans-

$$\text{formation de Box-Cox : } \varphi_\lambda(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \ln(x) & \text{sinon} \end{cases}$$

$$\text{En effet, } \frac{x^\lambda - 1}{\lambda} = \frac{e^{\lambda \log(x)} - 1}{\lambda} \approx \frac{(1 + \lambda \log(x) + \frac{1}{2} \lambda^2 \log(x)^2 + \dots) - 1}{\lambda} \rightarrow \log(x),$$

quand  $\lambda \rightarrow 0$

On obtient pour différentes valeurs de  $\lambda$  les histogrammes suivants :

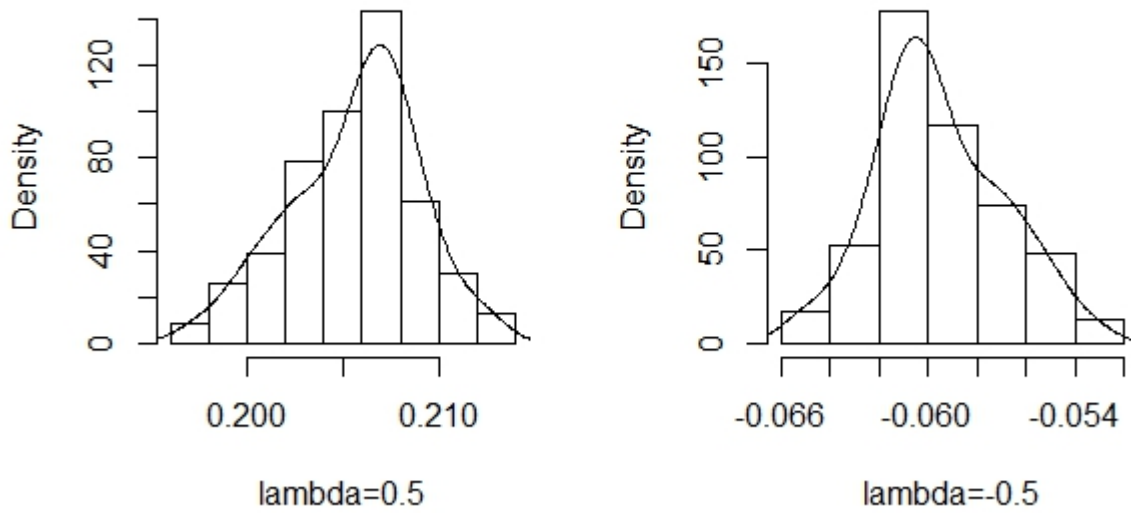


FIGURE 12 – Histogrammes des Tmoy

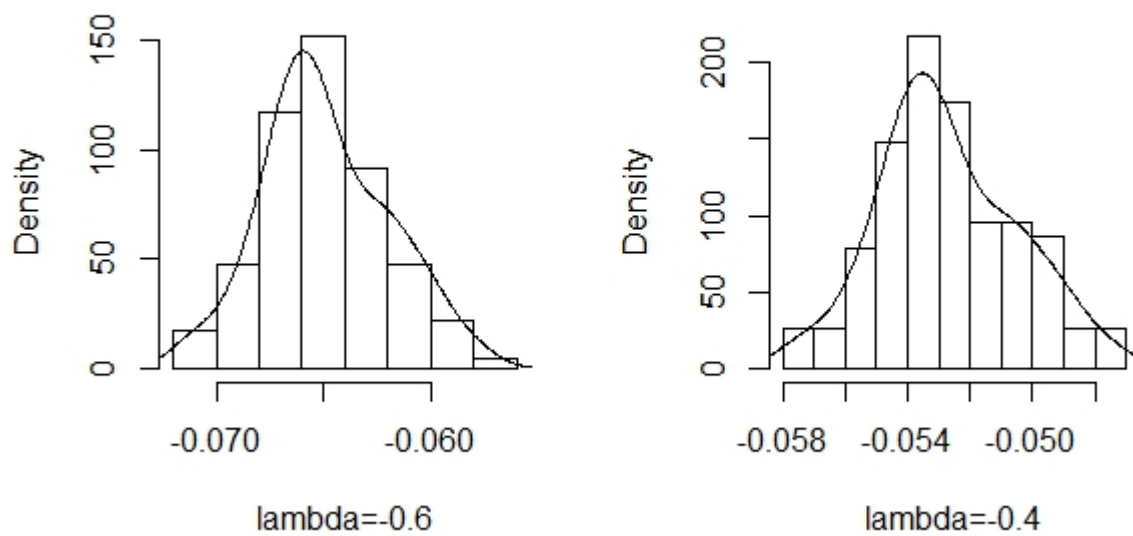


FIGURE 13 – Histogrammes des Tmoy

Après cette transformation, on réalise que chacun des histogrammes montrent également un mélange de lois normales (figure 12 et figure 13).

## 2.2 Distribution des couples de la température des années consécutives

Notons les couples consécutifs  $X = (T_i, T_{i+1})$ ,  $Y = (T_i, T_{i+1})$ ,  $Z = (T_i, T_{i+1})$ ,  $i = 1, \dots, 114$ , respectivement pour les températures minimales, moyennes et maximales.

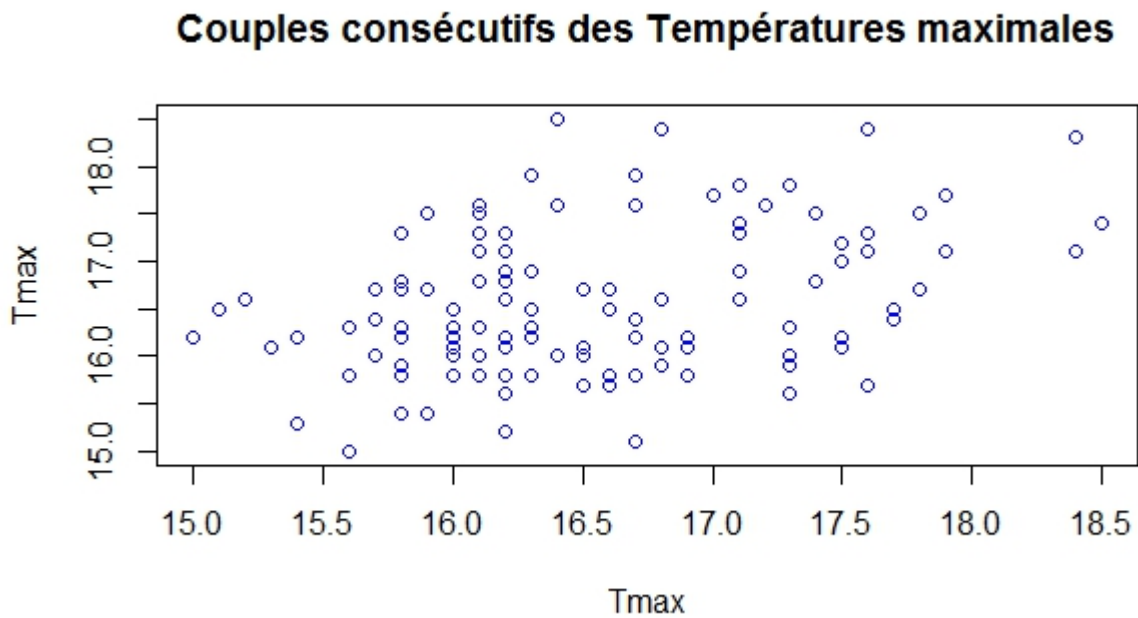


FIGURE 14 – Couples  $X = (T_i, T_{i+1})$ , températures maximales

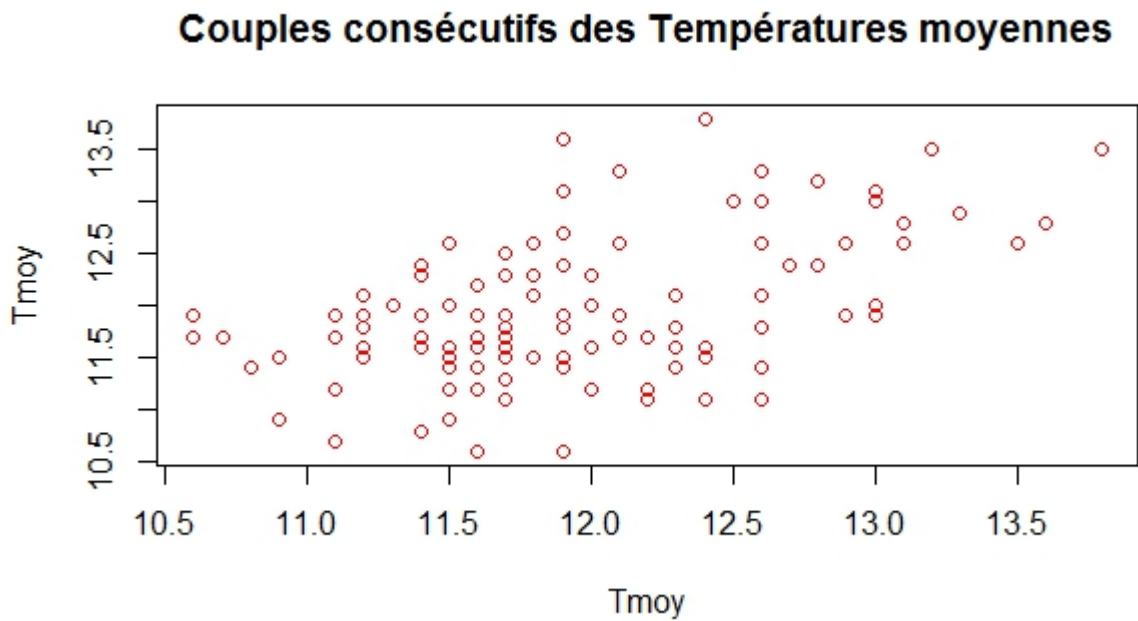


FIGURE 15 – Couples  $Y[T_i, T_{i+1}]$ , températures moyennes

## Couples consécutifs des Températures minimales

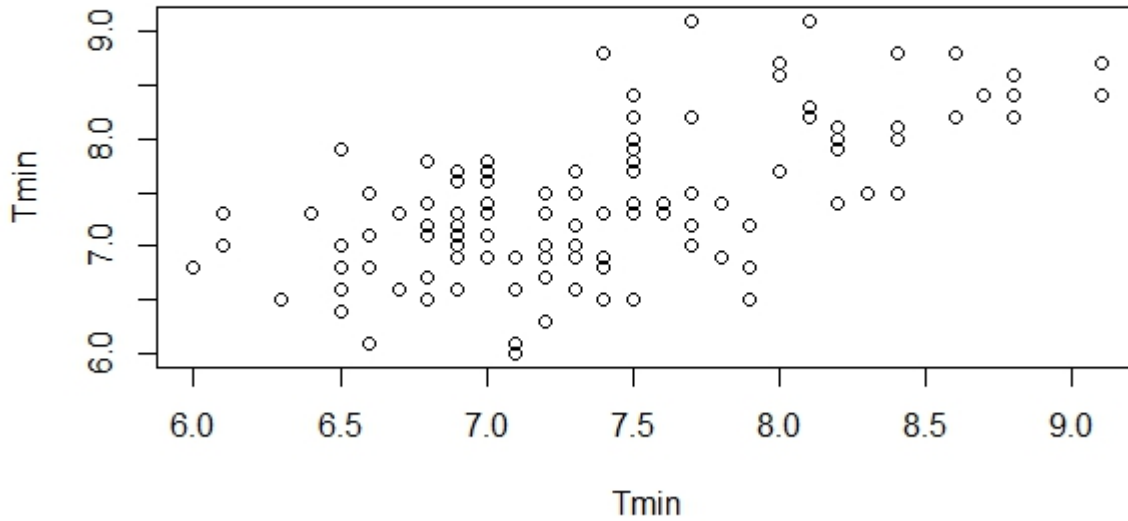


FIGURE 16 – Couples  $Z[T_i, T_{i+1}]$ , températures minimales

A partir de la représentation des couples consécutifs, nous pouvons trouver la loi de probabilité des  $T_i$ , et aussi celle des  $T_{i+1}$  sachant  $T_i$  et par conséquent faire des prévisions et estimé une température avenir.

En se basant par exemple sur le cas des températures minimales, on a :

$$\mathbb{P}(T_i = 7.5) = 1/9$$

$$\mathbb{P}(T_{i+1} = 6.5 | T_i = 7.5) = 1/9$$

$$\mathbb{P}(T_{i+1} = 8 | T_i = 7.5) = 7/9$$

$$\mathbb{P}(T_{i+1} = 8.5 | T_i = 7.5) = 1/9$$

$$\mathbb{P}(T_{i+1} \in [7.5, 8.5] | T_i = 7.5) = 2/3$$

### 2.3 Analyses sur les décennies

L'étude sur les décennies est très intéressante dans la mesure où, il sera alors possible de différencier les décennies froides et les décennies chaudes, et ainsi de dire par exemple si les premières décennies ont été plus froides que les dernières. On considère les 110 premières années, à partir desquelles on construit une matrice de 10 observations et 11 variables( $v_1...v_{11}$ ), chaque variable correspond à une décennie. Notre analyse va donc s'effectuer sur les onze premières décennies comme l'indique le tableau ci-dessous.

	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$	$v_7$	$v_8$	$v_9$	$v_{10}$	$v_{11}$
1	12,2	11.4	11.9	12.2	10.9	12.1	11.8	11.7	11.2	13	13.1
2	11,1	12.3	12.4	11.2	10.9	11.7	12.6	11.6	11.9	12	12.8
3	11.2	11.4	11.1	11.6	11.5	11.9	11.1	11.4	12.7	12.3	13.2
4	11.5	11.9	11.7	11.6	12.6	11.9	10.7	11.6	12.4	12.1	13.5
5	12	11.5	11.5	12.2	11.4	11.4	11.7	11.9	11.6	13.3	12.6
6	11.2	11.5	11.2	11.7	12.4	11.9	11.3	11.8	11.4	12.9	12.6
7	11.8	11.6	12.1	11.8	11.5	10.6	12	12.1	11.7	11.9	13.3
8	11.5	10.6	11.7	12.3	12.6	11.9	12	11.9	11.7	13.1	12.9
9	11.4	11.7	12.3	11.8	12.1	11.8	11.6	11.4	12.5	12.6	12.6
10	10.8	11.1	11.6	11.5	12.6	12.6	11.6	11.6	13	13	13



### 2.3.1 Analyse des corrélations significatives entre variables

Matrice de corrélation											
	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11
v1	1.0000	-0.0400	0.1405	0.7493	-0.4581	-0.4712	0.0794	0.4805	-0.6832	0.1960	0.1045
v2	-0.0400	1.0000	0.4543	-0.6370	-0.4583	-0.2917	0.0699	-0.2717	0.0123	-0.6608	0.0388
v3	0.1405	0.4543	1.0000	-0.1573	-0.3323	-0.2949	0.6378	0.0114	-0.0833	-0.3931	-0.0486
v4	0.7493	-0.6370	-0.1573	1.0000	0.0114	-0.1368	0.0204	0.4760	-0.5421	0.6598	-0.1538
v5	-0.4581	-0.4583	-0.3323	0.0114	1.0000	0.3690	-0.5147	-0.0355	0.4051	0.2176	0.0406
v6	-0.4712	-0.2917	-0.2949	-0.1368	0.3690	1.0000	-0.2796	-0.6122	0.4002	0.4516	-0.0942
v7	0.0794	0.0699	0.6378	0.0204	-0.5147	-0.2796	1.0000	0.3448	-0.3765	-0.0116	-0.3682
v8	0.4805	-0.2717	0.0114	0.4760	-0.0355	-0.6122	0.3448	1.0000	-0.6626	0.1447	0.0155
v9	-0.6832	0.0123	-0.0833	-0.5421	0.4051	0.4002	-0.3765	-0.6626	1.0000	-0.2439	0.2474
v10	0.1960	-0.6608	-0.3931	0.6598	0.2176	0.4516	-0.0116	0.1447	-0.2439	1.0000	-0.5519
v11	0.1045	0.0388	-0.0486	-0.1538	0.0406	-0.0942	-0.3682	0.0155	0.2474	-0.5519	1.0000

FIGURE 17 – Matrice de corrélation

D'après la matrice de corrélation des variables, on observe :

-Des corrélations positives fortes entre les variables v1 et v4, v3 et v7, v4 et v10 ;

La corrélation forte entre v3 et v7 peut s'expliquer par les moyennes de chacune qui sont sensiblement proche à 0.11 degré (voir tableau sur les moyennes)

-Des corrélations négatives fortes entre les variables v1 et v9, v2 et v4, v2 et v10, v8 et v9.

Entre v1 et v9 il y a un important écart de température mais aussi entre v2 et v10 (différence de plus de 1 degrés), ceci peut expliquer la corrélation négative entre ses variables

### 2.3.2 Analyse des moyennes et des variances

Variabes	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$	$v_7$	$v_8$	$v_9$	$v_{10}$	$v_{11}$
Moyenne	11.47	11.50	11.75	11.79	11.85	11.78	11.64	11.70	12.01	12.62	12.96
Variance	0.429	0.452	0.432	0.351	0.691	0.513	0.527	0.226	0.600	0.509	0.316

Les moyennes des variables montrent que la température croit d'une décennie à l'autre dans l'ordre chronologique, malgré une légère perturbation entre  $v_6$  et  $v_8$  qu'on peut considérer ici comme des décennies de transition. De façon globale on peut dire que la température croit au fur et à mesure qu'on avance dans les décennies; De plus, la différence de température entre la représentation graphique des variables  $v_1$  et  $v_{11}$  (figures 15 et 16) corrobore cette évolution de la température au fil des années, en effet on remarque que la température maximale pendant la première décennie est de 12.4 degrés, elle est inférieure à la température minimale obtenue à la onzième décennie, soit 10.8 degrés. Ces résultats viennent appuyer ceux obtenus plutôt comme quoi nous sommes en présence d'une nette hausse de la température.

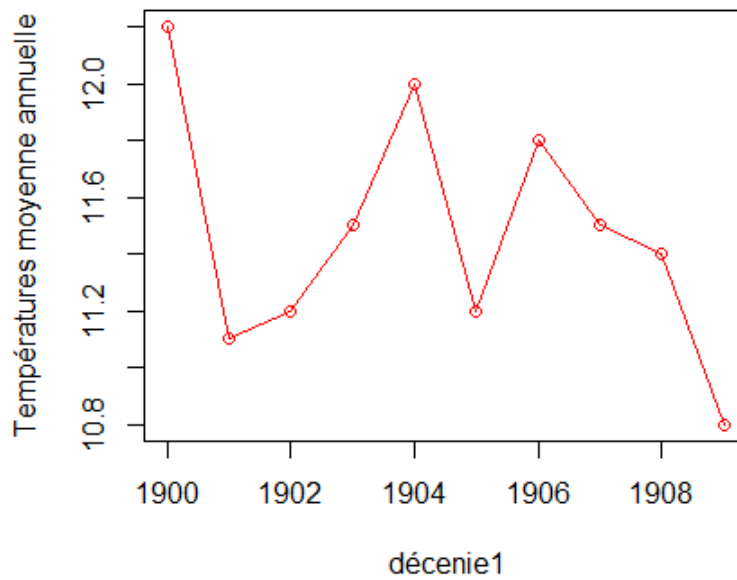


FIGURE 18 – Courbe représentant l'évolution de la température dans la décennie 1

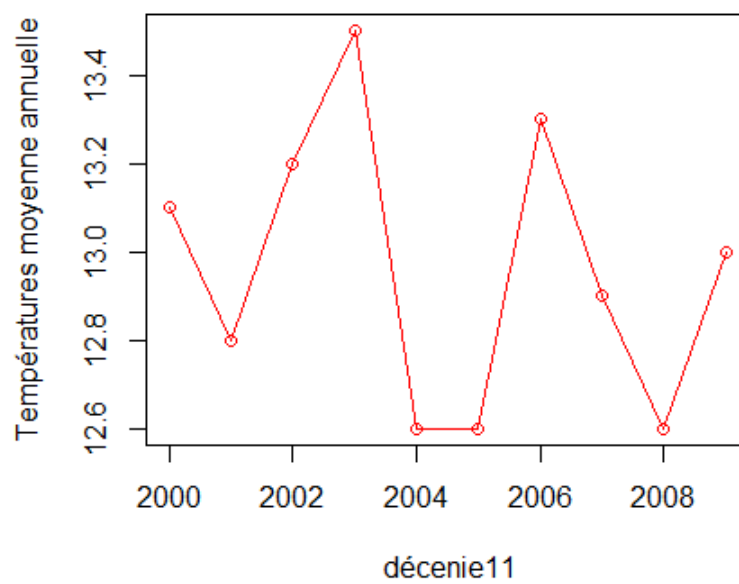


FIGURE 19 – Courbe représentant l'évolution de la température dans la décennie 11

### 3 Analyse en composantes principales

Effectuer une analyse en composantes principales (ACP) sur ce jeu de données (les décennies) permettra de vérifier les résultats obtenus précédemment :

#### 3.1 Valeurs propres

Pour choisir le nombre d'axes à retenir pour cette analyse, on a plusieurs techniques à savoir le critère du coude (éboulis des valeurs propres), la règle de Kaiser, et aussi le diagramme des valeurs propres.

Valeurs propres de la matrice de corrélation				
	Valeur propre	Différence	Proportion	Cumulé
1	3.54170608	0.50500714	0.3220	0.3220
2	3.03669894	1.40006066	0.2761	0.5980
3	1.63663828	0.64580341	0.1488	0.7468
4	0.99083487	0.24761756	0.0901	0.8369
5	0.74321731	0.21370382	0.0676	0.9045
6	0.52951349	0.16082835	0.0481	0.9526
7	0.36868513	0.22760634	0.0335	0.9861
8	0.14107879	0.12945169	0.0128	0.9989
9	0.01162711	0.01162711	0.0011	1.0000
10	0.00000000	0.00000000	0.0000	1.0000
11	0.00000000		0.0000	1.0000

FIGURE 20 – Valeurs propres

On remarque ici que la première valeur propre explique à elle seule environ 32% de l'inertie totale (variance totale), et les quatre premiers axes eux expliquent 84% de l'inertie totale.

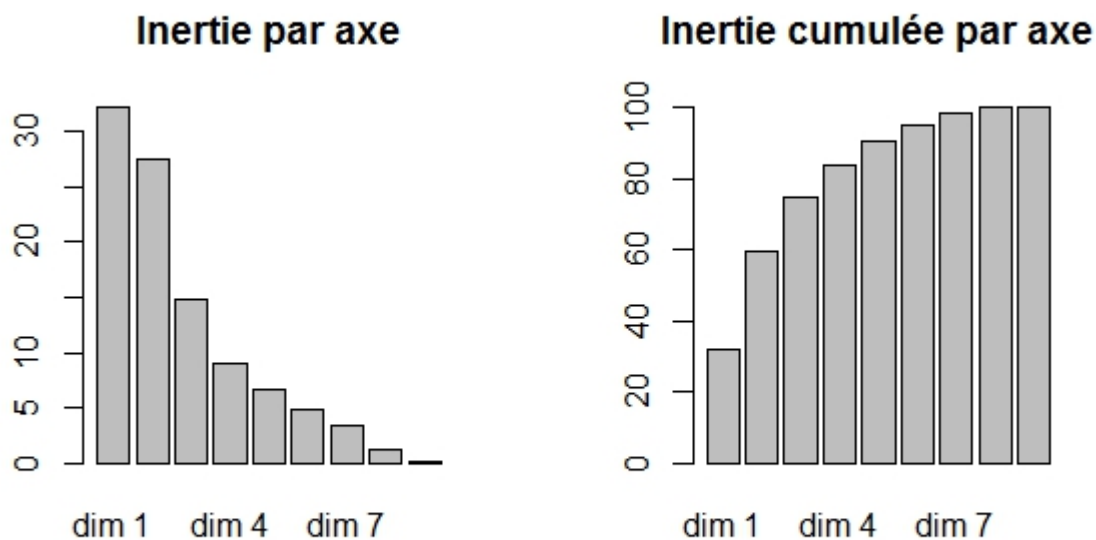


FIGURE 21 – Valeurs des inerties et des inerties cumulées

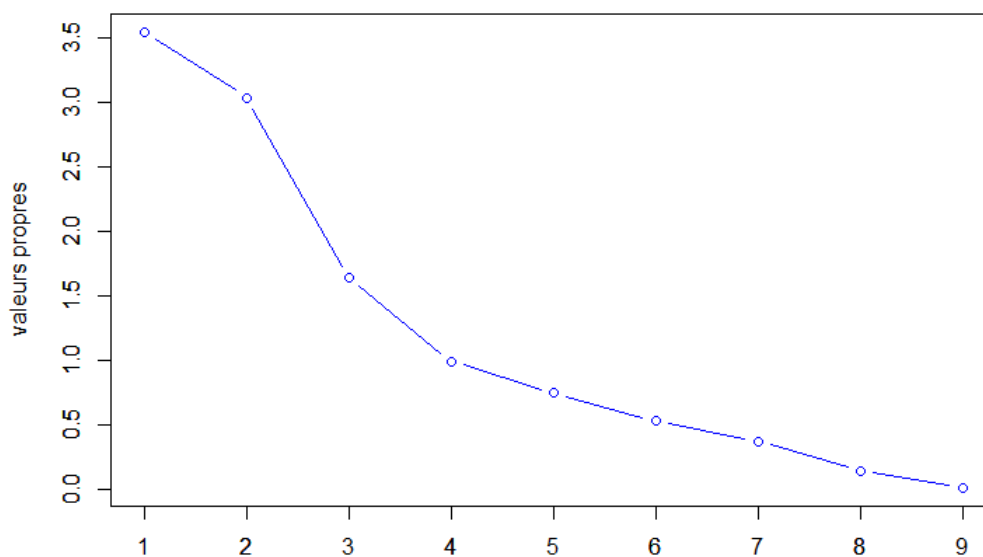


FIGURE 22 – Éboulement des valeurs propres

Les valeurs des inerties et des inerties cumulées, et aussi l'éboulement des valeurs propres (règle du coude) observés dans les figures 9 et 10, ou même la règle de Kaiser (valeur propre :  $\lambda \geq 1$ ) nous incite ici à retenir les quatre premiers axes factoriels.

Dans la suite on va étudier ces axes et aussi la qualité de représentation de nos variables sur ces axes.

## 3.2 Axes factoriels

### 3.2.1 Description des axes factoriels par les variables

<i>Dim.1</i>	$v_1$	$v_8$	$v_4$	$v_9$
corr	0.82	0.76	0.71	-0.87
contrib	19.2	16.5	14.5	21.4
$\cos^2$	0.68	0.60	0.51	0.75

<i>Dim.2</i>	$v_{10}$	$v_3$	$v_2$	<i>Dim.3</i>	$v_{11}$
corr	0.84	-0.69	-0.84	corr	-0.8181465
contrib	23.20	15.34	23.35	conttib	40.90
$\cos^2$	0.70	0.62	0.73	$\cos^2$	0.67

#### -Premier axe factoriel

Les variables  $v_9, v_1, v_4$  et  $v_8$  contribuent respectivement de façon significative à la construction du 1<sup>er</sup> axe, et d'après les valeurs de leur  $\cos^2$  proche de 1, on peut dire que ces variables sont bien représentées sur cet axe ; ceci ce confirme par leur représentation dans le cercle de corrélation (figure 20) où ces variables sont assez proches du bord du cercle, de plus  $v_9$  est corrélée négativement de façon significative à l'axe 1 tant dis que  $v_1$  est corrélée positivement à l'axe 1.  $v_4$  et  $v_8$  sont également corrélées positivement avec le premier axe. On remarque ici que l'axe 1 oppose les premières décennies aux dernières décennies, avec  $v_8$  comme décennie de transition ; ce qui valide la remarque faite plus haut à partir des moyennes, on voit aussi que les décennies corrélées positivement avec l'axe 1 ont plutôt une température moyenne autour de 11.50 degrés et celles corrélées négativement ont une température moyenne au dessus de 12 degrés. On peut donc conclure que l'axe 1 oppose les décennies froides (à droite du cercle) au décennies plus chaudes (à gauche du cercle).

#### -Deuxième axe factoriel

Sur le 2<sup>ieme</sup> axe principal,  $v_2, v_{10}, v_3, v_4$  et  $v_5$  ont une forte contribution

Cet axe est corrélé avec les variables  $v_{10}$  positivement et avec les variables  $v_2$  et  $v_3$  négativement

Seules les variables  $v_{10}$  et  $v_2$  sont bien représentées dans le cercle de corrélation de part leur  $\cos^2$  proche de 1 (figure 21) donc l'axe 2 semble également opposé les premières décennies aux dernières décennies. mais ici les décennies corrélées positivement ont plutôt une température moyenne au dessus de 12 et celles corrélées négativement ont une température moyenne autour de 11.50.

#### -Troisième et quatrième axes factoriels

Seule  $v_{11}$  est corrélée négativement avec l'axe 3 et aucune des variables n'est bien représentée sur l'axe 4 ; il n'est pas possible de faire une interprétation de ces axes.

**Variables factor map (PCA)**

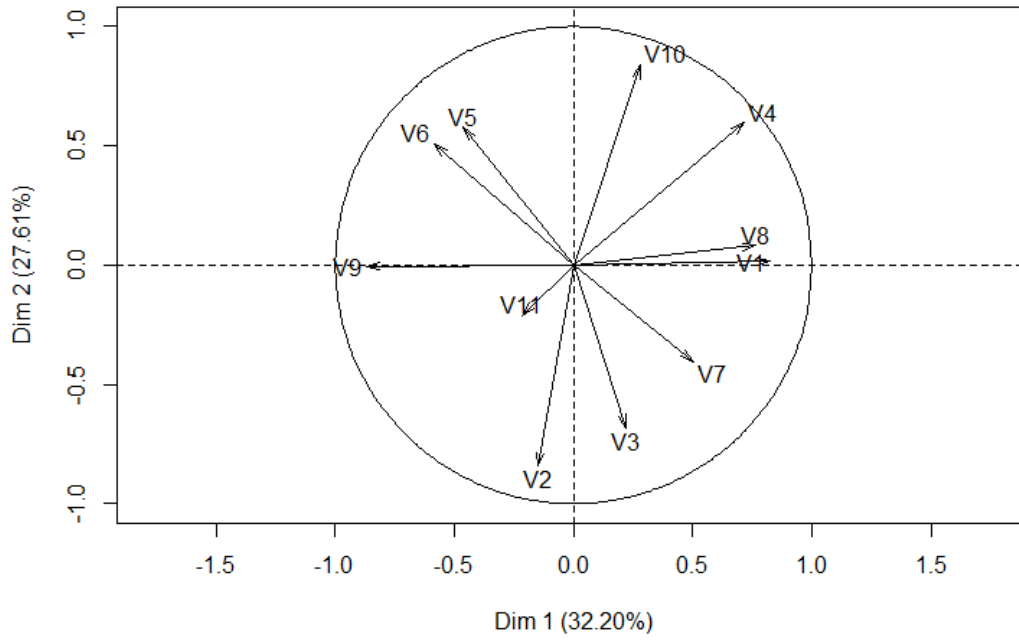


FIGURE 23 – Variables dans le plan(1,2)

**Variables factor map (PCA)**

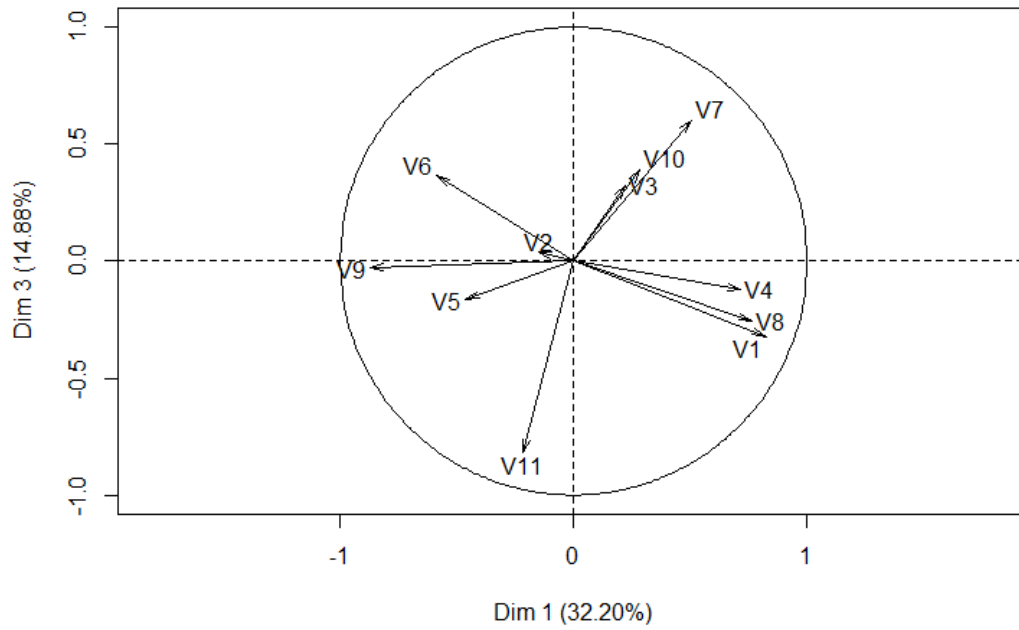


FIGURE 24 – Variables dans le plan(1,3)

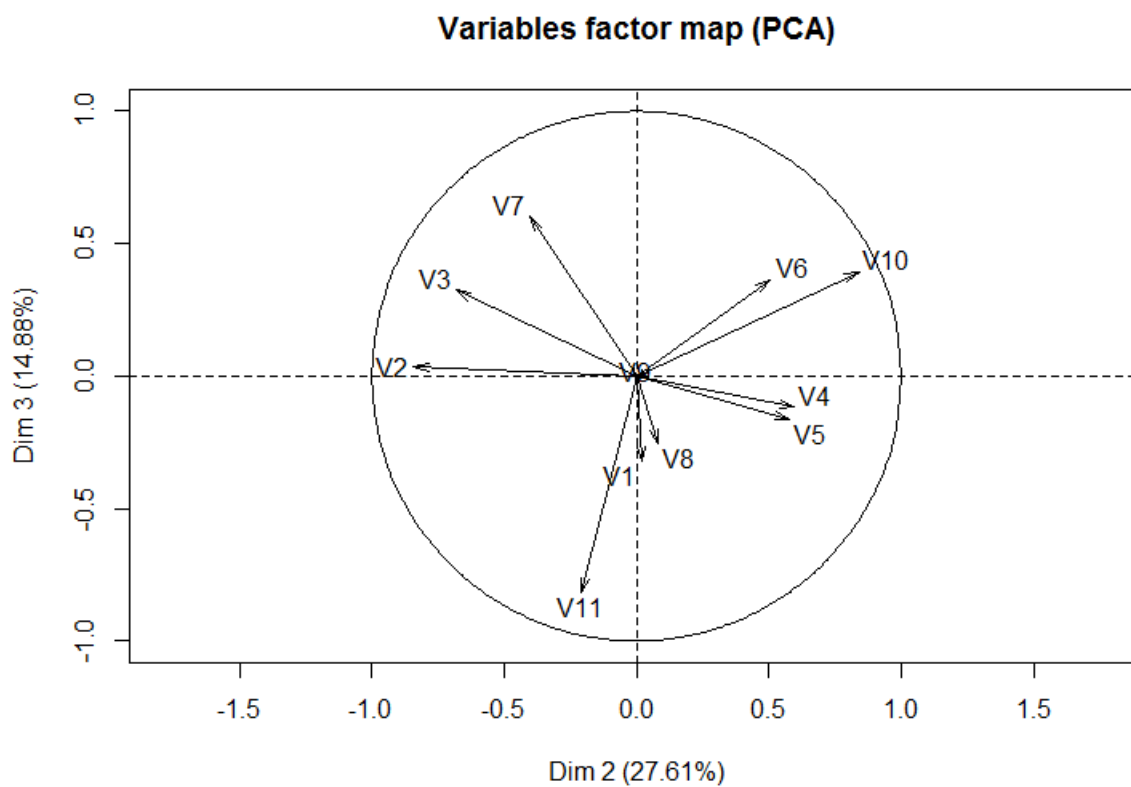


FIGURE 25 – Variables dans le plan(2,3)

En conclusion, après avoir effectué une ACP sur les 11 premières décennies soit de 1900 à 2009, l'analyse des axes a permis de conclure que les décennies récentes sont plus chaudes que les plus anciennes, donc on à assiste à une hausse de la température entre France sur cette période et par conséquent sur les 116 années.



## 4 Régression linéaire

La régression linéaire permet de calculer la tendance linéaire contenu dans la température annuelle moyenne. Elle permet de représenter la température comme la somme d'une partie linéaire et d'un résidu, c'est-à-dire sous la forme  $T_i = \beta_o + \beta_1 i + e_i$ .

Les calculs donnent  $\beta_o = -12.21$ ,  $\beta_1 = 0.01$ ,  $R^2 = 0.371$  et  $R^2_{ajust} = 0.3655$ . **Ainsi la tendance linéaire ne contient que 37% du signal température. Le reste se trouve dans les résidus. Pourtant la tendance linéaire reflète aussi le réchauffement (pente positive  $\beta_1 = 0.01$ ).**

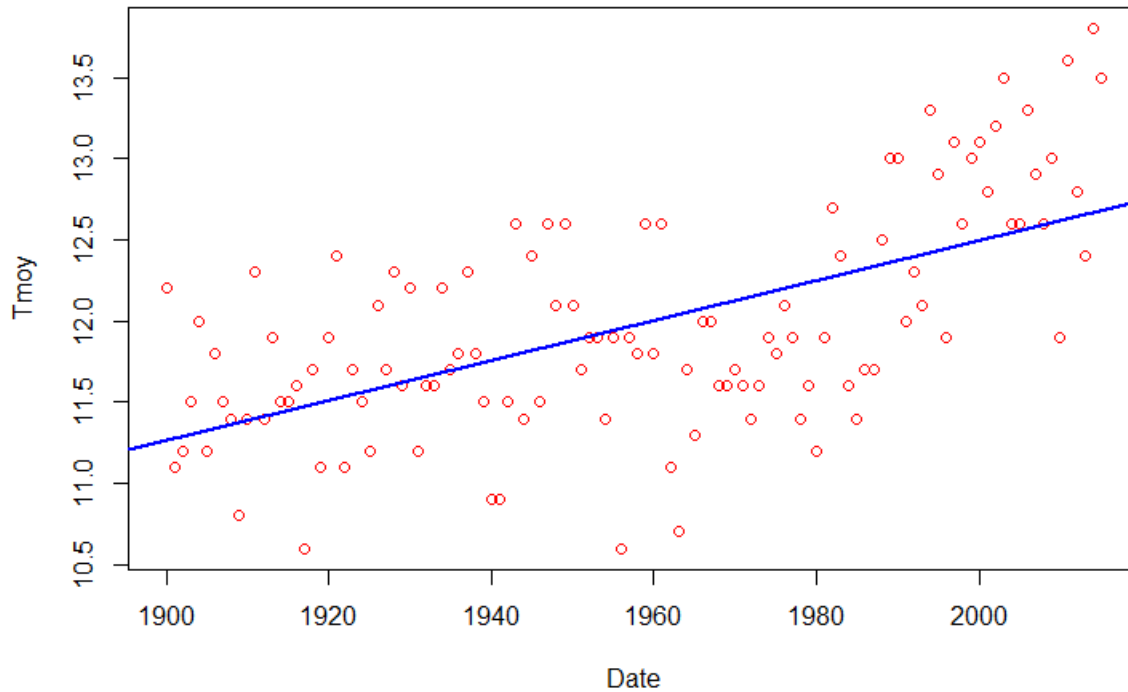


FIGURE 26 – Températures moyenne annuelle de 1900 à 2015

#### 4.1 Régression linéaire sur la période de 1900 à 1980

$$Y_i^1 = \beta_o^1 + \beta_1^1 x_i + e_i^1,$$

On obtient :  $\beta_o^1 = 5.84$ ,  $\beta_1^1 = 0.003$ ,  $R^2 = 0.02$  et  $R^2_{ajust} = 0.009$ .

Les résultats montrent un réchauffement plus faible que la période globale.

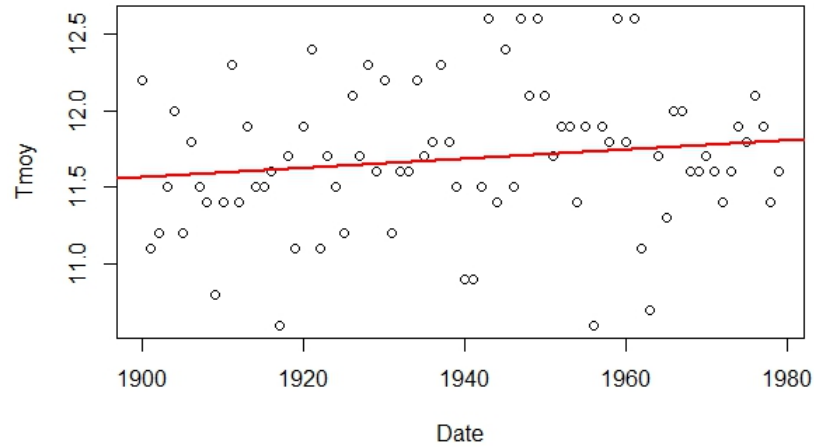


FIGURE 27 – Températures moyenne annuelle de 1900 à 1980

#### 4.2 Régression linéaire sur la période de 1981 à 2015

$$Y_i^2 = \beta_o^2 + \beta_1^2 x_i + e_i^2,$$

On obtient :  $\beta_o^2 = -64.18$ ,  $\beta_1^2 = 0.04$ ,  $R^2 = 0.38$  et  $R^2_{ajust} = 0.036$ . Les résultats montrent un réchauffement plus important.

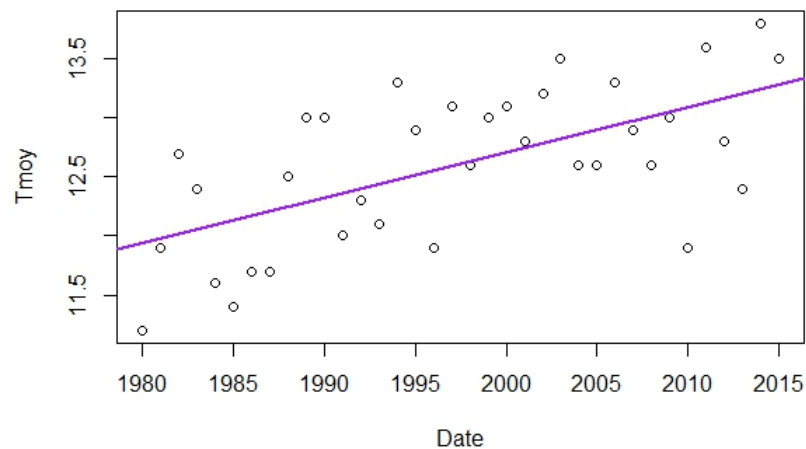
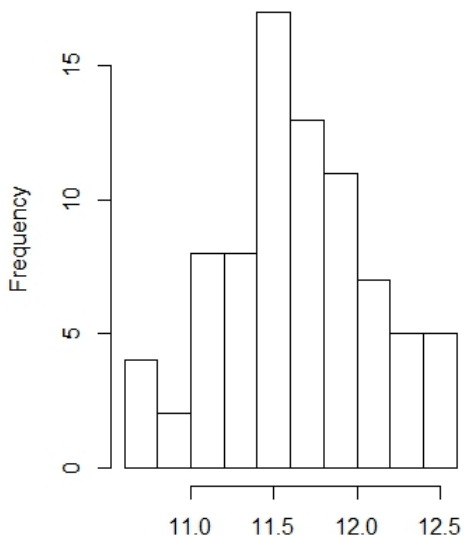


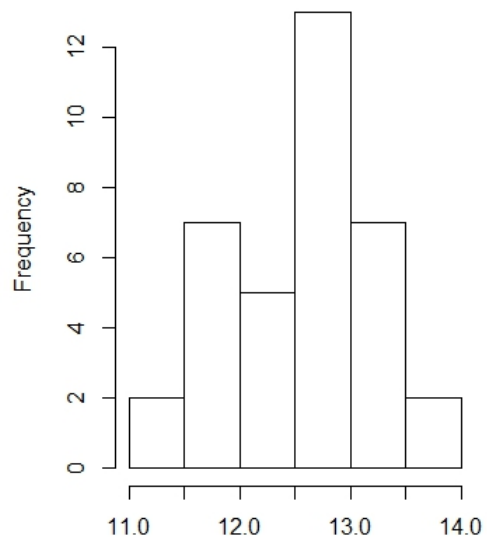
FIGURE 28 – Températures moyenne annuelle de 1980 à 2015

**Températures annuelles moyennes**



Histogramme de la période 1900 à 1980

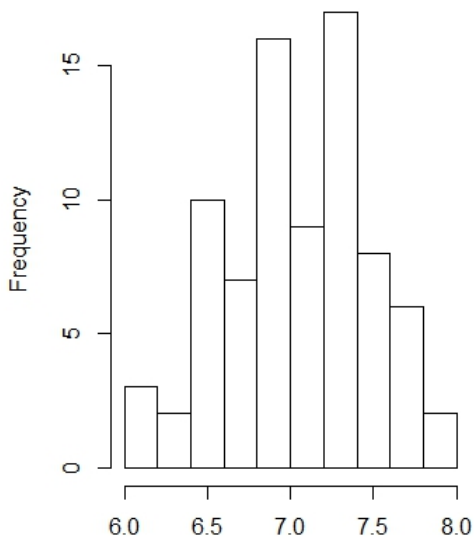
**Températures annuelles moyennes**



Histogramme de la période 1981 à 2015

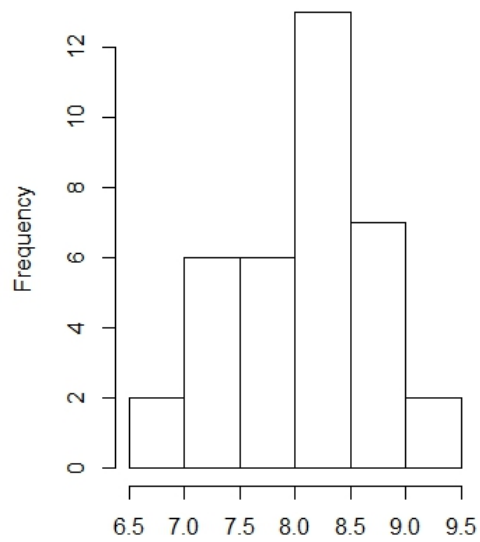
FIGURE 29 – Températures moyenne annuelle

**Températures annuelles minimales**



Histogramme de la période 1900 à 1980

**Températures annuelles minimales**



Histogramme de la période 1981 à 2015

FIGURE 30 – Températures minimales

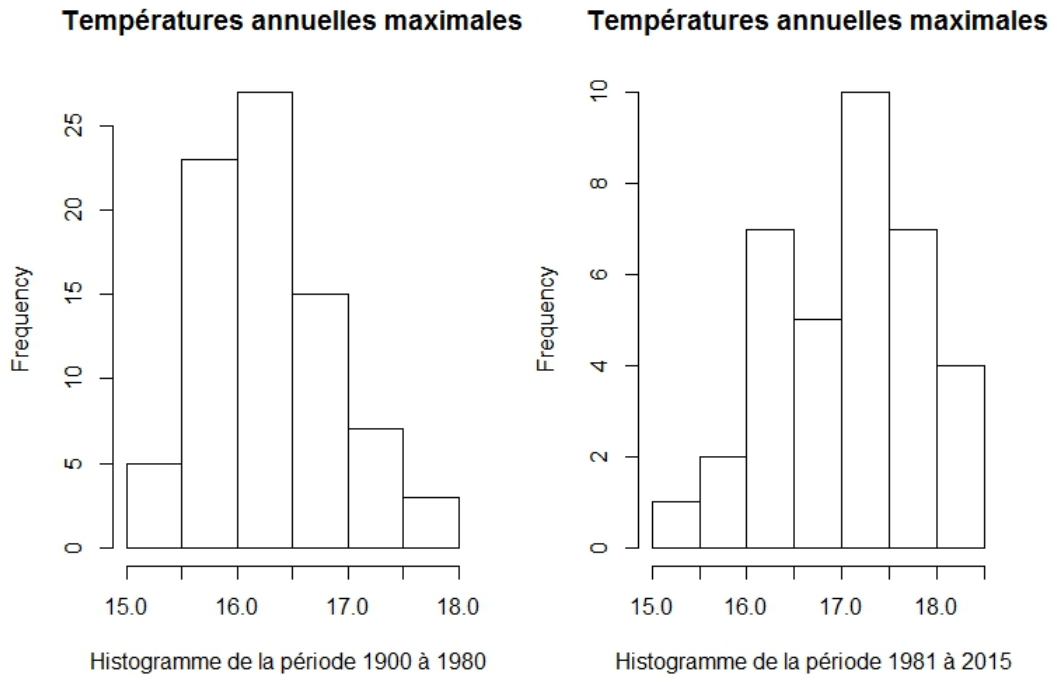


FIGURE 31 – Températures maximales

Ces résultats confirment le fait qu'on assiste à une nette augmentation de la température à partir des années 1980. L'analyse des histogrammes (figures 30, 31 et 32) montre que à chaque fois le mode de la série augmente lorsqu'on passe de la première à la seconde période. En effet en prenant l'exemple sur les températures moyennes, le mode entre 1900 et 1980 est de 11.5, tandis que entre 1981 et 2015 le mode est compris entre [12.5, 13]. A travers la valeur du mode qui augmente sur ces 2 périodes, on conclut également qu'on assiste à une hausse de la température lorsqu'on passe de la première à la seconde période.

## 5 Splines cubiques

Nous avons montré que les droites ne peuvent expliquer que 37% de la dynamique de la température. Dermoune et Preda (2017) ont proposé d'expliquer cette dynamique à l'aide de l'espace des splines cubiques. En effet une spline cubique est représentée par une courbe qui a une forte variabilité avec laquelle on peut par exemple relier la température entre deux années consécutives. Durant une année on modélise la température par un polynôme de degré 3. Quand on change d'année, on change également de polynôme.

On souhaite à partir d'une spline cubique  $s_i$  faire une estimation de la température  $T_i$  sur un intervalle continue  $[0, 115]$ , en utilisant les températures annuelles moyennes observées, avec  $i = 1, \dots, n$ .

Plus précisément Dermoune et Preda ont proposé le modèle suivant :

$T_i = s(i) + e_i$  avec  $s$  une spline cubique aléatoire dont les noeuds sont les années observées

( $t_1 = 0, \dots, t_{n+1} = 115$ ,  $n = 115ans$ ), et  $e_i$  les erreurs liées aux mesures.  $t_1$  correspond à la température observée en 1900,  $t_2$  à celle de 1901 et ainsi de suite jusque  $t_{n+1}$  qui correspond à la température observée en 2015.

### 5.1 Définition

L'espace  $S = S_3(t_1, \dots, t_{n+1})$  des splines cubiques définies par les noeuds  $t_1 < \dots < t_{n+1}$  est un sous espace de  $C^2([t_1, t_{n+1}])$  (espace des fonctions de classe  $C^2$  sur l'intervalle  $[t_1, t_{n+1}]$ ). Un élément  $s \in S$  est un polynôme de troisième degré sur chaque intervalle  $[t_i, t_{i+1}[$ , c'est-à-dire  $s(t) = p_i + (t - t_i)q_i + (t - t_i)^2u_i/2 + (t - t_i)^3v_i/6$  pour chaque  $t \in [t_i, t_{i+1}[$ .

Les constantes  $p_i = s(t_i)$ ,  $q_i = s'(t_i)$ ,  $u_i = s''(t_i)$ ,  $v_i = s'''(t_i)$ .

La condition  $s \in C^2$  implique les contraintes suivantes :

$$p_i + q_i h_i + u_i h_i^2/2 + v_i h_i^3/6 = p_{i+1}, \quad i = 1, \dots, n,$$

$$q_i + u_i h_i + v_i h_i^2/2 = q_{i+1}, \quad i = 1, \dots, n,$$

$$u_i + v_i h_i = u_{i+1}, \quad i = 1, \dots, n,$$

où  $h_i = t_{i+1} - t_i$ . Il est bien connu que l'ensemble des splines de cubiques  $S$  est un espace vectoriel de dimension  $n + 3$ [1].

#### 5.1.1 Paramétrisation et base d'un espace vectoriel

Soit  $E$  un espace vectoriel de dimension  $p$ . Définir une paramétrisation  $\theta : v \in E \rightarrow (\theta_1(v), \dots, \theta_p(v)) \in R^p$  linéaire est bijective est équivalent à définir une base de  $E$ . En effet à chaque paramétrisation  $\theta$  il existe une unique base  $B(\theta) = (b_1, \dots, b_p)$  de  $E$  telle que  $v = \sum_{i=1}^p \theta_i(v)b_i$ ,  $\forall v \in E$ .

D'après l'ouvrage *Journal of Multivariate Analysis*[2], les auteurs **A. Dermoune** et **C. Preda** ont défini des paramétrisations de l'espace des splines cubiques qu'on explicitera dans la suite.

$$\Theta_{002} = (s(t_1), s(t_2), s''(t_1), \dots, s''(t_{n+1})),$$

$$\Theta_{012} = (s(t_1), s'(t_1), s''(t_1), \dots, s''(t_{n+1})),$$

$$\Theta_{0123} = (s(t_1), s'(t_1), s''(t_1), s'''(t_1), \dots, s'''(t_n))$$

dont les bases correspondantes respectives sont :

$$\mathbf{B}_{002} = (\mathbf{b}_1^{002}, \dots, \mathbf{b}_{n+3}^{002}),$$

$$\mathbf{B}_{012} = (\mathbf{b}_1^{012}, \dots, \mathbf{b}_{n+3}^{012}),$$

$$\mathbf{B}_{0123} = (\mathbf{b}_1^{0123}, \dots, \mathbf{b}_{n+3}^{0123}),$$

Ces bases semblent être un bon choix pour séparer les paramètres fixes  $(p_1, p_2)$ ,  $(p_1, q_1)$  de  $(u_1, \dots, u_{n+1})$  aléatoire.

Les notations 002, 012 et 0123 se justifient par le fait qu'on note :

$$s(t_1) = s^{(0)}(t_1), s(t_2) = s^{(0)}(t_2)$$

$$s'(t_1) = s^{(1)}(t_1)$$

$$s''(t_1) = s^{(2)}(t_1), s''(t_{n+1}) = s^{(2)}(t_{n+1})$$

$$s'''(t_1) = s^{(3)}(t_1), s'''(t_n) = s^{(3)}(t_n).$$

Il en résulte donc que pour  $s \in S$ ,

$$s = s(t_1)\mathbf{b}_1^{002} + s(t_2)\mathbf{b}_2^{002} + \sum_{i=1}^{n+1} s''(t_i)\mathbf{b}_{2+i}^{002}, \text{ dans } \mathbf{B}_{002}$$

$$s = s(t_1)\mathbf{b}_1^{012} + s'(t_1)\mathbf{b}_2^{012} + \sum_{i=1}^{n+1} s''(t_i)\mathbf{b}_{2+i}^{012}, \text{ dans } \mathbf{B}_{012}$$

$$s = s(t_1)\mathbf{b}_1^{0123} + s'(t_1)\mathbf{b}_2^{0123} + s''(t_1)\mathbf{b}_3^{0123} + \sum_{i=1}^{n+1} s''(t_i)\mathbf{b}_{3+i}^{0123}, \text{ dans } \mathbf{B}_{0123}$$

### 5.1.2 Modèle pour la température : Dermoune, Preda

$T_i = s(i) + w_i$ ,  $i = 1, \dots, n+1$ , où  $s(t) = a + bt + \{s(t) - a - bt\}$  est une spline aléatoire Gaussienne. Sa moyenne est la droite  $a + bt$  (appelée effet fixe), la composante  $\{s(t) - a - bt\}$  est de moyenne nulle et sa loi est donnée par

$$\int_{t_1}^{t_{n+1}} |s''(t)|^2 dt = \text{energie de } s'' = \sum_{i=1}^n (u_i + u_i u_{i+1} + u_{i+1}^2) h_i / 3.$$

La signification de  $a, b$  dépend du choix de la paramétrisation. Par exemple  $a = s(t_1)$ ,  $b = s(t_2)$ , ou bien  $a = s(t_1)$ ,  $b = s'(t_1)$ , ou bien  $a = \sum_{i=1}^{n+1} s(t_i)$ ,  $b = \sum_{i=1}^{n+1} s'(t_i)$ .

### 5.1.3 Construction des Bases $\mathbf{B}_{0123}$ , $\mathbf{B}_{012}$ , $\mathbf{B}_{002}$

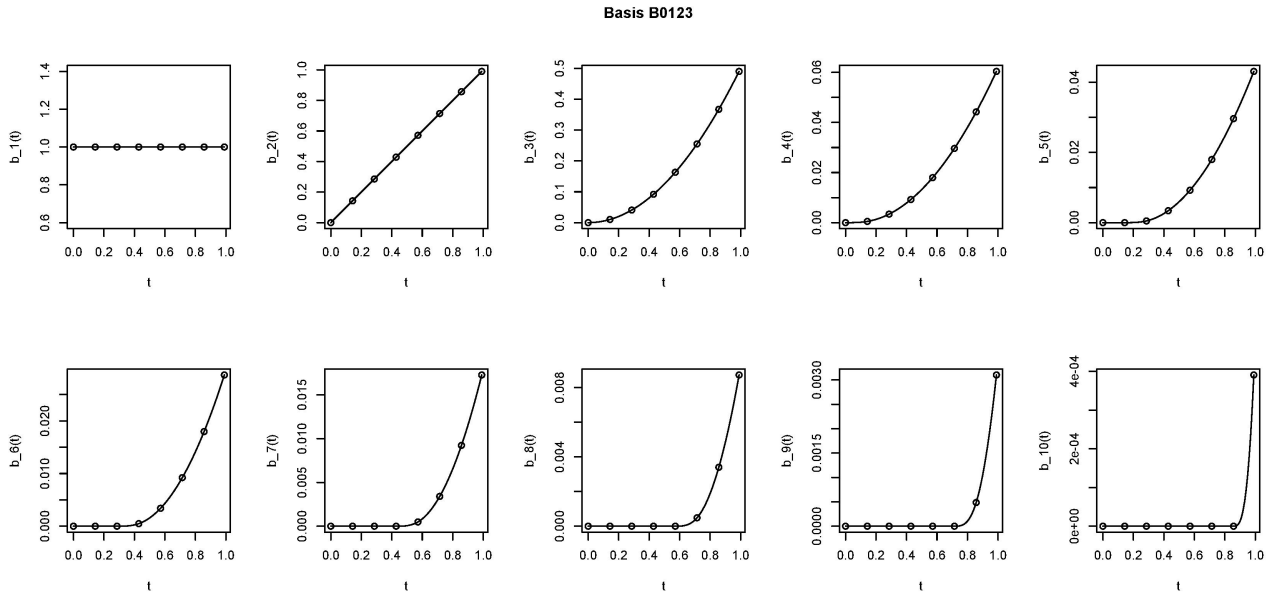


FIGURE 32 – Représentation de la base  $B_{0123}$ , pour  $n = 7$  dans l'intervalle  $[0,1]$

**Basis B012**

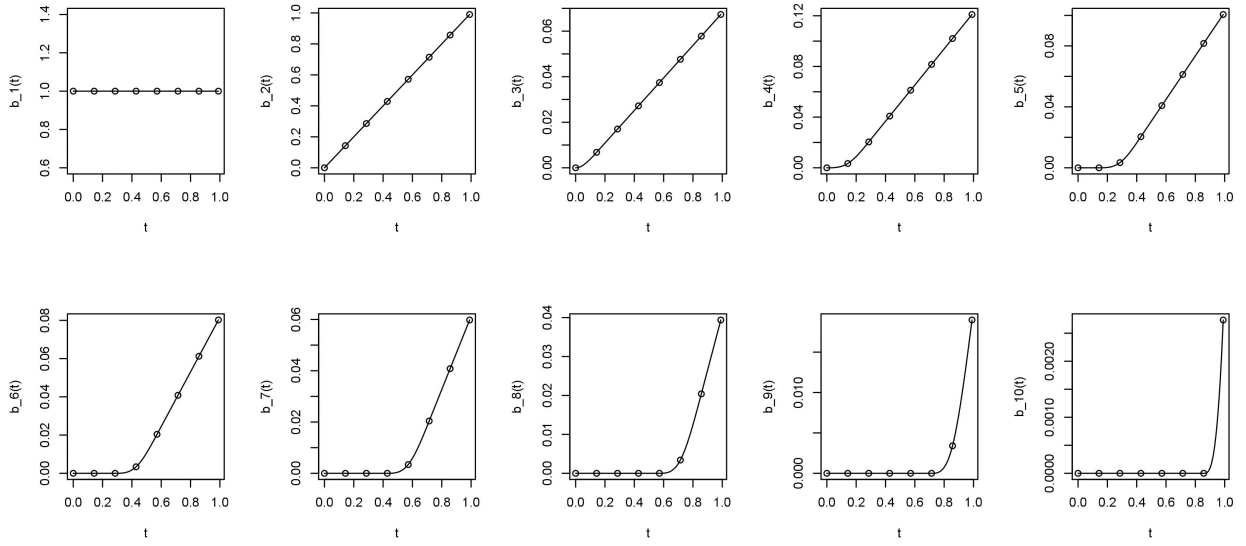


FIGURE 33 – Représentation de la base  $B_{012}$ , pour  $n = 7$  dans l'intervalle  $[0,1]$

**Basis B002**

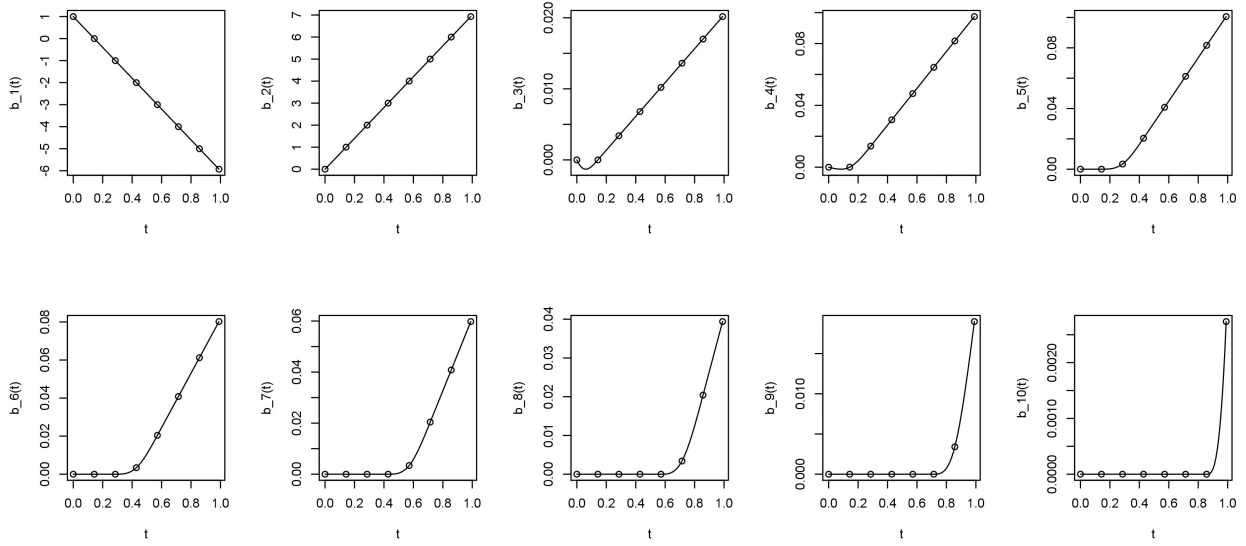


FIGURE 34 – Représentation de la base  $B_{002}$ , pour  $n = 7$  dans l'intervalle  $[0,1]$

### 5.1.4 Construction de la Base $\mathbf{B}_{202}$

$\Theta_{202} = (s''(t_1), s(t_1), \dots, s(t_{n+1}), s''(t_{n+1}))$ , dont la base correspondante s'écrit :  
 $\mathbf{B}_{202} = (\varphi_1, \dots, \varphi_{n+3})$ .

La notation 202 se justifie par  $s''(t_1) = s^{(2)}(t_1), s(t_1) = s^{(0)}(t_1), \dots, s(t_{n+1}) = s^{(0)}(t_{n+1}), s''(t_{n+1}) = s^{(2)}(t_{n+1})$ .  
 La spline cubique  $s$  s'écrit donc dans la base  $\mathbf{B}_{202}$  de façon suivante :

$$s = s''(t_1)\varphi_1 + \sum_{i=1}^{n+1} s(t_i)\varphi_{1+i} + s''(t_{n+1})\varphi_{n+3}$$

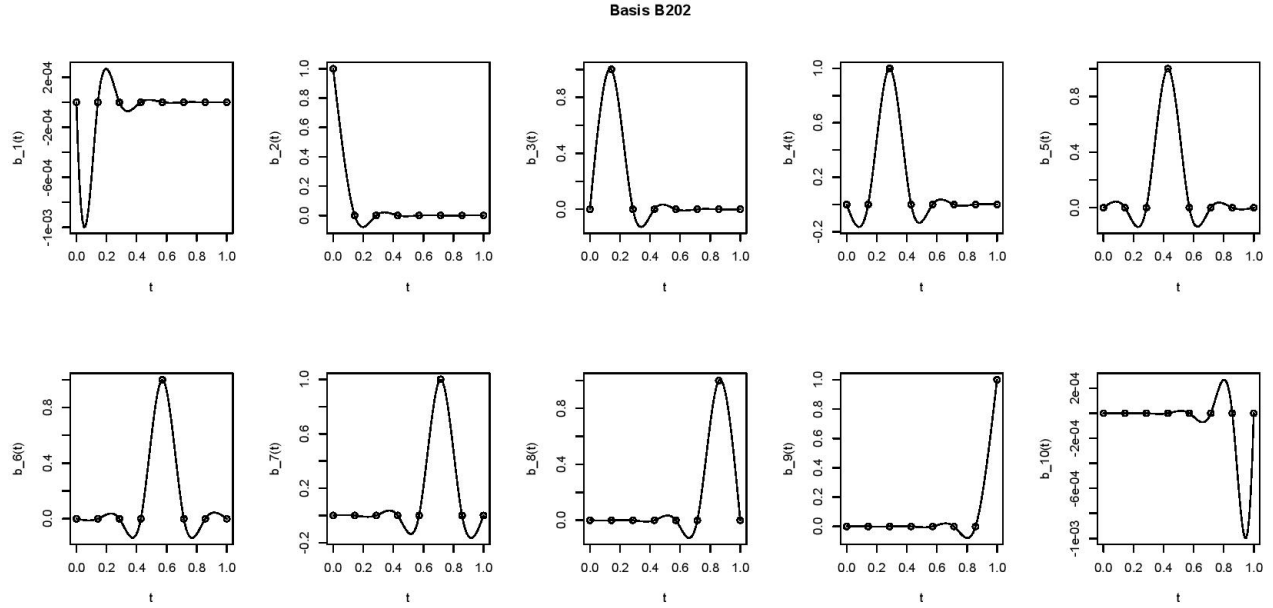


FIGURE 35 – Représentation de la base  $B_{202}$ , pour  $n = 7$  dans l'intervalle  $[0,1]$

D'après la représentation graphique, on remarque bien qu'on n'obtient pas de droite sur aucune des représentations ce qui pose un problème dans la mesure où on souhaite représenter la spline cubique  $s$  dans cette base  $\mathbf{B}_{s202}$ .  
 Pour remédier à ce problème, les auteurs ont modifié cette base en la dérivant pour obtenir la nouvelle base  $\mathbf{B}_{s202}$ .

Après avoir tracé ils ont obtenus la représentation suivante (figure 37) :



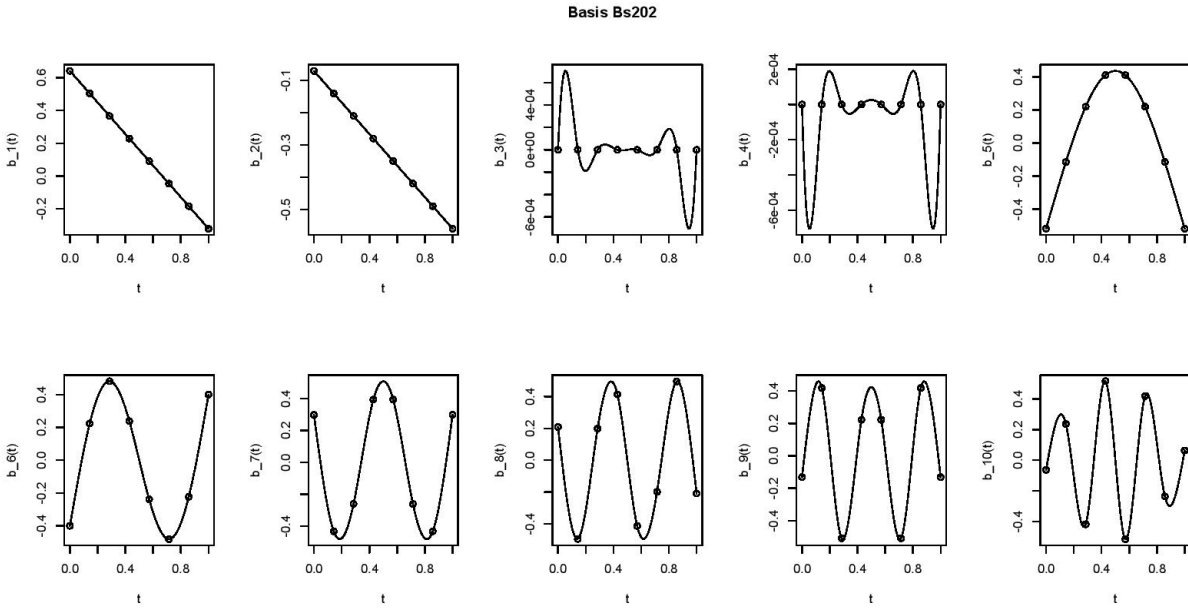


FIGURE 36 – Représentation de la base  $B_{s202}$ , pour  $n = 7$  dans l'intervalle  $[0,1]$   
 Les représentations obtenues sont à présent appropriées pour modéliser au mieux la spline cubique  $s$ .

En représentant dans les bases obtenues la spline cubique  $s$  en utilisant les températures annuelles moyennes observées, ils ont obtenus pour bases  $B_{002}$  et  $B_{s202}$  les représentations suivantes :

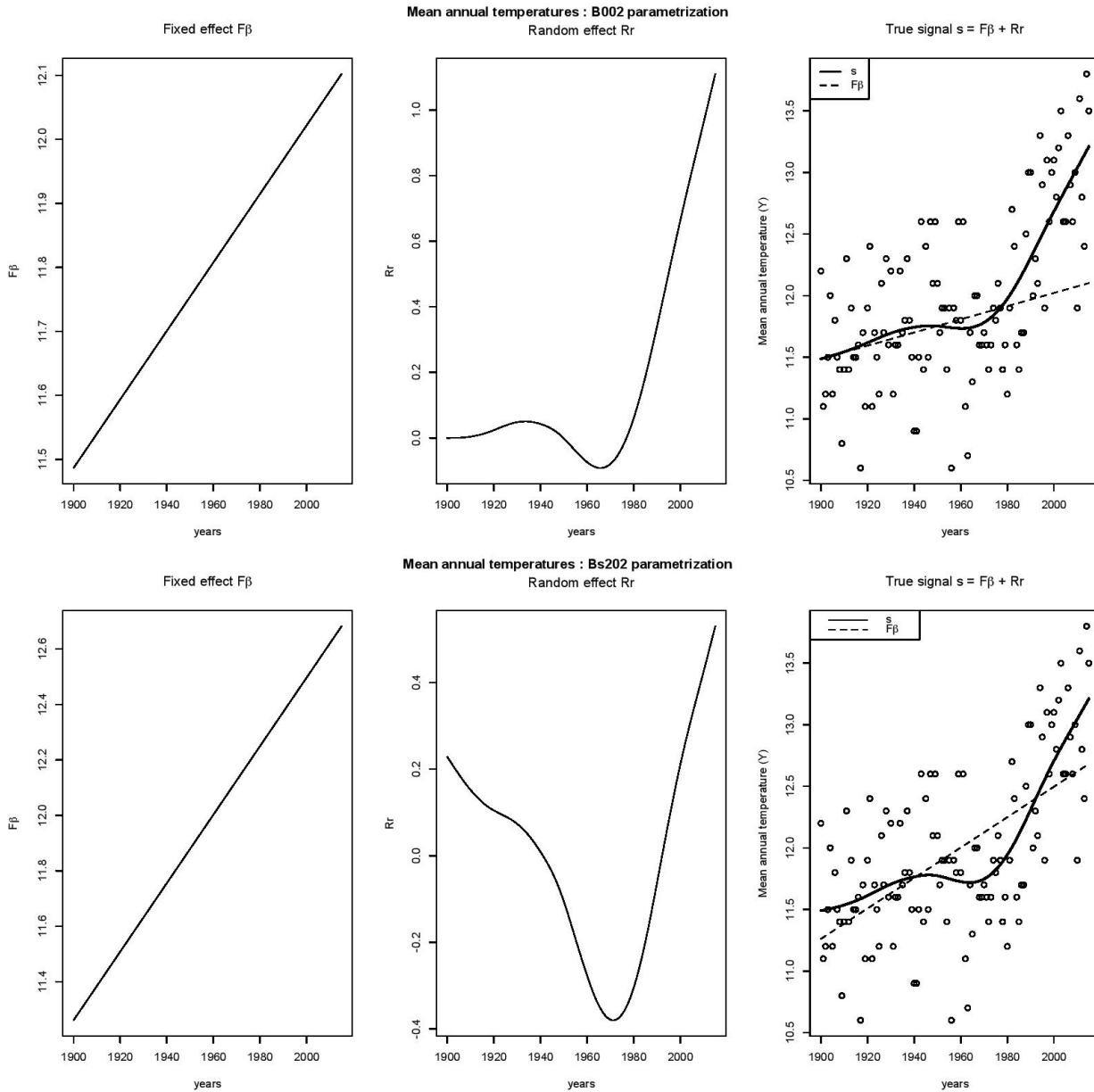


FIGURE 37 – Températures annuelles moyennes : paramétrisation Bs202 et B002

On remarque clairement à partir de ces représentations que la modélisation des températures observées est mieux représentée. Cette modélisation présente bien deux effets pour chacune des bases. En effet, la partie fixe est modélisée par une droite de pente strictement positive qui évolue de façon chronologique ; L'effet aléatoire représenté à partir des courbes qui quant à elles présentent des variations entre 1900 et 1968 environ, mais à partir de 1980 elles croient très rapidement ce qui traduit ici une hausse considérable de la température sur ces dernières années.

## Mean annual temperatures : parametrizations Bs202 and B002

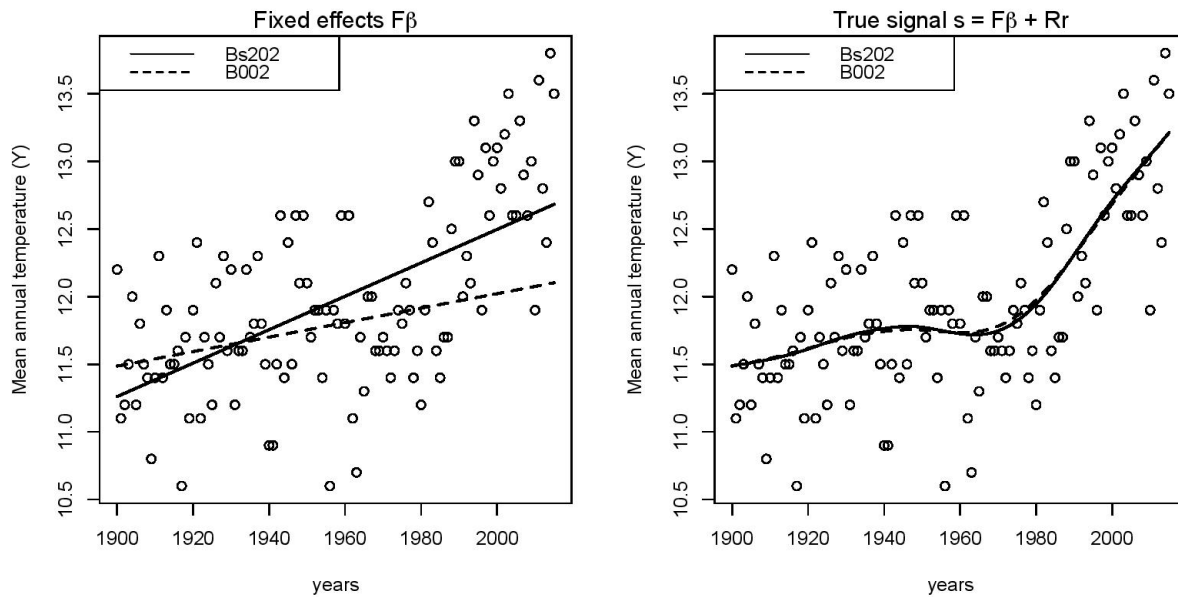


FIGURE 38 – Bs202 versus B002

Après avoir superposé les différentes modélisations pour les deux bases précédentes, on remarque que pour l'effet, les droites n'ont pas la même pente. On en déduit donc que l'effet fixe et l'effet aléatoire distinctement dépendent de la base choisie pour la modélisation, tandis que la somme des deux se superposent parfaitement comme indiqué sur le graphe de droite de la figure 36.

## 6 Conclusion

Suite à notre travail encadré de recherche, nous avons tenté de modéliser et aussi de comprendre comment à évolué la température en France entre 1900 et 2015. Il était question pour nous qu'à partir de plusieurs analyses statistiques qu'on puisse parvenir à faire des modélisations pertinentes qui permettent d'arriver à une même conclusion. Pour cela nous nous sommes attardé sur plusieurs techniques d'analyses statistiques notamment : statistique descriptive, analyse en composante principale, régression linéaire. Il en ressort de toutes ces analyses qu'on assiste bien à un réchauffement net surtout à partir des années 80.

Néanmoins ces techniques se sont révélées inappropriées pour modéliser la température, car la température étant continue et à forte variabilité. Nous avons donc introduit la notion de spline cubique qui s'est avérée très efficace pour la modélisation de la température de part sa grande variabilité. A partir des splines, nous avons donc obtenus des résultats très satisfaisants qui ont également corroborer la conclusion tirée à partir des techniques effectuées précédemment.

Donc à partir des données observées de la température en France entre 1900 et 2015, on peut affirmer sans ambiguïté qu'on nous en présence d'un réchauffement climatique.

## Références

- [1] C.de Boor, A Practical Guide to Splines, Springer, Berlin, 1978.
- [2] A.Dermoune, C.Preda, Journal of Multivariate Analysis 154 (2017) 162-176.

## **Remerciements**

Je tiens à remercier M. Dermoune qui, tout au long de ce travail, a su me guider avec passion vers l'aboutissement de ce projet. Ainsi tout au long des rendez vous, j'ai pris le plaisir à travailler en sa compagnie. De plus, je le remercie d'avoir accepté de formaliser un sujet de TER que j'ai trouvé très intéressant.

## 7 Annexes : code R

Code source TER

```
n = 115
x = mfTmin températures minimales
y = mfTmoytempératuresmoyennes
z = mfTmax températures maximales
```

Moyennes

```
mx = mean(x)
my = mean(y)
mz = mean(z)
```

Variances

```
vx = var(x)
vy = var(y)
vz = var(z)
```

Médianes, 1er quartile et 3ième quartile

```
summary(x)
summary(y)
summary(z)
```

```
Y = mf[4 :4] matrice des Tmoy
Z = mf[3 :3] matrice des Tmax
X = mf[2 :2] matrice des Tmin
```

Boîtes à moustaches

```
boxplot(mf[,1] mf[,2], xlab="Températures minimales", ylab="années", range=0)
boxplot(mf[,1] mf[,3], xlab="Températures maximales", ylab="années", range=0)
boxplot(mf[,1] mf[,4], xlab="Températures moyennes", ylab="années", range=0)
```

```
boxplot(X, main="Boîte à moustaches pour Tmin")
boxplot(Y, main="Boîte à moustaches pour Tmoy")
boxplot(Z, main="Boîte à moustaches pour Tmax")
```

Histogrammes

```
par(mfrow=c(1,2))
n1 = rnorm(115, mx, vx*vx) simulation de loi normale de paramètres la moyenne et l'écart-type de l'échantillon
d1 = density(n1) densité de la loi normale
dex = density(x) densité de l'échantillon
```

```
hist(x, main=" ", xlab="Tmin", freq=FALSE, ylim=c(0,2))
lines(d1, col="red")
lines(dex, col="green")
```

```
n2 = rnorm(115, my, vy*vy) simulation de loi normale de paramètres la moyenne et l'écart-type de l'échantillon
d2 = density(n2) densité de la loi normale
dey = density(y) densité de l'échantillon
```

```
hist(y, main=" ", xlab="Tmoy", freq=FALSE, ylim=c(0,2.5))
lines(d2, col="blue")
lines(dey, col="green")
```

```
n3 = rnorm(115, mz, vz*vz) simulation de loi normale de paramètres la moyenne et l'écart-type de l'échantillon
d3 = density(n3) densité de la loi normale
```

dez=density(z) densité de l'échantillon

```
hist(z,main=" ",xlab="Tmax",freq=FALSE,ylim=c(0,1.5))
lines(d3)
lines(dez,col="green")
plot(d3)
```

suite dérivées premières

```
i=1 :115
X1=X[i+1,]-X[i,]
m1=mean(X1)
s1=sd(X1)
nor1=rnorm(115,m1,s1)
d1=density(nor1)
d11=density(X1)
hist(X1, main="Dérivées premières",xlab = "Tmoy ",freq = FALSE,ylim=c(0,0.8))
lines(d1,col="blue")
lines(d11,col="red")
```

$X11=X1/X[i,] \ ((Ti+1)-Ti)/Ti$  pour les températures moyennes  
hist(X11)

```
Y1=Y[i+1,]-Y[i,]
m2=mean(Y1)
s2=sd(Y1)
nor2=rnorm(115,m2,s2)
d2=density(nor2)
d22=density(Y1) hist(Y1, main="Dérivées premières",xlab = "Tmax ",freq = FALSE, ylim = c(0,0.6))
lines(d2,col="green")
lines(d22)
```

```
Z1=Z[i+1,]-Z[i,]
m3=mean(Z1)
s3=sd(Z1)
nor3=rnorm(115,m3,s3)
d3=density(nor3)
d33=density(Z1)
hist(Z1, main="Dérivées premières",xlab = "Tmin ",freq = FALSE, ylim=c(0,0.6))
lines(d3,col="purple")
lines(d33,col="blue")
```

suite dérivées secondes

```
j=1 :114
c2=X[j+2,]-2*X[j+1,]+X[j,]
hist(c2, main="Dérivées secondes",xlab = "Tmoy ")
e2=Y[j+2,]-2*Y[j+1,]+Y[j,]
hist(e2, main="Dérivées secondes",xlab = "Tmax ")
b2=Z[j+2,]-2*Z[j+1,]+Z[j,]
hist(b2, main="Dérivées secondes",xlab = "Tmin ")
```

suite dérivées troisièmes

```
k=1 :113
c3=X[k+3,]-3*X[k+2,]+3*X[k+1,]-X[k,]
hist(c3, main="Dérivées troisièmes",xlab = "Tmoy")
e3=Y[k+3,]-3*Y[k+2,]+3*Y[k+1,]-Y[k,]
hist(e3,main="Dérivées troisièmes",xlab = "Tmax")
b3=Z[k+3,]-3*Z[k+2,]+3*Z[k+1,]-Z[k,]
hist(b3, main="Dérivées troisièmes",xlab = "Tmin")
```

pourcentages  $((ti+1)-ti)/ti$

```
par(mfrow=c(1,2))
X11=X1/X[i,]
hist(X11,freq=FALSE)
```

```
Y11=Y1/Y[i,]
hist(Y11,freq=FALSE)
```

```
Z11=Z1/Z[i,]
hist(Z11, freq=FALSE)
```

couples consécutifs des températures

```
X2=cbind(X[i,],X[i+1,]) couples consécutifs des Températures minimales
Y2=cbind(Y[i,],Y[i+1,]) couples consécutifs des Températures moyennes
Z2=cbind(Z[i,],Z[i+1,]) couples consécutifs des Températures maximales
```

```
plot(X2)
plot(Y2)
plot(Z2)
```

Transformée de Box-cox pour voir si c'est possible de transformer les données en gaussiennes

```
Tmin
h1<-function(lambda)
f=((X[i,]lambda - 1)/X[i,])
return(f)
```

```
Tmoy
h2 <- function(lambda)g = ((Y[i,]lambda - 1)/Y[i,])return(g)
Tmax
h3 <- function(lambda)h = ((Z[i,]lambda - 1)/Z[i,])return(h)
histogrammepourcoxpourTmoy
```

```
bc1 = h2(0.5)
bc2 = h2(-0.5)
bc3 = h2(-0.6)
bc4 = h2(-0.4)
bc5 = h2(-1)
bc6 = h2(1)
```

```
par(mfrow=c(1,2))
hist(bc1,xlab="lambda=0.5",main=" ")
hist(bc2,xlab="lambda=-0.5",main=" ")
hist(bc3,xlab="lambda=-0.6",main=" ")
hist(bc4,xlab="lambda=-0.4", main=" ")
hist(bc5,xlab="lambda=-1", main=" ")
hist(bc6,xlab="lambda=1", main=" ")
```

regression linéaire

```
par(mfrow=c(1,1))
```

```
t=1 :116
```

```
plot(t+1899,y, type="p", col = "black", main = "Temperature moyenne annuelle de 1900 à 2015 en France"
, ylim = c(min(y), max(y)), ylab = "Température moyenne annuelle (Y)", xlab="années")
```

```
tt=1 :81 plot(tt+1899,y[1 :81], type="p", col = "blue", main = "Temperature moyenne annuelle de 1900
à 1980 en France" , ylim = c(min(y), max(y)), ylab = "Température moyenne annuelle (Y)", xlab="années")
```

```
ttt=1 :35 plot(ttt+1899,y[82 :116], type="p", col = "red", main = "Temperature moyenne annuelle
```



```
de 1980 à 2015 en France" , ylim = c(min(y), max(y)), ylab = "Température moyenne annuelle (Y)",  
xlab="années")
```

```
regression sur les températures moyennes de 1900 à 2015
```

```
reg1=mf[,-c(2,3)]  
all1=lm(formula=Tmoy Date, data=reg1)  
summary(all1)
```

```
with( reg1, plot(Tmoy Date,type="p",col="red" ))  
abline(all1, col="blue",lwd=2)
```

```
hist(reg1[,2])
```

```
r1=residuals(all1) calcul des résidus de même variance
```

```
test si résidus gaussiens  
shapiro.test(r1)  
qqnorm(r1)  
qqline(r1)  
ks.test(r1,mean(r1),sd(r1))
```

```
hist(r1,main="Températures annuelles moyennes", xlab="Histogramme des résidus de 1900 à 2015")
```

```
de 1900 à 1980
```

```
reg2=mf[-c(81 :116),]  
all2=lm(formula=Tmoy Date, data=reg2)  
summary(all2)
```

```
with( reg2, plot(Tmoy Date,type="p",col="black" ))  
abline(all2, col="red",lwd=2)  
hist(reg2[,2])
```

```
r2=residuals(all2) calcul des résidus de même variance
```

```
test si résidus gaussiens  
shapiro.test(r2)  
qqnorm(r2)  
qqline(r2)  
ks.test(r2,mean(r2),sd(r2))
```

```
hist(r2,main="Températures annuelles moyennes", xlab="Histogramme des résidus de 1900 à 1980")
```

```
de 1981 à 2015
```

```
reg3=mf[-c(1 :80),]  
all3=lm(formula=Tmoy Date, data=reg3)  
summary(all3)
```

```
with( reg3, plot(Tmoy Date,type="p",col="black" ))  
abline(all3, col="purple",lwd=2)  
hist(reg3[,2])
```

```
r3=residuals(all3) calcul des résidus de même variance
```

```
test si résidus gaussiens  
shapiro.test(r3)  
qqnorm(r3)  
qqline(r3)  
ks.test(r3,mean(r3),sd(r3))
```

```
hist(r3,main="Températures annuelles moyennes", xlab="Histogramme des résidus de 1981 à 2015")
```

```

library(FactoMineR)

v=array(Y[,1],c(10,11)) matrice 11 variables et 10 ind correspondant aux Tmoy des 110 lières années
colMeans(v)

w=scale(v, center=TRUE, scale = TRUE)matrice centrée réduite de v
cov(w)matrice covariance

acp=PCA(w) acp
acpeigd'aprslargledeKaiserouducoudeonpeutretenirles4premieressvalpro

barplot(acpeig[,1],main = "valeurspropres", names.arg = paste("dim", 1 : nrow(acpeig)))

dimdesc(acp) description des axes factoriels par les variables

pairs(v[1 :10,1-12]) corrélation entre les variables
summary(acp)

par(las=2)
plot(v[1 :10,10], type = "b", axes = T, ylab = "variable i", xlab = " ", bty="o")
axis(2)
axis(1,1 :10,rownames(v)[1 :10])
par(las=0)
plot(acp, choix="var", axes = c(1,3))
plot(acp, choix="var", axes = c(1,2))
round(cor(v[,c(1,8,4,9)]),2) corrélation entre v1,v8,v4,v10 avec 2 chiffres après la virgule
acpvar
acpind

```