



Une analyse statistique de l'évolution des températures depuis un siècle à Rennes et Sydney

Marie-Madeleine Derriennic, Yves Derriennic

► To cite this version:

Marie-Madeleine Derriennic, Yves Derriennic. Une analyse statistique de l'évolution des températures depuis un siècle à Rennes et Sydney. 2020. hal-02888189

HAL Id: hal-02888189

<https://hal.science/hal-02888189>

Preprint submitted on 8 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Une analyse statistique de l'évolution des températures depuis un siècle à Rennes et Sydney

Marie-Madeleine Derriennic*

Dépt. de Math. INSA, Rennes.

Yves Derriennic†

Dépt. de Math. Fac. des Sciences,

UBO, Brest.

Abstract

Annual counts of "hot" or "cold" days since 1925 at Rennes (Fr.) and Sydney (Aus.) are presented. An analysis based on elementary statistical tests is performed in order to try to separate randomness from tendencies. Numbers of "hot" days and "cold" days for the same year appear to be uncorrelated. Random fluctuations around linear trends appear and are shown to be (plausibly) normally distributed. Fluctuations dominate linear trends except for Sydney "cold" days numbers. In both locations the decline of numbers of "cold" days is more striking than the rise of numbers of "hot" days, especially at Sydney; this suggests that the current warming occurs mainly by the rise of low temperatures.

Résumé Des dénombrements annuels de jours "chauds" ou "froids" depuis 1925 à Rennes (Fr.) et Sydney (Aus.) sont présentés. Une analyse fondée sur des tests statistiques élémentaires est réalisée en vue d'essayer de séparer le hasard des tendances. Les nombres de jours "chauds" et jours "froids" d'une même année sont non corrélés. Des fluctuations au hasard autour de tendances linéaires apparaissent et sont, de façon plausible, normalement distribuées. Ces fluctuations dominent les tendances linéaires sauf pour les nombres de jours "froids" de Sydney. Dans les deux lieux la baisse des nombres de jours "froids" est plus frappante que la hausse des nombres de jours "chauds", spécialement à Sydney; ceci suggère que le réchauffement en cours se produit principalement par l'élévation des basses températures.

I. Introduction.

Les stations météorologiques de Rennes (Fr.) et Sydney (Aus.) donnent des relevés de température détaillés depuis longtemps: Rennes depuis 1925 et Sydney depuis 1850. Le travail qui suit présente une analyse statistique de données extraites de ces relevés.

Aujourd'hui et depuis plusieurs années le grand public est sollicité de façon pressante par une multiplicité de nouvelles et rapports sur l'évolution climatique. Mais, loin des opinions toutes faites, que peut-on voir en approfondissant un peu la lecture de relevés de mesures météo effectuées depuis cent ans? Serait-il possible de saisir ce que représente quantitativement le réchauffement climatique de façon directe et intuitive? L'objet du présent travail est d'apporter des éléments de réponse à ces questions.

Tout d'abord sont présentés des tableaux rassemblant des dénombrements de jours "chauds" ou "froids" par année depuis 1925, en ces deux lieux, Rennes et Sydney, aux antipodes l'un de l'autre. Ces tableaux apparaissent aux parties III et VI après la présentation à la partie II

*madeleine.derriennic@gmail.com

†yves.derriennic@univ-brest.fr

⁰30 juin 2020

des totaux par décennie depuis 1930. Dans les autres parties IV-VIII est proposée une analyse statistique comparée de ces tableaux. Le but de l'étude n'est pas de désigner des causes ni de faire des prévisions, il est descriptif. Il s'agit de dégager des traits remarquables contenus dans les données mais n'apparaissant pas dans une lecture rapide. Une question guide l'analyse: quelle part de hasard peut on déceler dans ces séquences de valeurs couvrant presque un siècle? Ici le hasard n'est donc envisagé que du point de vue phénoménologique.

Les méthodes de statistique mathématique employées sont élémentaires: elles sont du niveau de l'enseignement donné aujourd'hui au Lycée ou en première année de Licence de Sciences naturelles ou de Sciences Humaines. Elles ne se réduisent pas à des calculs de moyennes qui trop souvent sont considérés comme les seuls principes de la statistique. La partie IV consacrée à l'analyse de la liste des nombres annuels de jours "chauds" de Rennes expose en détail la démarche et les raisonnements; les parties V, VII et VIII reprennent pour les autres suites, jours "froids" de Rennes, jours "chauds" ou "froids" de Sydney, la même approche, mais de façon abrégée, les détails étant renvoyés en Appendice. La partie X propose une synthèse de ces différentes analyses. La partie XI contient des remarques méthodologiques et des réflexions sur quelques ouvrages de climatologie.

En anticipant un peu sur les conclusions de la partie XII, on peut avancer que des tableaux similaires pour un échantillon de lieux plus nombreux et bien choisis pourraient être formés et analysés; par exemple dans [PoBrMa] p.112 on trouve déjà un diagramme figurant les nombres annuels de journées estivales à Paris et à Toulouse depuis 1950. Une image d'ensemble apparaîtrait et ses contrastes donneraient des informations importantes.

II. Conventions de terminologie et Tableau des nombres des jours "chauds" ou "froids" à Rennes ou à Sydney totalisés par décennie depuis 1930.

Nous appelons "chaud" un jour dont la température maximale a atteint ou dépassé 30° et "très chaud" un jour dont la température maximale a été plus de 35° . Les jours "très chauds" sont évidemment comptés parmi les jours "chauds". Pour les jours "froids" la convention est différente pour Rennes et pour Sydney. Pour Rennes nous appelons "froid" un jour dont la température minimale a atteint ou est descendue sous 0° et "très froid" un jour où cette température minimale est passée sous -5° ; les jours "très froids" font évidemment partie des jours "froids". A Sydney le climat est différent, il n'y a presque jamais de gel; aussi pour obtenir des nombres significatifs, un jour est appelé "froid" si sa température minimale a atteint ou est descendue sous $+10^{\circ}$ et "très froid" si cette température minimale est passée sous $+5^{\circ}$.

Dans les comptages annuels de nombres de jours "chauds" ou "froids", pour respecter la continuité des saisons, a été adoptée la règle de comptage par saison chaude ou saison froide. Il faut donc comprendre que le nombre de jours "froids" de l'année n à Rennes est, ici, le nombre de jours "froids" comptés entre le mois d'Octobre de l'année n et le mois d'Avril de l'année $n + 1$. A Sydney, dans l'hémisphère Sud, c'est l'inverse: le nombre de jours "chauds" de l'année n à Sydney est le nombre de jours "chauds" comptés entre le mois d'Octobre de l'année n et le mois d'Avril de l'année $n + 1$. Les relevés disponibles pour Rennes débutent en 1925 mais ceux des années 1940 (à partir de Juin) à 1944 manquent en raison de l'Occupation.

Pour amorcer cette étude nous présentons d'abord les totaux décennaux: Tableau 1. Pour la raison qui vient d'être dite, pour Rennes les nombres de la décennie 1940-49 manquent. Ce tableau montre déjà plusieurs traits remarquables.

Le contraste des climats de Rennes et Sydney est frappant. Sydney, situé dans l'hémisphère Sud, au bord de l'océan Pacifique, proche de la zone tropicale, est soumis à des températures très différentes de celles de Rennes. Par exemple il n'y a eu à Sydney que quatre jours pendant la dernière décennie dont la température minimale est passée sous $+5^{\circ}$!

**Tableau 1. Nombres de jours "chauds" ou "froids"
à Rennes et à Sydney totalisés par décennie, depuis 1930.**

	$t_{\min} \leq -5^\circ$	$t_{\min} \leq 0^\circ$	$t_{\max} \geq 30^\circ$	$t_{\max} \geq 35^\circ$
RENNES	"tr.froid"	"froid"	"chaud"	"tr.chaud"
2010-2019	27	276	80	8
2000-2009	27	338	93	9
1990-1999	23	302	103	8
1980-1989	39	369	52	0
1970-1979	20	370	50	7
1960-1969	65	415	28	1
1950-1959	45	458	40	2
1940-1949	données manquantes			
1930-1939	46	380	40	2

	$t_{\min} \leq +5^\circ$	$t_{\min} \leq +10^\circ$	$t_{\max} \geq 30^\circ$	$t_{\max} \geq 35^\circ$
SYDNEY	"tr.froid"	"froid"	"chaud"	"tr.chaud"
2010-2019	4	654	236	55
2000-2009	5	730	198	40
1990-1999	5	722	144	30
1980-1989	13	728	135	35
1970-1979	19	849	155	38
1960-1969	16	874	162	23
1950-1959	20	844	134	31
1940-1949	48	1029	160	39
1930-1939	60	1008	155	36

Le réchauffement entre les périodes de trente ans 1950-1979 et 1990-2019 apparaît évident pour les deux villes. A Rennes le nombre de jours "chauds" par an est passé "en gros" de 4 à 9; à Sydney ce nombre est passé de 15 à 19. Pour les jours "très froids" à Rennes ($t_{\min} \leq -5^\circ$) le nombre est passé "en gros" de 4 à 2 par an; à Sydney (jours "très froids": $t_{\min} \leq +5^\circ$) il est passé de 2 par an à 1 tous les deux ans! Se dessine aussi la phase de refroidissement des années 1950-70 qui a suivi le premier réchauffement de 1910-40 ([Ka]p.64).

Mais des différences attirent l'attention. A Rennes la décennie la plus chaude semble être 1990-1999; depuis apparaît une certaine stabilisation sinon décroissance des températures. A Sydney le réchauffement semble avoir commencé "par le bas", c'est à dire par la hausse des températures basses d'hiver, dès les années 1940; les hautes températures d'été n'ont monté que depuis 20 ans et c'est la dernière décennie, 2010-2019, qui est nettement la plus chaude.

Pour approfondir ces observations il faut considérer les tableaux des nombres annuels, ce que nous allons faire dans les parties suivantes.

III. Tableau des nombres annuels des jours "chauds" ou "froids" à Rennes depuis 1925.

Rappelons que les années 1940-44 manquent. Rappelons aussi que le nombre de jours "froids" de l'année n est le nombre de jours "froids" de la saison froide des années $n \sim (n + 1)$, donc dans la première ligne du tableau ci-après le nombre de jours "froids" indiqué pour l'année 1925 est le nombre de jours "froids" relevé durant la saison froide 1925-26 et dans la dernière ligne le nombre de jours "froids" de 2019 est le nombre de la saison froide 2019-20.

Tableau 2. Nombres de jours "chauds" ou "froids" par an à RENNES depuis 1925. (les cases vides sont des 0)

	$t_{\min} \leq -5^{\circ}$ "tr.froid"	$t_{\min} \leq 0^{\circ}$ "froid"	$t_{\max} \geq 30^{\circ}$ "chaud"	$t_{\max} \geq 35^{\circ}$ "tr.chaud"
1925	3	46		
26	2	32	6	
27	4	29		
28	12	67	2	
29	1	26	5	
1930	3	36	4	
31	4	46		
32	6	40	6	2
33	9	62	10	
34	2	31	7	
35		27	4	
36		29		
37	1	33	3	
38	6	24	4	
39	15	52	2	
40	absence de données			
41				
42				
43				
44				
45	11	49	3	
46	18	56	3	
47	4	38	18	1
48	9	55	3	
49	8	43	19	1
1950	2	44	2	
51		40		
52	2	67	6	2
53	8	41	4	
54		57		
55	19	71	8	
56	4	31	2	
57	3	39	7	
58	1	41		
59	6	27	11	
1960		23	2	
61	6	49	9	1
62	32	69	1	
63	6	54	1	
64	4	50	5	
65	6	32		
66	3	24		
67		38	1	
68	6	34	4	
69	2	42	5	
1970	10	39	2	
71		25	6	
72		50	1	

	$t_{\min} \leq -5^{\circ}$ "tr.froid"	$t_{\min} \leq 0^{\circ}$ "froid"	$t_{\max} \geq 30^{\circ}$ "chaud"	$t_{\max} \geq 35^{\circ}$ "tr.chaud"
1973		34	6	
74		17	1	
75	1	42	9	2
76		29	22	5
77	1	37	1	
78	6	62		
79	2	35	2	
1980	3	46	2	
81		25	3	
82		34	7	
83	2	46	10	
84	13	51	6	
85	10	52	2	
86	9	47	4	
87	2	19	4	
88		26	1	
89		23	13	
1990	8	36	21	5
91	4	42	9	
92		27	8	
93		27	2	
94		14	7	
95	2	42	23	1
96	8	36	8	
97	1	30	10	
98		21	5	2
99		27	10	
2000		19	5	
01	1	28	11	1
02	2	24	2	
03	1	32	23	6
04	2	35	11	
05	3	56	15	
06	1	14	17	2
07	4	37	1	
08	7	42	1	
09	6	51	7	
2010	5	52	6	
11	8	27	4	
12		34	4	
13		12	10	
14	2	33	1	
15		19	8	1
16	9	34	10	2
17	3	30	13	1
18		22	11	
2019		13	13	4

Les nombres de jours "froids" sont plus élevés que ceux de jours "chauds", cela est dû à la convention de départ dans la définition de "chaud" ou "froid" et ne donne donc pas d'information.

Dans la suite on se propose de faire apparaître quelques traits remarquables présents dans ces données mais cachés à la première lecture. En effet ce qui saute aux yeux c'est seulement la grande variabilité des nombres présentés avec une certaine tendance à la hausse des nombres de jours "chauds" et baisse des nombres de jours "froids".

Tout d'abord voici, Figure 1, la distribution croisée des nombres annuels de jours "chauds" ou "froids" à Rennes depuis 1925: pour chaque année entre 1925 et 2019 un point est placé avec pour abscisse son nombre de jours "chauds" et pour ordonnée son nombre de jours "froids"; les points rouges sont des points doubles. Les valeurs "jours chauds" sont comprises entre le minimum $\min_C = 0$ (cette valeur est prise 10 fois dont 9 fois dans la première moitié de la liste avant 1972) et le maximum $\max_C = 23$; leur médiane est 4,5. Pour les valeurs "jours froids" $\min_F = 12$, $\max_F = 71$ et la médiane est 35,5.

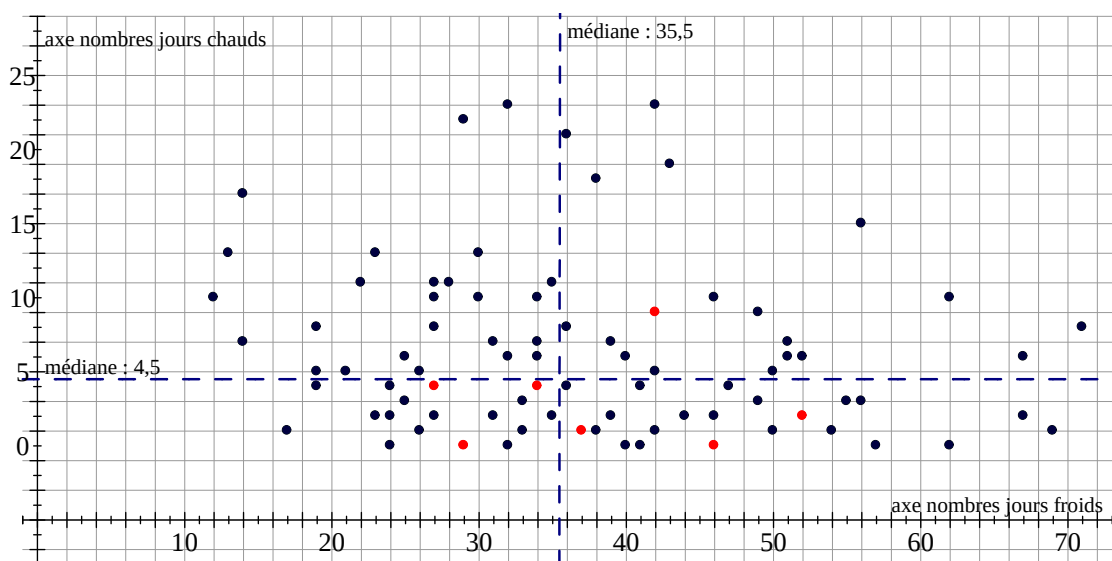


Figure 1. Distribution croisée "chauds"- "froids", Rennes, 1925-2019 . En pointillés : droite horizontale médiane des nombres de jours "chauds", droite verticale médiane des nombres de jours "froids". 7 points doubles en rouge; 10 points sur la droite d'ordonnée 0 : axe horizontal abaissé pour la lisibilité.

Cette figure montre déjà un trait remarquable: aucune liaison significative n'apparaît entre le nombre de jours "chauds" d'un été et le nombre de jours "froids" de l'hiver qui suit! La répartition des 90 années dans les quatre cases délimitées par les médianes est: $\frac{25}{20} \mid \frac{20}{25}$.

Cette répartition serait tout à fait plausible si les 90 points avaient été répartis au hasard dans les quatre cases, en respectant les liaisons fixées par la définition des médianes qui font qu'il n'y a qu'un seul degré de liberté. Dans l'Appendice 1 (n°1) cette figure subdivisée en 9 cases délimitées par les terciles est redonnée: ce tableau confirmera l'observation actuelle. On peut donc considérer qu'il n'y a pas de corrélation entre le nombre de jours "chauds" d'un été et le nombre de jours "froids" de l'hiver suivant (le coefficient de corrélation linéaire vaut 0,17). Les deux listes de valeurs vont faire l'objet de deux analyses séparées.

IV. Analyse statistique de la liste des nombres annuels de jours "chauds" à Rennes ($t_{\max} \geq 30^\circ$) depuis 1925 avec exposé détaillé des méthodes employées.

Avant d'aller plus loin revenons sur la question déjà abordée dans l'introduction: pourquoi entreprendre une analyse statistique du Tableau 1? La hausse des températures est évidente, pourquoi vouloir en dire plus? Mais la variabilité apparente des listes présentées est grande: le hasard pourrait-il être la "cause" principale de celle-ci? Précisons un peu le sens de cette question.

Les fluctuations de la liste des nombres annuels de jours "chauds" à Rennes depuis 1925 pourraient-elles provenir tout simplement d'une succession de chances et malchances se répétant de façon plus ou moins constante, ou bien pourrait-on discerner un changement au cours du temps de la loi statistique qui les gouverne? Un changement de régime de la loi statistique est-il requis pour qu'il y ait une probabilité non négligeable que des variations comparables à celles observées se produisent? De façon un peu plus savante, serait-il plausible qu'une telle liste provienne d'un "processus stochastique stationnaire"? Tenter de répondre à ces questions est l'objet de cette partie.

Dans un premier temps ne sera étudiée que la liste de nombres de jours "chauds" à Rennes; les nombres de jours "très chauds" seront considérés ensuite. Cette partie sera la plus longue car elle indique en détail la démarche suivie. Celle-ci sera reprise, sans les explications, aussi bien pour les jours "froids" de Rennes que pour les données de Sydney dans les parties suivantes.

Le nombre de jours "chauds" de l'année n est noté C_n (pour les jours "froids" dans les parties suivantes on notera F_n). Chaque symbole qui porte l'indice C est relatif à la liste des nombres de jours "chauds" (F pour "froids"). La lettre n est utilisée pour représenter le numéro de l'année; comme les années 1940-1944 manquent il n'y a que 90 années à considérer et donc l'indice n circule entre 1 et 90. Dans cette numérotation l'interruption 1940-1944 sera simplement ignorée: le numéro $n = 1$ représente l'année 1925 et $n = 90$ l'année 2019. L'interruption 1940-1944 ne sera prise en compte que pour la détermination de la "droite des moindres carrés".

L'analyse va se décomposer en plusieurs parties.

A) Pour la suite $(C_n)_{1 \leq n \leq 90}$ on calcule les caractéristiques empiriques usuelles: le minimum $\min_C$, le maximum $\max_C$, la moyenne M_C , la médiane med_C , l'écart-type σ_C , les quartiles inférieur Q_C^i et supérieur Q_C^s . On trouve:

$$\min_C = 0 ; \max_C = 23 ; M_C = 6,1 ; \sigma_C = 5,6 ; \text{med}_C = 4,5 ; Q_C^i = 2 ; Q_C^s = 9.$$

La valeur 0 est prise 10 fois, mais une seule fois depuis 1970. La distribution des C_n est déséquilibrée vers les grandes valeurs, non seulement à cause du grand nombre de zéros, dû à la définition conventionnelle de "chaud", mais aussi parce que, par exemple, 6 valeurs dépassent $M_C + 2\sigma_C = 17,3$.

Pour décrire la tendance moyenne de la liste des valeurs $(C_n)_{1 \leq n \leq 90}$ on cherche sa droite d'approximation au sens des "moindres carrés". Pour cela, dans un repère orthonormé xOy la liste $(C_n)_{1 \leq n \leq 90}$ est représentée par le nuage des 90 points de coordonnées $(n, C_n)_{1 \leq n \leq 15}$ et $(n+5, C_n)_{16 \leq n \leq 90}$; ici les abscisses représentent les années écoulées (pour 1925 $n = 1$, pour 2019 $n = 90$) et les ordonnées les nombres C_n observés; le décalage de 5 est causé par l'absence de données pour les 5 années 1940-44. La "droite des moindres carrés" de ce nuage est la droite d'équation $y = ax + b$ dont les paramètres a et b sont choisis de façon à minimiser la quantité $\sum_{n=1}^{15} (C_n - an - b)^2 + \sum_{n=16}^{90} (C_n - a(n+5) - b)^2$. On sait que la droite "horizontale" $y = m$ qui minimise $\sum_{n=1}^{90} (C_n - m)^2$ a pour ordonnée $m = M_C$, la moyenne de la liste $(C_n)_{1 \leq n \leq 90}$. Donc la droite "des moindres carrés" précise la notion de moyenne: la droite horizontale qui approche le mieux le nuage de points a pour ordonnée la moyenne, mais parmi toutes les droites celle qui approche le mieux le nuage, au sens des moindres carrés, est celle appelée "droite des moindres carrés": $y = a_C x + b_C$ dont la pente est notée a_C et l'ordonnée à l'origine b_C .

En arrondissant un peu les valeurs, on trouve:

$$y = 0,07x + 2,7 : \text{pente } a_C = 0,07.$$

La pente donne l'information principale: en moyenne le nombre annuel de jours "chauds" a augmenté en 95 ans de $0,07 \times 95 = 6,6$. L'ordonnée à l'origine $b_C = 2,7$ donne une information complémentaire; elle dépend du choix de la position $x = 0$, donc ici du choix $n = 1$ pour 1925. L'équation précédente indique que depuis 1925, "en gros", le nombre annuel de jours "chauds" de Rennes est passé de 2 ou 3 à 8 ou 9. Cela donne la représentation graphique suivante:

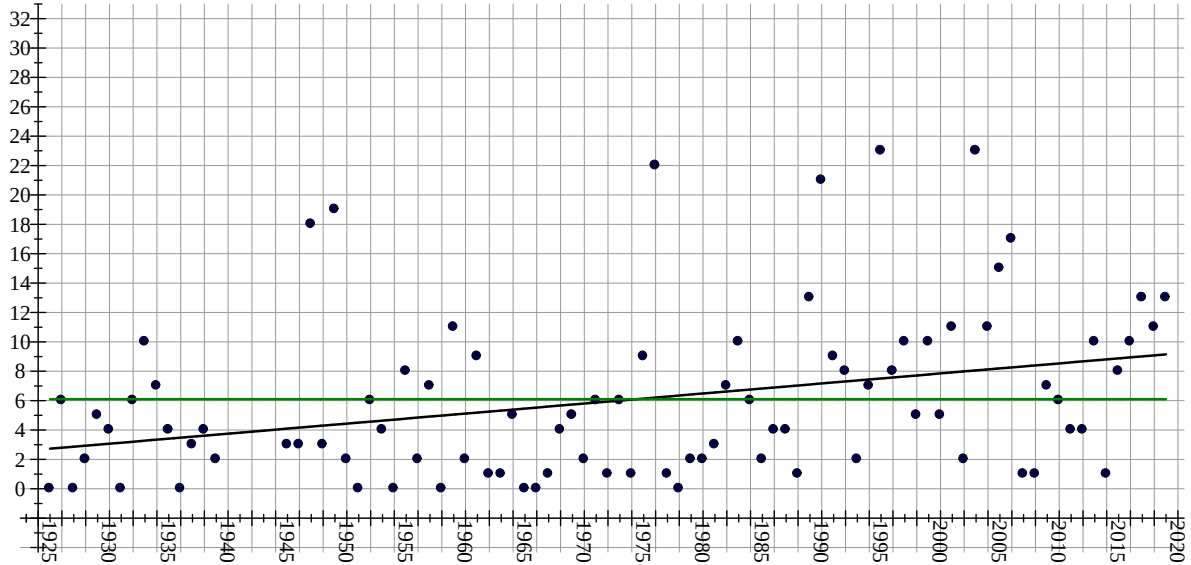


Figure 2. Nombres de jours "chauds", $t_{\max} \geq 30^\circ$, à Rennes en fonction de l'année depuis 1925, avec droite des moindres carrés en noir de pente $a_C = 0,07$. En vert la droite horizontale d'ordonnée la moyenne $M_C = 6,1$. Axe horizontal abaissé pour la lisibilité.

B) Nous abordons maintenant la question principale posée au début de cette partie: le hasard pourrait-il provoquer les variations observées de la liste $(C_n)_{1 \leq n \leq 90}$ des nombres annuels de jours "chauds" à Rennes depuis 1925? Pour n'employer ici que des méthodes élémentaires, cette question ne sera envisagée que sous une forme simplifiée, à savoir: pourrait-on produire avec une probabilité significative une liste telle que la liste $(C_n)_{1 \leq n \leq 90}$ en effectuant 90 tirages au sort successifs, stochastiquement indépendants, selon une distribution invariante au court du temps?

La statistique mathématique élémentaire propose plusieurs méthodes pour répondre à cette question c'est à dire pour tester l'hypothèse: "la suite $(C_n)_{1 \leq n \leq 90}$ a été obtenue par une suite de tirages indépendants suivant une même distribution; autrement dit, c'est une suite indépendante et stationnaire (i.e. homogène)". Cette hypothèse sera notée $\mathcal{H}(C)$:

$\mathcal{H}(C)$: "la suite $(C_n)_{1 \leq n \leq 90}$ forme une suite indépendante et stationnaire".

La distribution invariante évoquée dans l'hypothèse $\mathcal{H}(C)$ n'est en rien précisée, les méthodes courantes supposant la "normalité" des distributions considérées sont donc exclues. Les trois tests statistiques que nous allons utiliser, ne supposent pas de condition sur cette distribution invariante.

Dans un premier temps $\mathcal{H}(C)$ sera soumise au "test de la médiane", Test 1, aussi appelé "test des séries" et à un test "d'équilibre", Test 2, variante du test de la médiane. Le premier pas consiste à remplacer la liste $(C_n)_{1 \leq n \leq 90}$ par la suite des lettres P ou F, Pile ou Face, en écrivant P si $C_n > \text{med}_C$ ou F si $C_n < \text{med}_C$. La suite obtenue est donc de longueur 90.

La voici, le respect de l'ordre de la suite est essentiel:

F, P, FF, P, FF, PPP, FFFFFFFF, P, F, P, FF, P, FF, P, F, P, F, P, F, P, FF, P, FFFF, P, F, P, F, P, F
PP, FFFFF, PPP, FFFF, PPPP, F, PFFFFFFF, F, PPPP, FF, PP, FF, P, F, PFFFF

Cette suite comporte 45 P et 45 F, l'égalité de ces deux nombres provient de la définition même de la médiane med_C . Sous l'hypothèse $\mathcal{H}(C)$, c'est à dire en supposant $\mathcal{H}(C)$ vraie, la suite écrite ne devrait pas être distinguable d'une suite de même longueur obtenue par le jeu de Pile ou Face (c'est à dire par 90 jets successifs d'une pièce de monnaie ordinaire donnant la probabilité 1/2 à chaque face) vérifiant la condition d'égalité des nombres de Pile et Face. En effet la distribution inconnue supposée invariante n'intervient que par sa médiane qui partage en deux moitiés la suite des valeurs; ici on assimile la médiane théorique à la médiane empirique. Or le Calcul des Probabilités établit la quantification des probabilités *a priori* de nombreuses propriétés d'une suite de Pile ou Face. Deux d'entre elles vont être utiles.

Test 1 (de la médiane). Pour "tester" l'hypothèse $\mathcal{H}(C)$ nous allons considérer le nombre de "séries" composant la suite de P ou F écrite ci-dessus et la longueur de la plus longue d'entre elles; on appelle série une séquence d'une même lettre (de longueur ≥ 1). Ici la suite entière est formée de 44 séries, indiquées par les virgules, et la plus longue série comporte 8 lettres, série de P au milieu de la seconde ligne. Or il est établi que:

"Pour une suite de longueur N obtenue par le jeu de Pile ou Face, sous la condition d'égalité des nombres de Pile et Face, il y a une probabilité a priori d'environ 0,95 que la longueur de la plus longue série $\tau(N)$ et le nombre de séries $\nu(N)$ vérifient les deux inégalités:

$$\tau(N) < 3,3(1 + \log_{10} N) \text{ et } \nu(N) > \frac{N}{2} - \sqrt{N} \quad \textbf{Test 1}$$

(à condition que N soit assez grand, $N \geq 50$)" ([Aïv] chap.3, [Sie] chap. 4).

Ici, pour la suite de P ou F écrite ci-dessus $N = 90$ et en notant ν_C le nombre de séries, τ_C la longueur de la plus longue on trouve: $\nu_C(90) = 44$, $\tau_C(90) = 8$. Le premier seuil vaut $3,3(1 + \log_{10} 90) = 9,7 > 8 = \tau_C(90)$ et la seconde inégalité s'écrit $\nu_C(90) = 44 > \frac{90}{2} - \sqrt{90} = 35,5$. Donc pour ce Test 1, au niveau de confiance 0,95, l'hypothèse $\mathcal{H}(C)$ n'est pas rejetée, elle reste acceptable.

L'hypothèse $\mathcal{H}(C)$ porte à la fois sur l'indépendance et la stationnarité. Nous testons les deux propriétés ensemble. Cela est critiquable mais permet de rester dans des méthodes élémentaires. Ce test de la médiane privilégie l'indépendance sur la stationnarité. Puisque la stationnarité est la propriété critique la plus importante si l'on pense au problème du réchauffement, on doit aussi tester l'hypothèse $\mathcal{H}(C)$ par une méthode privilégiant la stationnarité.

Test 2 (d'équilibre). Pour la liste des P ou F écrite ci-dessus on calcule ε l'écart entre le nombre de P de la première moitié et celui de la seconde moitié. On trouve $\varepsilon = 29 - 16 = 13$ (il y a plus de P dans la seconde moitié). Or il est établi que:

"Pour une suite de longueur N obtenue par le jeu de Pile ou Face, sous la condition d'égalité des nombres de Pile et Face, il y a une probabilité a priori d'environ 0,95 que l'écart $\varepsilon(N)$ entre le nombre de Pile de la première moitié et celui de la seconde moitié vérifie l'inégalité:

$$\varepsilon(N) < \sqrt{N} \quad \textbf{Test 2}$$

(à condition que N soit assez grand)" ([F]).

Ici la valeur de cet écart noté $\varepsilon_C(90)$ est $\varepsilon_C(90) = 13 > \sqrt{90} = 9,5$. Donc par ce Test 2, au niveau de confiance 0,95 l'hypothèse $\mathcal{H}(C)$ est rejetée de façon nette: elle devient inacceptable.

Dans un second temps nous allons utiliser un autre test classique: le test des signes, qui convient aussi pour comparer les deux moitiés de la liste $(C_n)_{1 \leq n \leq 90}$. Il s'agit à nouveau d'éprouver en priorité la stationnarité de cette suite, mais toujours en supposant aussi l'indépendance.

Test 3 (des signes). Une suite de signes + ou - est formée en posant + si $C_n > C_{n+45}$ ou - si $C_n < C_{n+45}$ et en omettant les cas $C_n = C_{n+45}$; la longueur de la suite de signes, notée S , est le nombre des différences $C_n - C_{n+45}$ non nulles. On obtient ici la suite de signes suivante:

- - - + + + - - + + - - + - - - + + + - - - - - + - + - - - - - - - - + - - - - -

Elle comporte $S = 41$ termes. L'hypothèse $\mathcal{H}(C)$ sera testée d'après le nombre ρ de signes + qui ici vaut 12, car il est établi que:

"Pour une suite de + et - de longueur S obtenue complètement au hasard, par le jeu de Pile ou Face, il y a la probabilité a priori 0,95 que le nombre ρ de signes + vérifie l'inégalité:

$$\left| \rho - \frac{S}{2} \right| < \sqrt{S} \quad \text{Test 3}$$

(pour S assez grand)" ([Bor] ch. 4, [F]).

Sous l'hypothèse $\mathcal{H}(C)$, la suite de + ou - écrite ci-dessus ne devrait pas être distinguable d'une suite au hasard. Or en notant $\rho_C(41)$ le nombre de signes + observés dans cette suite, on a $\rho_C(41) = 12$, d'où $\left| \rho_C(41) - \frac{41}{2} \right| = |12 - 20,5| = 8,5 > \sqrt{41} = 6,4$ ce qui signifie donc que, au niveau de confiance 0,95, l'hypothèse $\mathcal{H}(C)$ doit être rejetée.

Il résulte de cette courte étude que la propriété d'indépendance stochastique des termes de la suite $(C_n)_{1 \leq n \leq 90}$ manifestée par leur grande variabilité pourrait éventuellement être acceptée mais qu'il ne peut pas être considéré comme plausible que cette suite soit indépendante et stationnaire, c'est à dire qu'elle soit le résultat d'un hasard maintenu dans un régime invariant depuis 1925.

C) Pour essayer de dégager un "modèle" plausible d'évolution pour la suite $(C_n)_{1 \leq n \leq 90}$ des nombres annuels de jours "chauds" à Rennes depuis 1925 l'analyse statistique du paragraphe **B** va être reprise pour la suite des valeurs $(d_n)_{1 \leq n \leq 90}$ obtenues en retranchant à C_n la valeur de la droite des "moindres carrés" pour la même année n . Ces différences d_n sont définies par les équations:

$$\text{pour } n \leq 15, d_n = C_n - a_C n - b_C \text{ et pour } 15 < n \leq 90, d_n = C_n - a_C(n + 5) - b_C,$$

avec $a_C = 0,07$ et $b_C = 2,7$ (le décalage de 5 étant dû à l'intervalle 1940-44). Cette suite se calcule immédiatement à partir de la suite $(C_n)_{1 \leq n \leq 90}$. Ces différences d_n représentent les fluctuations qui restent après avoir privé de sa partie linéaire l'évolution de la suite initiale $(C_n)_{1 \leq n \leq 90}$ des nombres de jours "chauds". La suite $(d_n)_{1 \leq n \leq 90}$ sera donc appelée aussi suite des fluctuations. Son analyse a pour but de dégager la part qui pourrait revenir au hasard. Sa moyenne est nulle, cela résulte de la définition de la droite des "moindres carrés". Avec les mêmes notations que précédemment dans lesquelles l'indice C est remplacé par d , les caractéristiques de cette liste sont:

$$\min_d = -7,8 ; \max_d = 15,8 ; M_d = 0 ; \sigma_d = 5,3 ; \text{med}_d = -0,8 ; Q_d^i = -3,1 ; Q_d^s = 2,2.$$

Les trois tests statistiques de la partie précédente **B** vont être utilisés pour éprouver la nouvelle hypothèse $\mathcal{H}(d)$ suivante: "la suite des "fluctuations" $(d_n)_{1 \leq n \leq 90}$ provient d'une suite de tirages indépendants suivant une distribution invariante". Autrement dit:

$$\mathcal{H}(d) : \text{"la suite } (d_n)_{1 \leq n \leq 90} \text{ forme une suite indépendante et stationnaire"}.$$

La distribution invariante évoquée dans cette hypothèse $\mathcal{H}(d)$ est, comme précédemment, inconnue.

Pour le test de la médiane on forme la liste des P ou F suivant la position de la valeur par rapport à la médiane comme on l'a fait pour la suite (C_n) ; on obtient:

F, P, FF, PP, F, PPPP, F, PP, FFF, P, F, P, FF, PP, F, P, F, P, F, P, F, P, FF, P, FFFFFFF, P, F, P, F, PP, FFFFF, PPP, FFFF, PPPP, F, PPPP, F, P, F, P, F, PPPP, FFFFFFF, P, F, F, PPPP

La longueur de cette suite est $N = 90$. Le nombre de séries, indiquées par les virgules, est $\nu_d(90) = 47$; la plus longue a pour longueur $\tau_d(90) = 6$ (deux séries de F, l'une à la fin de la première moitié et l'autre à la fin de la seconde). Pour le niveau de confiance 0,95 on trouve:

Test 1: $3,3(1 + \log_{10} 90) = 9,7 > 6 = \tau_d(90)$ et $\nu_d(90) = 47 > \frac{90}{2} - \sqrt{90} = 35,5$;

d'où acceptation de $\mathcal{H}(d)$.

Pour le Test 2, la différence absolue entre les nombres de P de la première moitié et de la seconde moitié de la suite écrite ci-dessus donne l'écart $\varepsilon_d(90) = |20 - 24| = 4$. On trouve alors:

Test 2: $\varepsilon_d(90) = 4 < \sqrt{90} = 9,5$; d'où acceptation de $\mathcal{H}(d)$.

Pour le test des signes, la liste de signes obtenue en comparant d_n et d_{n+45} est la suivante:

- - + + + + + + + + - + + - - - + + + - - - + - + - + - + - - + + + - - + - +

Sa longueur est $S = 45$. Le nombre de signes + est $\rho_d(45) = 23$. On obtient alors:

Test 3: $\left| \rho_d - \frac{S}{2} \right| = 23 - \frac{45}{2} = 0,5 < \sqrt{S} = \sqrt{45} = 6,7$; d'où acceptation de $\mathcal{H}(d)$.

L'hypothèse $\mathcal{H}(d)$ n'est pas rejetée, loin s'en faut: les acceptations sont très larges. Des points de vue de l'indépendance et de la stationnarité elle apparaît acceptable.

Ceci amène à considérer que la suite $(d_n)_{1 \leq n \leq 90}$ pourrait être prise, au moins dans une première approche, pour une suite de valeurs provenant du hasard, c'est à dire pour une suite de valeurs observées prises par des variables aléatoires indépendantes et de même distribution. Cette distribution est difficile à identifier précisément; elle est centrée (i.e. de moyenne nulle) mais dissymétrique, assez étalée surtout du côté positif. La dissymétrie de cette loi provient pour une part de la fréquence de la valeur minimale 0 dans la suite $(C_n)_{1 \leq n \leq 90}$.

En résumé cela rend plausible de considérer la suite $(C_n)_{1 \leq n \leq 90}$ des nombres annuels de jours "chauds" à Rennes depuis 1925 comme obéissant au modèle

$$C_n = 0,07n + 2,7 + d_n ; n = 1 \text{ pour } 1925$$

où la suite des fluctuations d_n est une suite de valeurs aléatoires stochastiquement indépendantes suivant une même distribution centrée. C'est un modèle d'évolution stochastique: une fonction linéaire, indiquant une tendance, plus une fluctuation aléatoire, indépendante et stationnaire. La partie X abordera la question de la "normalité", ou plutôt de la "normalité tronquée", de la distribution des fluctuations $(d_n)_{1 \leq n \leq 90}$.

D) Il est tentant de donner une valeur prédictive à l'équation que nous venons d'écrire, et d'affirmer par exemple qu'il faut s'attendre à voir le nombre de jours "chauds" par an à Rennes vers 2050 être encore "en moyenne" supérieur de deux unités, car $0,07 \times 30 = 2,1$. Ce serait une erreur de principe: cette étude n'a qu'une valeur descriptive, comme la plupart des études statistiques portant sur des données dont le relevé est unique, non répétable. Le "modèle" formulé ci-dessus est conçu pour permettre une compréhension synthétique de l'évolution passée depuis presque cent ans; il peut suggérer des questions ou des hypothèses. Il ne prédit pas le futur: les coefficients de la partie linéaire, a_C et b_C , sont extraits des observations passées. D'ailleurs en suivant la même démarche à partir de la liste (C_n) considérée seulement sur les 50 dernières années, c'est à dire de 1970 à 2019 on peut arriver à un modèle sensiblement différent qui donnerait une autre "prédiction".

Explicitons cela un peu en détail.

Reprenons la liste des nombres annuels de jours "chauds" à Rennes seulement à partir de l'année 1970. C'est donc une liste de longueur 50 qui sera notée $(C'_m)_{1 \leq m \leq 50}$. Elle est incluse dans la liste initiale: $C'_m = C_{m+40}$; le nouvel indice m numérote les années de 1970 à 2019.

Pour cette nouvelle liste la droite d'approximation au sens des "moindres carrés" d'équation a pour équation: $y = 0,10x + 5,0$. C'est à dire: pente $a'_C = 0,10$ et ordonnée à l'origine $b'_C = 5,0$ mais ici l'origine de l'axe des abscisses correspond au choix $x = 1$ pour l'année 1970 (Figure 3). En utilisant cette nouvelle équation pour prédire l'avenir on annoncerait donc pour 2050 une augmentation moyenne de trois unités du nombre de jours chauds par an au lieu des deux de la "prévision" précédente. Cela montre le caractère précaire de ce genre de "prévision". En fait la différence des pentes des deux "droites de moindres carrés" qui sont d'une part $a_C = 0,07$ pour la période 1925-2019 et d'autre part $a'_C = 0,10$ pour la période 1970-2019, met autant en évidence la période de refroidissement qu'a connu la Terre de 1940 à 1970 ([Ka] p.64) qu'une accélération du réchauffement. Si la droite d'approximation trouvée pour la période 1970-2019 était prolongée vers le passé, on trouverait en 1925 une différence de presque 2 avec la droite de la période 1925-2019: c'est une autre manière de montrer que l'extrapolation n'est pas nécessairement valide.

Voici la représentation graphique:

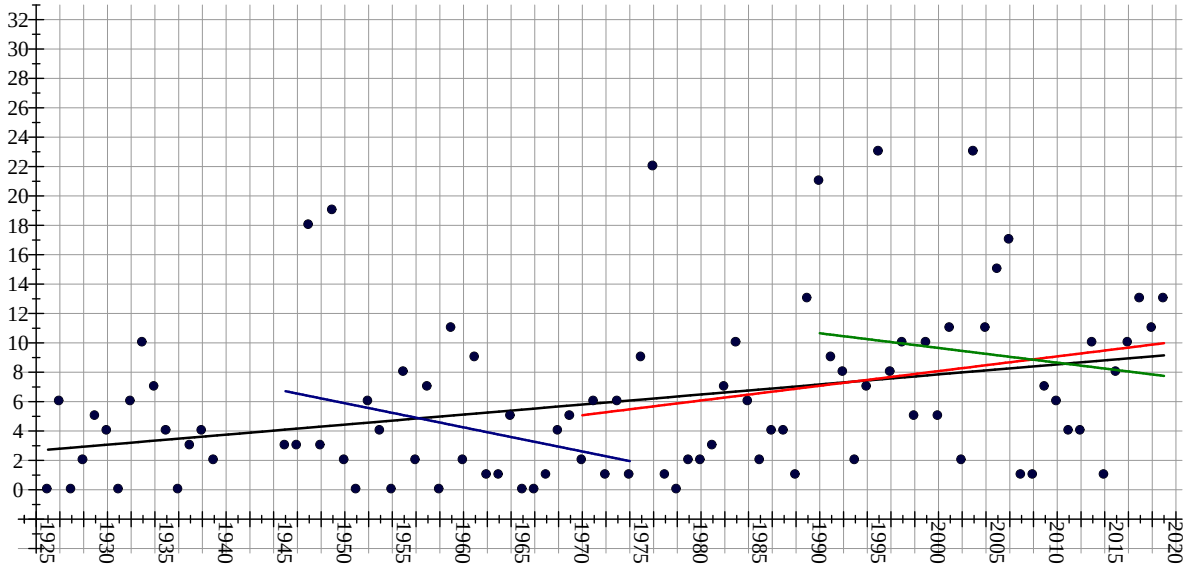


Figure 3. Nombres de jours "chauds", $t_{\max} \geq 30^\circ$, à Rennes, en fonction de l'année, avec droites des moindres carrés : noire 1925-2019 ; rouge 1970-2019 ; bleue 1945-1974 ; verte 1990-2019. Axe horizontal abaissé pour la lisibilité.

A la Figure 3 sont tracées aussi les droites des "moindres carrés" pour les deux périodes de 30 ans: 1945-1974 et 1990-2019. La première en bleu a la pente $-0,16$, et la seconde, en vert, la pente $-0,10$. Ces deux nouvelles droites sont décroissantes et cela fait douter de la validité d'une analyse basée sur une durée de 30 ans seulement.

V. Analyse statistique résumée de la liste des nombres annuels de jours "froids" à Rennes ($t_{\min} \leq 0^\circ$) depuis 1925.

Comme indiqué déjà, un jour "froid" à Rennes est par convention un jour dont la température minimale a atteint ou est descendue sous 0° . L'analyse statistique de la liste des nombres de jours "froids" $(F_n)_{1 \leq n \leq 90}$ suivra la même démarche que celle de la liste des nombres de jours "chauds", simplement le symbole C , pour "chaud", est remplacé par F , pour "froid". Le symbole d pour les fluctuations sera remplacé par g . La liste $(F_n)_{1 \leq n \leq 90}$ a été donnée au Tableau 2.

Pour alléger l'exposé un résumé seulement des résultats de l'étude sera donné ici. Dans l'Appendice 2 tous les détails des calculs et raisonnements sont indiqués.

La liste $(F_n)_{1 \leq n \leq 90}$ a les caractéristiques:

$$\min_F = 12 ; \max_F = 71 ; M_F = 37,2 ; \sigma_F = 13,4 ; \text{med}_F = 35,5 ; Q_F^i = 27 ; Q_F^s = 46.$$

Ses valeurs sont très supérieures à celles de la suite $(C_n)_{1 \leq n \leq 90}$; cela ne comporte aucune information car c'est le résultat de notre convention définissant "chaud" ou "froid". La droite d'approximation au sens des "moindres carrés" a pour équation:

$$y = -0,17x + 45,7 : \text{pente } a_F = -0,17.$$

La pente est négative, le nombre de jours froids a tendance à diminuer. La représentation graphique est à la Figure 4.

Pour tester l'hypothèse:

$$\mathcal{H}(F) : \text{"la suite } (F_n)_{1 \leq n \leq 90} \text{ forme une suite indépendante et stationnaire"}$$

les trois tests de la partie IV sont à nouveau employés. L'hypothèse $\mathcal{H}(F)$ est rejetée de façon très nette: toujours au même niveau de confiance 0,95, chacun des trois tests conclut au rejet.

On forme donc la liste $(g_n)_{1 \leq n \leq 90}$ des fluctuations: ce sont les différences notées g_n entre F_n et la valeur de la droite d'approximation, soit $g_n = F_n + 0,17n - 45,7$ pour $n \leq 15$ et $g_n = F_n + 0,17(n + 5) - 45,7$ pour $15 < n \leq 90$. Les caractéristiques de cette liste sont:

$$\min_g = -20,1 ; \max_g = 30,6 ; M_g = 0 ; \sigma_g = 12,6 ; \text{med}_g = -0,3 ; Q_g^i = -10,4 ; Q_g^s = 8,4.$$

De nouveau les trois tests de la partie IV sont utilisés pour éprouver l'hypothèse:

$$\mathcal{H}(g) : \text{"la suite } (g_n)_{1 \leq n \leq 90} \text{ forme une suite indépendante et stationnaire"}.$$

Cette hypothèse est acceptée: chacun des trois tests conclut à l'acceptation. On en déduit qu'un modèle plausible représentant l'évolution du nombre annuel de jours "froids" à Rennes depuis 1925 est donnée par l'équation:

$$F_n = -0,17n + 45,7 + g ; n = 1 \text{ pour } 1925,$$

où $(g_n)_{1 \leq n \leq 90}$ forme une suite de variables aléatoires indépendantes et de même distribution centrée (à la partie X la "normalité" éventuelle de cette distribution sera recherchée).

Comme dans l'étude des nombres de jours "chauds" on peut aussi reprendre la liste des nombres de jours "froids" tronquée de la période 1925-1969 et considérer la liste $(F'_m)_{1 \leq m \leq 50}$ des nombres de jours "froids" par an à Rennes depuis 1970: $F'_m = F_{m+40}$.

Sa droite d'approximation a pour équation: $y = -0,22x + 38,7$; pente $a'_F = -0,22$ avec $x = 1$ pour 1970.

Voici la représentation graphique:

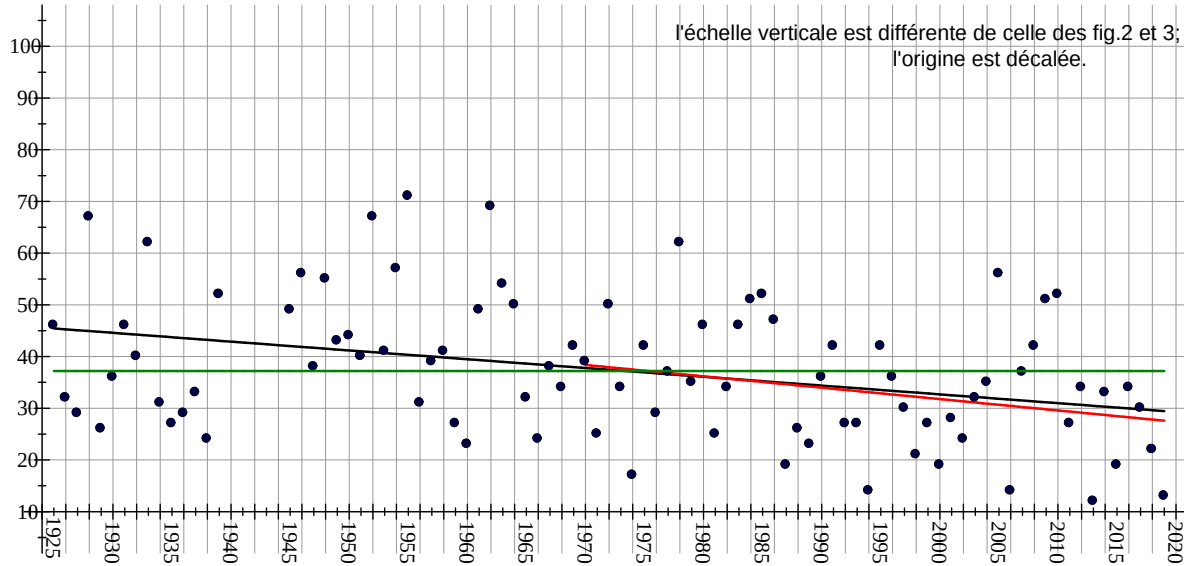


Figure 4. Nombres de jours "froids" à Rennes, $t_{\min} \leq 0^\circ$, en fonction de l'année, avec droites des moindres carrés : noire 1925-2019 ; rouge 1970-2019 . En vert la droite horizontale d'ordonnée la moyenne $M_F = 37,2$.

Bien que la Figure 4 montre une forte ressemblance entre les deux évolutions, depuis 1925 ou depuis 1970, on remarque qu'en extrapolant les équations jusqu'en 2050 une différence de 3 jours "froids" par an apparaît entre les deux. La disparition des jours "froids" tendrait à s'accélérer. Mais, comme cela a déjà été dit, cette étude ne donne pas de validité à cette sorte d'extrapolation. Si l'on prolongeait vers le passé la droite d'approximation trouvée pour la période 1970-2019, en 1925 apparaîtrait aussi une différence de 2 ou 3 avec la droite trouvée pour la période 1925-2019.

VI. Tableau des nombres annuels des jours "chauds" ou "froids" à Sydney depuis 1925.

Pour Sydney la convention définissant un jour "froid", $t_{\min} \leq +10^\circ$, et un jour "très froid", $t_{\min} \leq +5^\circ$, n'est plus celle de Rennes. Le nombre de jours "chauds" de l'année n est le nombre de jours "chauds" comptés entre le mois d'Octobre de l'année n et le mois d'Avril de l'année $n + 1$; donc dans la première ligne du Tableau 3 figurent les nombres de jours "chauds" ou "très chauds" relevés durant la saison chaude 1925-1926 et dans la dernière ligne les nombres de la saison chaude 2019-2020.

Tableau 3. Nombres de jours "chauds" ou "froids" par an à SYDNEY depuis 1925.

(les cases vides sont des 0)

| | $t_{\min} \leq +5^{\circ}$
"tr.froid" | $t_{\min} \leq +10^{\circ}$
"froid" | $t_{\max} \geq 30^{\circ}$
"chaud" | $t_{\max} \geq 35^{\circ}$
"tr.chaud" |
|------|--|--|---------------------------------------|--|
| 1925 | 1 | 109 | 19 | 9 |
| 26 | | 91 | 22 | 7 |
| 27 | 8 | 125 | 7 | 3 |
| 28 | 1 | 81 | 17 | 7 |
| 29 | 12 | 107 | 10 | 5 |
| 1930 | | 83 | 17 | 1 |
| 31 | 9 | 102 | 12 | 7 |
| 32 | 12 | 98 | 19 | 2 |
| 33 | 10 | 99 | 11 | 2 |
| 34 | 3 | 103 | 13 | 1 |
| 35 | 9 | 114 | 10 | 2 |
| 36 | 2 | 115 | 20 | 6 |
| 37 | 1 | 94 | 13 | 1 |
| 38 | 7 | 100 | 19 | 7 |
| 39 | 7 | 100 | 21 | 7 |
| 1940 | 1 | 101 | 18 | 3 |
| 41 | 6 | 108 | 25 | 6 |
| 42 | 2 | 94 | 20 | 4 |
| 43 | 10 | 116 | 11 | 2 |
| 44 | 6 | 113 | 21 | 6 |
| 45 | 4 | 91 | 17 | 8 |
| 46 | | 117 | 16 | 5 |
| 47 | 7 | 85 | 6 | |
| 48 | 7 | 111 | 17 | 2 |
| 49 | 5 | 93 | 9 | 3 |
| 1950 | 1 | 57 | 7 | 4 |
| 51 | | 84 | 19 | 4 |
| 52 | 2 | 78 | 6 | |
| 53 | 4 | 102 | 19 | 5 |
| 54 | 1 | 95 | 10 | 1 |
| 55 | 2 | 92 | 8 | |
| 56 | 3 | 97 | 9 | 3 |
| 57 | 2 | 87 | 26 | 8 |
| 58 | 4 | 70 | 11 | 1 |
| 59 | 1 | 82 | 19 | 5 |
| 1960 | 1 | 84 | 14 | 1 |
| 61 | 1 | 95 | 11 | 1 |
| 62 | 3 | 86 | 10 | 1 |
| 63 | 1 | 82 | 18 | 2 |
| 64 | 2 | 85 | 18 | 3 |
| 65 | 2 | 83 | 21 | 3 |
| 66 | 3 | 100 | 16 | 5 |
| 67 | 1 | 78 | 22 | 3 |
| 68 | 2 | 99 | 22 | 3 |
| 69 | | 82 | 10 | 1 |
| 1970 | 4 | 111 | 15 | 1 |
| 71 | 5 | 89 | 7 | 1 |
| 72 | | 82 | 21 | 7 |

| | $t_{\min} \leq +5^{\circ}$
"tr.froid" | $t_{\min} \leq +10^{\circ}$
"froid" | $t_{\max} \geq 30^{\circ}$
"chaud" | $t_{\max} \geq 35^{\circ}$
"tr.chaud" |
|------|--|--|---------------------------------------|--|
| 1973 | | 62 | 10 | 1 |
| 74 | 2 | 95 | 14 | 3 |
| 75 | 1 | 73 | 11 | 2 |
| 76 | | 78 | 18 | 5 |
| 77 | 2 | 90 | 25 | 7 |
| 78 | 5 | 85 | 9 | 2 |
| 79 | | 84 | 25 | 9 |
| 1980 | | 69 | 14 | 4 |
| 81 | 2 | 72 | 14 | 4 |
| 82 | 2 | 91 | 19 | 4 |
| 83 | 4 | 72 | 12 | 5 |
| 84 | | 83 | 13 | 3 |
| 85 | | 81 | 20 | 5 |
| 86 | 4 | 73 | 8 | 3 |
| 87 | 1 | 65 | 13 | 3 |
| 88 | | 55 | 13 | 3 |
| 89 | | 67 | 9 | 1 |
| 1990 | | 84 | 26 | 7 |
| 91 | | 62 | 11 | 3 |
| 92 | | 79 | 11 | 2 |
| 93 | | 65 | 13 | 5 |
| 94 | | 88 | 9 | 3 |
| 95 | 3 | 64 | 10 | 1 |
| 96 | 1 | 67 | 11 | 1 |
| 97 | | 89 | 31 | 7 |
| 98 | 1 | 61 | 11 | 1 |
| 99 | | 63 | 11 | |
| 2000 | | 84 | 20 | 4 |
| 01 | | 68 | 18 | 4 |
| 02 | 1 | 76 | 18 | 5 |
| 03 | 1 | 74 | 22 | 5 |
| 04 | | 81 | 19 | 6 |
| 05 | | 66 | 28 | 3 |
| 06 | | 76 | 18 | 6 |
| 07 | 3 | 61 | 7 | |
| 08 | | 82 | 23 | 2 |
| 09 | | 62 | 25 | 5 |
| 2010 | 2 | 66 | 22 | 4 |
| 11 | | 65 | 9 | 1 |
| 12 | | 83 | 18 | 2 |
| 13 | | 50 | 16 | 4 |
| 14 | | 57 | 25 | 3 |
| 15 | 1 | 71 | 27 | 7 |
| 16 | | 53 | 34 | 11 |
| 17 | | 72 | 27 | 8 |
| 18 | 1 | 70 | 28 | 6 |
| 2019 | | 67 | 30 | 9 |

De nouveau on se propose de dégager quelques traits remarquables présents dans ces données, au delà de la grande variabilité des nombres présentés et de la tendance évidente à la hausse des nombres de jours "chauds" et baisse des nombres de jours "froids". Les notations utilisées seront les mêmes que pour Rennes avec la même signification; les parties étant séparées les confusions ne sont pas à craindre.

Tout d'abord la Figure 5 montre la distribution croisée des nombres annuels des jours "chauds" et des jours "froids" à Sydney depuis 1925; en abscisse le nombre de jours "chauds" et en ordonnée le nombre de jours "froids", les points rouges étant des points doubles. Les 95 valeurs "jours chauds" sont comprises entre le minimum $\min_C = 6$ et le maximum $\max_C = 34$; leur médiane est $\text{med}_C = 17$. Pour les 95 valeurs "jours froids", $\min_F = 50$, $\max_F = 125$ et la médiane est $\text{med}_F = 83$.

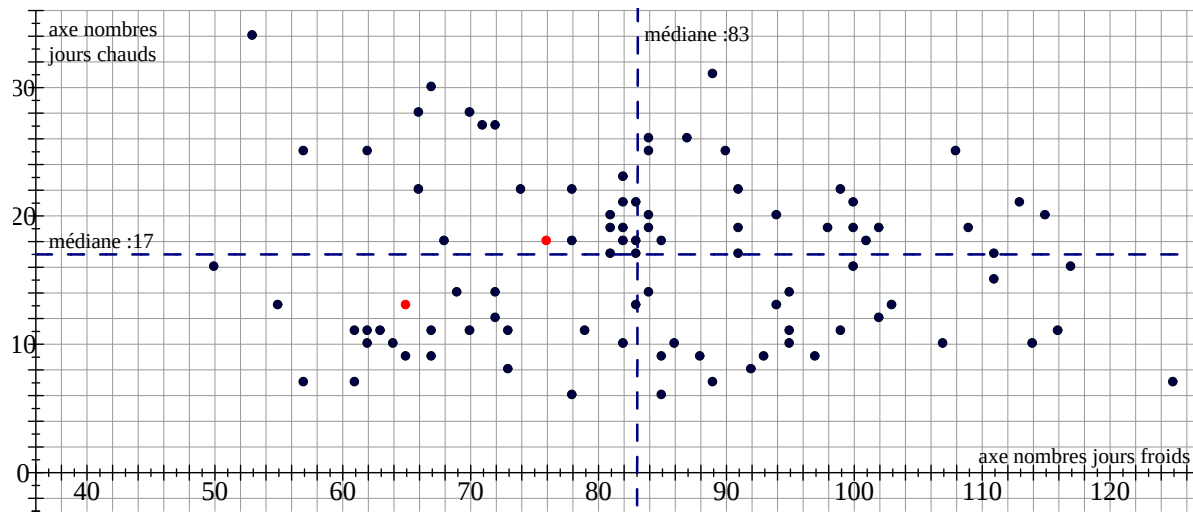


Figure 5. Distribution croisée "chauds"- "froids", Sydney, 1925-2019 . En pointillés : droite horizontale médiane des nombres de jours "chauds", droite verticale médiane des nombres de jours "froids". 2 points doubles en rouge.

Cette figure montre déjà un trait remarquable: comme pour Rennes, aucune liaison n'apparaît entre le nombre de jours "froids" d'un hiver et le nombre de jours "chauds" de l'été qui suit. La répartition des 95 années dans les quatre cases délimitées par les médianes est:

$\frac{23}{24} \mid \frac{23}{25}$ (tableau

légèrement approximatif car les points situés sur les médianes sont comptés pour 1/2 de chaque côté, le point central pour 1/4 et les totaux sont arrondis de 0,25; noter qu'il n'y a qu'un degré de liberté). Cette répartition est évidemment indistinguable d'une répartition au hasard ! L'Appendice 1 (n°2) redonne cette figure subdivisée en 9 cases délimitées par les terciles. Ce tableau confirmera l'observation actuelle: absence de corrélation apparente entre les nombre de jours "froids" et les nombres de jours "chauds" (le coefficient de corrélation linéaire vaut 0,07). Dans la suite les deux listes de valeurs seront analysées séparément.

VII. Analyse statistique résumée de la liste des nombres annuels de jours "chauds" à Sydney ($t_{\max} \geq 30^\circ$) depuis 1925.

L'analyse statistique de la liste des nombres de jours "chauds" $(C_n)_{1 \leq n \leq 95}$ de Sydney va suivre la même démarche que celle de la liste des nombres de jours "chauds" de Rennes (partie IV) avec des notations similaires. Les hypothèses testées seront notées $\mathcal{H}_S$ avec l'indice S . L'objectif de l'analyse est toujours le même: essayer de dégager la part du hasard. Pour alléger l'exposé, seulement un résumé des résultats sera présenté ici. L'Appendice 2 contient les détails des calculs et raisonnements.

La liste des nombres de jours "chauds" $(C_n)_{1 \leq n \leq 95}$ a pour caractéristiques, en reprenant les mêmes notations que pour Rennes, sans risque de confusion les parties étant disjointes:

$$\min_C = 6 ; \max_C = 34 ; M_C = 16,4 ; \sigma_C = 6,5 ; \text{med}_C = 17 ; Q_C^i = 11 ; Q_C^s = 21.$$

La valeur $\text{med}_C = 17$ est prise 4 fois. La droite d'approximation au sens des "moindres carrés" a pour équation:

$$y = 0,07x + 13,1 ; \text{pente } a_C = 0,07.$$

La représentation graphique est à la Figure 6. Pour tester l'hypothèse:

$$\mathcal{H}_S(C) : \text{"la suite } (C_n)_{1 \leq n \leq 95} \text{ forme une suite indépendante et stationnaire"}$$

les trois tests de la partie IV sont à nouveau employés. Leur conclusion est que, toujours au niveau de confiance 0,95, l'hypothèse $\mathcal{H}_S(C)$ n'est pas rejetée. Chacun des trois tests conclut à l'acceptation mais parfois de façon un peu limite. On pourrait donc considérer que le modèle à poser est simplement que la suite $(C_n)_{1 \leq n \leq 95}$ est une suite au hasard conforme à l'hypothèse $\mathcal{H}_S(C)$. Ce résultat peut être vu comme une conséquence de la faible valeur de la pente $a_C = 0,07$, donc de la tendance linéaire à l'augmentation, par rapport à la taille des valeurs elles mêmes, $M_C = 16,4$, et à l'amplitudes des variations, $\max_C - \min_C = 28$ et $\sigma_C = 6,5$. A la partie X on reviendra sur la distribution plausible de la suite $(C_n)_{1 \leq n \leq 95}$ et sur son éventuelle normalité.

La liste $(d_n)_{1 \leq n \leq 95}$ des fluctuations, c'est à dire des $d_n = C_n - 0,07n - 13,1$ différences entre C_n et la droite d'approximation, possède les caractéristiques:

$$\min_d = -11,7 ; \max_d = 14,7 ; M_d = 0 ; \sigma_d = 6,2 ; \text{med}_d = -0,4 ; Q_d^i = -4,4 ; Q_d^s = 4,1.$$

On reprend les trois tests de la partie IV pour éprouver l'hypothèse:

$$\mathcal{H}_S(d) : \text{"la suite } (d_n)_{1 \leq n \leq 95} \text{ forme une suite indépendante et stationnaire"}.$$

Cette hypothèse est acceptée de façon très nette par chacun des trois tests, de façon beaucoup plus nette que $\mathcal{H}_S(C)$. Cela conduit au modèle suivant pour représenter l'évolution du nombre annuel de jours "chauds" à Sydney depuis 1925:

$$C_n = 0,07n + 13,1 + d_n ; n = 1 \text{ pour } 1925$$

où $(d_n)_{1 \leq n \leq 95}$ forme une suite indépendante et stationnaire, centrée; la normalité de sa distribution sera discutée à la partie X.

Comme à la partie IV la liste des nombres de jours "chauds" tronquée de la période 1925-1969 peut aussi être considérée; c'est la liste $(C'_m)_{1 \leq m \leq 50}$ des nombres de jours "chauds" par an à Sydney depuis 1970: $C'_m = C_{m+45}$. Sa droite d'approximation a pour équation: $y = 0,23x + 11,4$; pente $a'_C = 0,23$ avec $x = 1$ pour 1970. La pente $a'_C = 0,23$ est nettement supérieure à la pente précédente $a_C = 0,07$. Voici la représentation graphique:

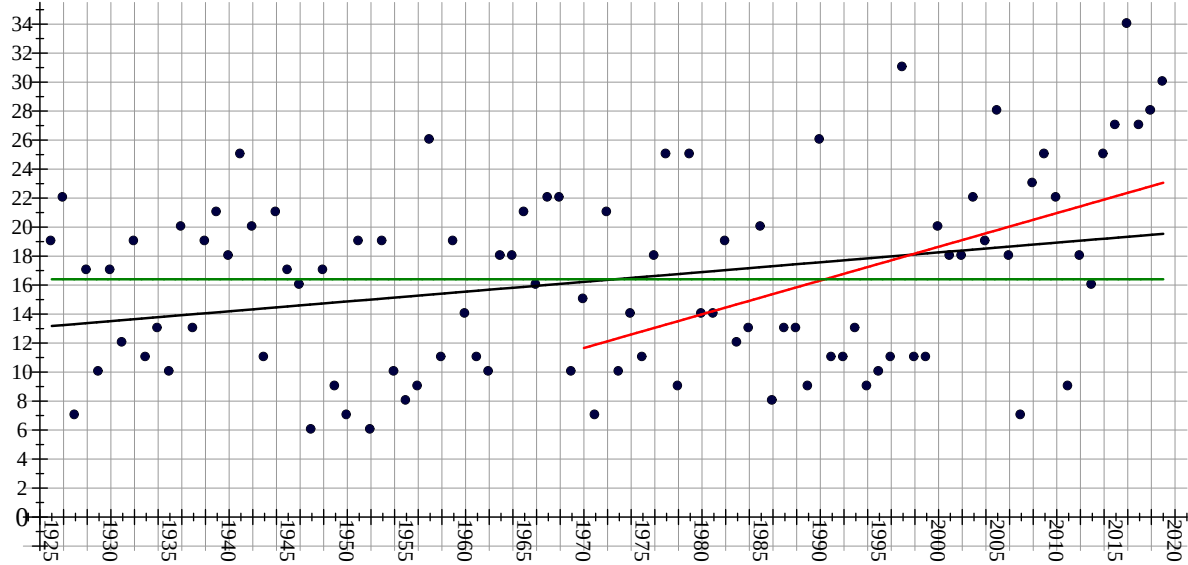


Figure 6. Nombre de jours "chauds" à Sydney en fonction de l'année; les droites d'approximation, depuis 1925 en noir et depuis 1970 en rouge. En vert la droite d'ordonnée la moyenne $M_C = 16,4$. Echelle verticale identique à celle des Fig. 2 et 3.

Au vu de la Figure 6 et de la différence des pentes a_C et a'_C , conclure à l'accélération du réchauffement semble évident, mais il faut encore se garder de donner une valeur prédictive à cette étude; d'ailleurs il est évident que prolonger vers le passé la droite d'approximation de la période 1970-2019, en rouge sur la Figure 6, ne donnerait pas une reconstitution valable. La forte pente a'_C met aussi en évidence la période de basses températures aux environs de 1970.

VIII. Analyse statistique résumée de la liste des nombres annuels de jours "froids" à Sydney ($t_{\min} \leq +10^\circ$) depuis 1925.

Suivant la même démarche on analyse maintenant la liste des nombres de jours "froids" $(F_n)_{1 \leq n \leq 95}$ de Sydney. Les détails des calculs sont à l'Appendice 2. Avec toujours les mêmes notations, cette liste $(F_n)_{1 \leq n \leq 95}$ a les caractéristiques:

$$\min_F = 50 ; \max_F = 125 ; M_F = 83,7 ; \sigma_F = 16,6 ; \text{med}_d = 83 ; Q_F^i = 70 ; Q_F^s = 95.$$

La droite d'approximation au sens des "moindres carrés" a pour équation:

$$y = -0,45x + 105,1 ; \text{pente } a_F = -0,45.$$

La représentation graphique est à la Figure 7. Pour tester l'hypothèse:

$$\mathcal{H}_S(F) : \text{"la suite } (F_n)_{1 \leq n \leq 95} \text{ forme une suite indépendante et stationnaire"}$$

les trois tests de la partie IV sont repris; leur conclusion est que, au même niveau de confiance 0,95, l'hypothèse $\mathcal{H}_S(F)$ est rejetée de façon très nette. Chacun des trois tests conclut au rejet.

Alors en formant la liste $(g_n)_{1 \leq n \leq 95}$ des fluctuations, c'est à dire des différences entre F_n et la valeur de la droite d'approximation, soit $g_n = F_n + 0,45n - 105,1$, on obtient une liste dont les caractéristiques sont:

$$\min_g = -36,5 ; \max_g = 26,4 ; M_g = 0 ; \sigma_g = 11,2 ; \text{med}_g = -0,9 ; Q_g^i = -7,7 ; Q_g^s = 7,4.$$

L'écart-type $\sigma_g = 11,2$ est beaucoup inférieur à $\sigma_F = 16,6$.

En utilisant de nouveau les trois tests pour éprouver l'hypothèse:

$\mathcal{H}_S(g) : "la suite (g_n)_{1 \leq n \leq 95} \text{ forme une suite indépendante et stationnaire}"$.

une acceptation large est obtenue, pour chacun des trois tests. On peut donc poser le modèle suivant, représentant l'évolution du nombre annuel de jours "froids" à Sydney depuis 1925:

$$F_n = -0,45n + 105,1 + g_n ; n = 1 \text{ pour } 1925$$

où $(g_n)_{1 \leq n \leq 95}$ forme une suite indépendante et stationnaire, centrée (la normalité de sa distribution sera discutée à la partie X).

Comme précédemment, peut aussi être étudiée la liste des nombres de jours "froids" tronquée de la période 1925-1969; c'est la liste $(F'_m)_{1 \leq m \leq 50}$ des nombres de jours "froids" par an à Sydney depuis 1970: $F'_m = F_{m+45}$. Sa droite d'approximation a pour équation: $y = -0,41x + 84,5$; pente $a'_F = -0,41$ avec $x = 1$ pour 1970. La pente $a'_F = -0,41$ est quasiment égale à la pente $a_F = -0,45$ trouvée pour la période 1925-2019. Voici la représentation graphique:

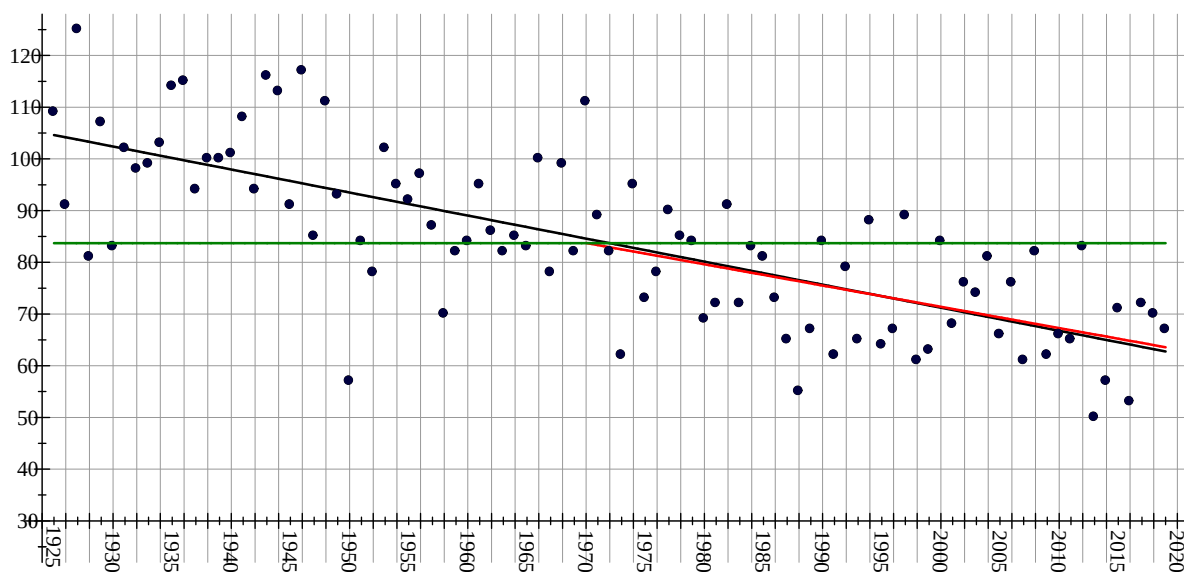


Figure 7. Nombres de jours "froids", à Sydney, en fonction de l'année, avec droites des moindres carrés : noire 1925-2019 ; rouge 1970-2019 . En vert la droite d'ordonnée la moyenne $M_F = 83,7$. L'échelle verticale est celle de la Fig.4; l'origine des ordonnées est translatée.

La droite d'approximation pour la période 1970-2019 représentée en rouge est presque confondue avec la droite de la période 1925-2019 tracée en noir.

IX. Sur les nombres de jours "très chauds" ou "très froids" à Rennes et à Sydney depuis 1925.

A première vue les nombres de jours qualifiés "très chauds" ou "très froids", donnés par les Tableaux 2 et 3, obéissent aux mêmes évolutions que les nombres de jours "chauds" ou "froids", aussi bien à Rennes qu'à Sydney.

Pour Rennes le grand nombre d'années sans aucun jour "très chaud" empêche une analyse détaillée. On observe que sur les 17 années, depuis 1925, durant lesquelles la température d'un jour au moins a atteint 35° , il n'y en a eu que 5 avant 1975. Les trois années ayant subi le plus grand nombre de jours "très chauds" sont 2003 avec 6 jours "très chauds", 1976 et 1990 avec 5, aucune année de la dernière décennie, ce qui ne confirme pas l'opinion courante à Rennes que les vagues de fortes chaleurs sont de plus en plus fréquentes. D'autre part sur les 26 années sans jour "très froid" il n'y en a que 6 avant 1970. Durant la dernière décennie, 2010-19, 5 années n'ont connu aucun jour "très froid", comme durant la décennie 1990-99 et la décennie 1970-79.

Pour Sydney c'est la liste des jours "très froids" ($t_{\min} \leq +5^\circ$) qui comporte le plus de 0: il y en a 35 dont 30 depuis 1970; durant la dernière décennie 7 années sans jour "très froid" comme durant la décennie 1990-99.

Par contre la liste des nombres de jours "très chauds" à Sydney peut-être analysée: sa droite d'approximation au sens des "moindres carrés" a pour pente 0,05 (inférieure à 0,07 la pente de la droite d'approximation des nombres de jours "chauds"). En reprenant la même démarche que précédemment pour tester l'hypothèse $\mathcal{H}_S(TC)$: "*la suite $(TC_n)_{1 \leq n \leq 95}$ des nombres de jours "très chauds" (depuis 1925) forme une suite indépendante et stationnaire*" celle-ci n'est pas rejetée. C'est le même résultat que pour l'hypothèse $\mathcal{H}_S(C)$ concernant la liste des jours "chauds". Il faut donc comprendre que sur presque cent ans la variation en tendance du nombre de jours "très chauds" est mineure par rapport à l'amplitude des fluctuations "au hasard" annuelles. Il est remarquable que le résultat est le même pour la liste des nombres de jours "très chauds" depuis 1970. L'hypothèse $\mathcal{H}'_S(TC)$: "*la suite $(TC_n)_{1 \leq n \leq 50}$ des nombres de jours "très chauds" (depuis 1970) forme une suite indépendante et stationnaire*" est acceptable pour les trois tests, alors que l'hypothèse analogue $\mathcal{H}'_S(C)$ pour la liste des nombres de jours chauds depuis 1970 est rejetée (voir Appendice 2). L'augmentation du nombre de jours "très chauds" à Sydney ne montre pas de tendance vraiment importante, sinon peut-être que le maximum de cette liste a été atteint en 2016 avec 11 jours "très chauds"; par exemple, le nombre de jours "très chauds" de l'été qui vient de s'écouler (2019-20) est 9, comme pour l'année 1979 et l'année 1925. A Sydney le réchauffement en cours depuis cent ans se produit surtout par le bas: ce sont les températures basses, prises durant les nuits d'hiver, qui deviennent de plus en plus rares.

X. Interprétation des modèles posés: présence de la loi normale et phénomènes extrêmes.

Les modèles que nous avons posés pour les nombres de jours "chauds" ou "froids", aussi bien pour Rennes que pour Sydney, sont des modèles stochastiques qui permettent de concevoir ces listes de nombres comme des sommes d'une fonction linéaire, croissante ou décroissante, et de termes aléatoires formant une réalisation d'une suite de variables aléatoires indépendantes et de même loi, c'est à dire de même distribution.

Pour les nombres de jours "chauds" de Rennes ou de Sydney les pentes des parties linéaires des modèles sont égales. Par contre pour les nombres de jours "froids" ces pentes sont très différentes. Une autre différence entre jours "chauds" et jours "froids" est que pour Rennes comme pour Sydney les modèles posés pour les jours "chauds" comportent des parties linéaires différentes pour les deux périodes 1925-2019 et 1970-2019 alors que pour les jours "froids", dans les deux cas, ces parties linéaires sont presque les mêmes, comme le montrent les Figures 3 et 6. Ces observations pourraient n'être que des conséquences du choix conventionnel des définitions de "chaud" et "froid". Pour obtenir une interprétation valable il faut maintenant essayer d'identifier la loi des termes aléatoires introduits. Avec une liste de 90 ou 95 valeurs réparties dans un intervalle de variation d'amplitude 25 ou 30 ou davantage, d'écart-type 5 ou plus, une identification précise est impossible. Nous allons cependant essayer de répondre à la question: la loi des fluctuations dans les différents modèles pourrait-elle être considérée comme une loi "normale"?

Commençons avec les jours "chauds" de Rennes. Examinons la liste $(d_n)_{1 \leq n \leq 90}$ des fluctuations. Elle est centrée en raison de la définition même de la droite d'approximation au sens des "moindres carrés"; ses caractéristiques ont été données à la partie IV. Pour tester la normalité de la distribution des $(d_n)_{1 \leq n \leq 90}$ on peut s'appuyer sur le test de Kolmogorov ([Bo] [S]), en ne prenant en compte que les positions de la médiane et des quartiles. On trouve d'une part que le quotient $\text{med}_d / \sigma_d = -0,15$ correspond au niveau 0,44 pour la loi normale standard (ou loi normale réduite); il n'y a pas de différence significative, au niveau de confiance 0,95, entre 0,44 et le niveau théorique 0,50 puisque la borne fixée par le test de Kolmogorov vaut $1,36/\sqrt{90} = 0,14$.

D'autre part $Q_d^i/\sigma_d = -0,59$ et $Q_d^s/\sigma_d = 0,42$ correspondent respectivement aux niveaux 0,28 et 0,34 pour la loi normale standard et il n'apparaît donc pas non plus de différence significative avec le niveau théorique 0,25. C'est un peu grossier puisqu'en fait la fonction de répartition empirique n'est comparée avec la fonction de répartition normale standard qu'en trois points: la médiane et les deux quartiles empiriques; mais ici nous recherchons un résultat qualitatif, une trop grande précision de calcul serait illusoire.

On peut aussi tester la normalité en considérant la valeur maximum de la liste. Ici pour $(d_n)_{1 \leq n \leq 90}$ c'est $\max_d = 15,8$ soit presque $3\sigma_d$. Pour la loi normale standard la valeur 3 n'est atteinte qu'avec la probabilité 0,0014. Mais pour un échantillon de taille 90 comme celui considéré ici, la probabilité de voir apparaître une valeur aussi grande est environ $1 - (1 - 0,0014)^{90} = 0,12$. Au niveau de confiance 0,95, l'hypothèse de normalité n'est donc pas rejetée. La liste $(d_n)_{1 \leq n \leq 90}$ présente une dissymétrie: elle est moins étalée du côté négatif, mais cela provient pour une bonne part du choix de la convention $t_{\max} \geq 30^\circ$ pour définir le jour "chaud" qui a engendré 10 valeurs de C_n bloquées à 0. En conclusion il n'est donc pas interdit de supposer que la loi commune des $(d_n)_{1 \leq n \leq 90}$ soit une loi normale centrée d'écart type environ 5 ou 6 mais tronquée de ses valeurs négatives de plus grandes valeurs absolues.

Dans chacune des autres situations - jours "froids" de Rennes, jours "chauds" ou "froids" de Sydney - ces mêmes arguments peuvent être réemployés: ils montrent que l'hypothèse de normalité pour les lois des fluctuations est acceptable. Le détail de cette analyse est donné dans l'Appendice 2. Ici nous indiquons seulement l'image qualitative qui s'en dégage.

Pour les listes de nombres de jours "chauds" de Rennes et de Sydney la différence des écarts-types est faible, d'autant plus que celui de Rennes est minoré du fait de la troncature de la loi des fluctuations mentionnée ci-dessus. Nous arrivons donc à cette conclusion assez remarquable: une fois privées de leurs parties linéaires dont les pentes sont égales, qui donnent les tendances, il est plausible que ces deux listes soient soumises à des fluctuations qui sont des valeurs observées d'une suite de variables aléatoires indépendantes, régies par une même loi normale centrée d'écart-type environ 6 (rappelons qu'ici l'unité a un sens simple: c'est le jour "chaud" par an).

L'intérêt principal de cette remarque sur la normalité des fluctuations est qu'elle conduit à une meilleure interprétation des modèles posés dans les parties précédentes.

Pour les nombres de jours "chauds" de Rennes et de Sydney depuis 1925 nous avons deux équations de formes voisines:

$$C_n = 0,07n + 2,7 + d_n ; \text{ avec } (d_n)_{1 \leq n \leq 90} \text{ v.a.i.i.d. normales centrées } \sigma_d \simeq 5$$

$$C_n = 0,07n + 13,1 + d_n ; \text{ avec } (d_n)_{1 \leq n \leq 95} \text{ v.a.i.i.d. normales centrées } \sigma_d \simeq 6$$

(v.a.i.i.d. pour variables aléatoires indépendantes et identiquement distribuées). La pente 0,07 dit que sur cent ans il y a eu "en moyenne" une augmentation de 7 du nombre annuel de jours "chauds", mais cette augmentation est accompagnée de fluctuations annuelles indépendantes et stationnaires dont l'ordre de grandeur dépasse sensiblement 7. Avec l'écart-type valant 6, un intervalle de confiance au niveau 0,95 pour ces fluctuations, que l'on peut supposer normales, est $[-12, +12]$. Distinguer la tendance de la fluctuation est donc difficile! La différence des termes constants rappelle que les nombres sont plus élevés à Sydney; en fait en ajoutant une constante d'environ 10 à la liste des C_n de Rennes on trouverait une liste indistinguishable de la liste des C_n de Sydney. Cette observation est assez surprenante et demanderait un approfondissement; elle indique que l'augmentation des températures élevées a été plutôt moins forte à Sydney qu'à Rennes. Voici deux figures illustrant ces observations:

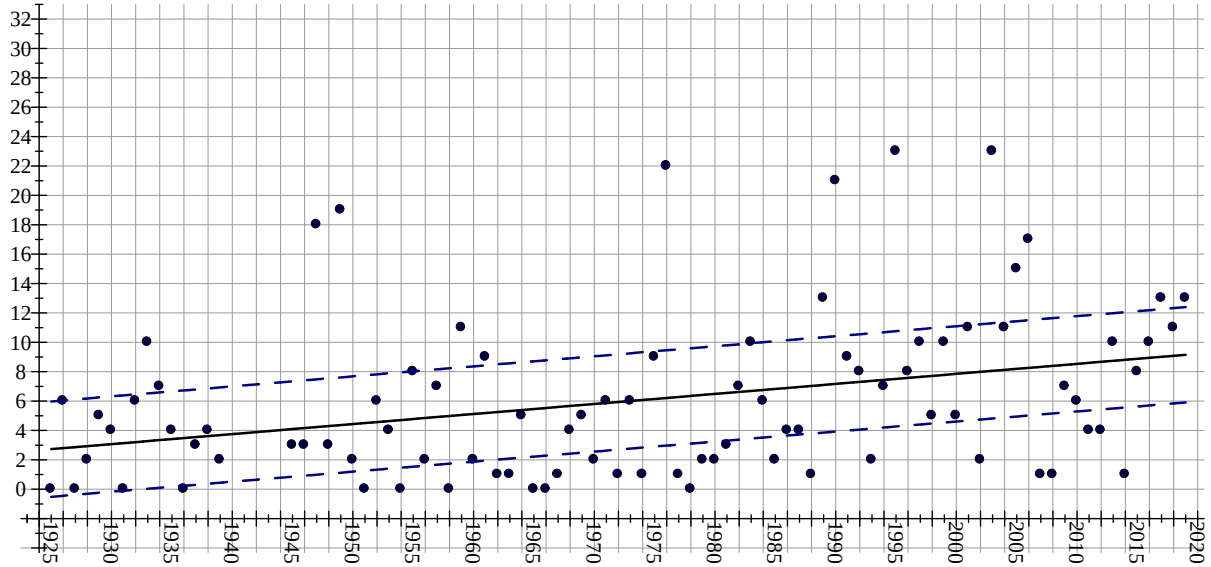


Figure 8. Nombres de jours "chauds" à Rennes. La "droite des moindres carrés" est tracée au centre d'une bande de largeur $\sim 6,5 \simeq 0,07 \times 95$ (variation tendancielle sur 95 ans). Celle-ci ne contient que la moitié des points : la "tendance" est absorbée par les fluctuations.

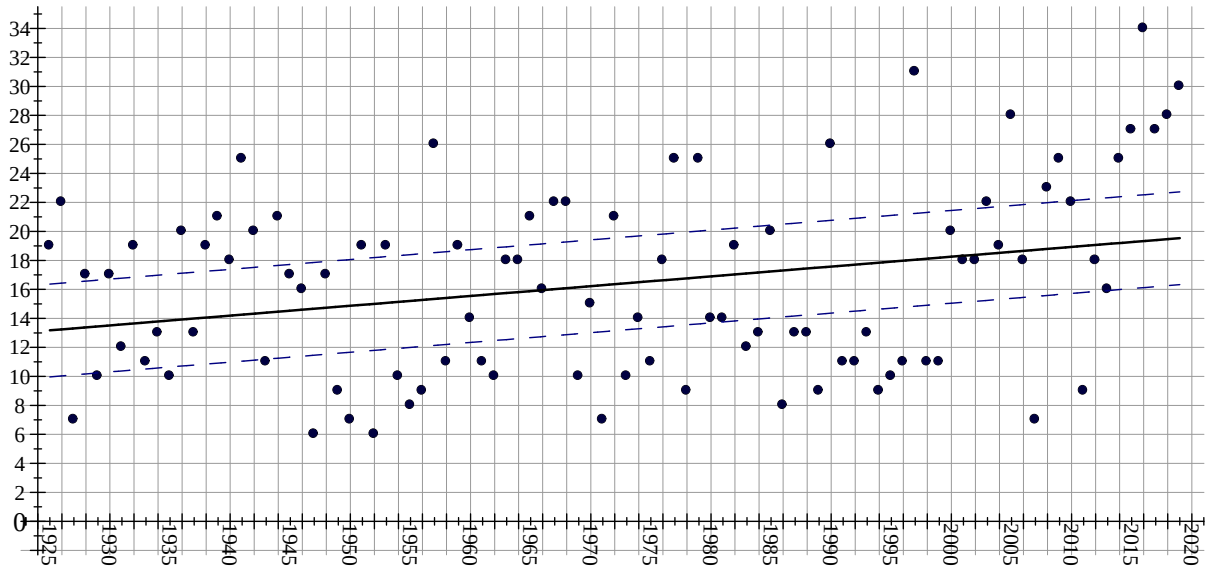


Figure 9. Nombres de jours "chauds" à Sydney. La "droite des moindres carrés" est tracée au centre d'une bande de largeur $\sim 6,5 \simeq 0,07 \times 95$ (variation tendancielle sur 95 ans). Celle-ci ne contient que le tiers des points : la "tendance" est absorbée par les fluctuations. Fig. 8 et 9 ont même échelle verticale.

Pour les listes de nombres de jours "froids" de Rennes et de Sydney on arrive aussi à la conclusion que les fluctuations de ces nombres, autour de leurs tendances linéaires principales, peuvent être considérées comme indépendantes et obéissant à une même loi normale centrée d'écart-type environ 12. En fait c'est l'écart-type des suites initiales des nombres de jours "chauds" ou "froids" qui donne, sans surprise, l'ordre de grandeur de l'écart-type, un peu inférieur, de la loi normale à envisager pour modéliser les fluctuations. Cependant Rennes et Sydney se distinguent nettement. Pour Rennes notre modèle pour les nombres de jours "froids" est:

$$F_n = -0,17n + 45,7 + g_n; \text{ avec } (g_n)_{1 \leq n \leq 90} \text{ v.a.i.d. normales centrées } \sigma_g \simeq 13$$

alors que pour Sydney:

$$F_n = -0,45n + 105,1 + g_n; \text{ avec } (g_n)_{1 \leq n \leq 95} \text{ v.a.i.i.d. normales centrées } \sigma_g \simeq 11.$$

(v.a.i.i.d. pour variables aléatoires indépendantes et identiquement distribuées). La première équation indique une baisse en moyenne de 17 unités pour le siècle passé accompagnée de fluctuations annuelles indépendantes et stationnaires dont l'intervalle de confiance est $[-2\sigma_g = -26, +2\sigma_g = +26]$. Comme pour les jours "chauds" la tendance linéaire est dominée, absorbée, par les fluctuations; elles sont donc difficiles à séparer. Par contre, pour Sydney la tendance linéaire à la baisse provoque une baisse de 45 unités, en moyenne sur un siècle, alors que les fluctuations annuelles restent à peu près dans le même intervalle $[-2\sigma_g = -22, +2\sigma_g = +22]$. Les fluctuations n'absorbent plus la tendance linéaire: celle-ci domine nettement.

Les deux Figures suivantes illustrent ce phénomène:

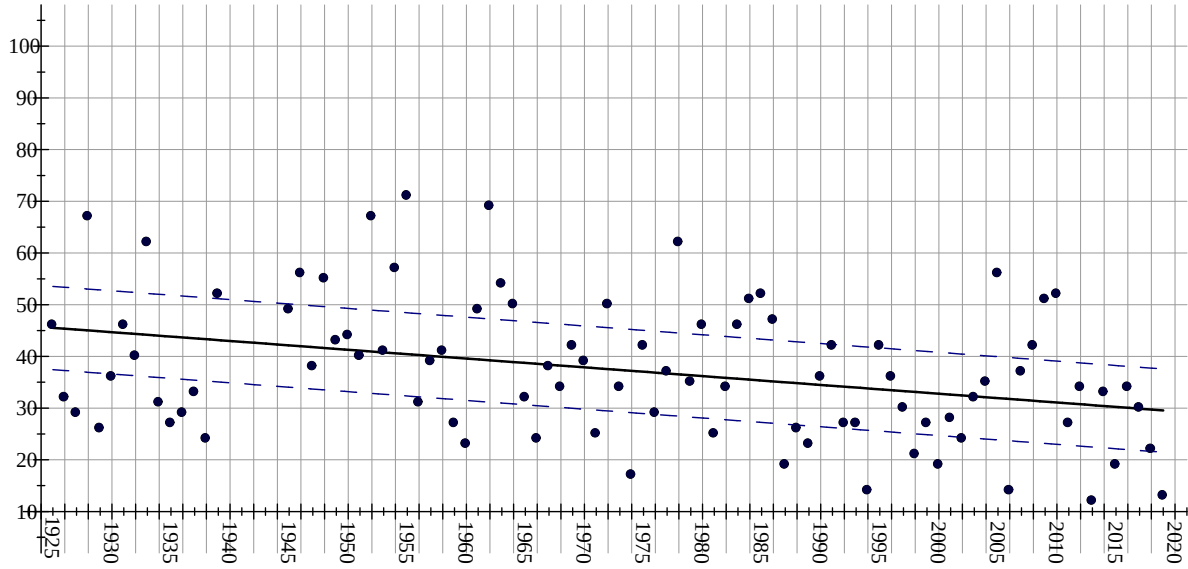


Figure 10. Nombres de jours "froids" à Rennes. La "droite des moindres carrées" est placée au centre d'une bande de largeur $\sim 16 \simeq 0,17 \times 95$ (variation tendancielle sur 95 ans). Cette bande ne contient que la moitié des points.

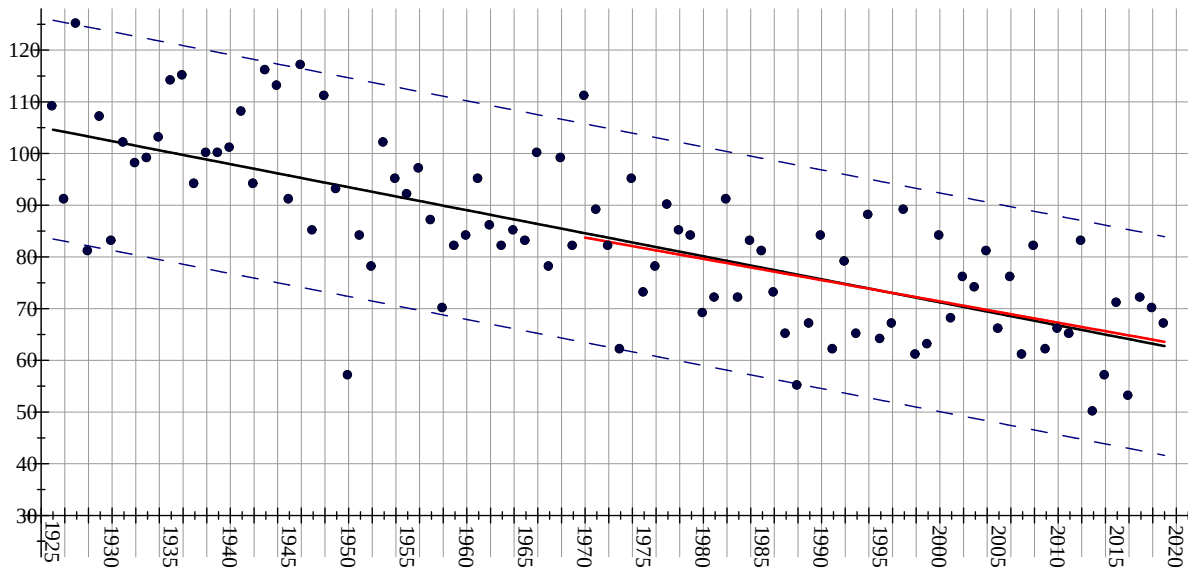


Figure 11. Nombres de jours "froids" à Sydney. La "droite des moindres carrées" est placée au centre d'une bande de largeur $\sim 43 \simeq 0,45 \times 95$ (variation tendancielle sur 95 ans). Cette bande contient presque tous les points. Fig. 10 et 11 ont même échelle verticale avec des origines différentes.

C'est une des principales observations de cette étude: la baisse du nombre annuel de jours "froids" à Sydney est d'un ordre de grandeur sensiblement supérieur à celui des fluctuations "au hasard" annuelles; on pourrait presque parler d'une baisse linéaire. De plus cette baisse domine l'augmentation du nombre annuel de jours "chauds". Cela suggère que l'amplitude annuelle des températures a tendance à diminuer; ce fait, qui demande une vérification, serait à rapprocher de la diminution de l'amplitude journalière des températures mentionnée dans le rapport [GIEC2013] (Chap.2, p.188).

La notion de phénomène extrême ne peut être discutée sans une précision sur le sens de cette expression. Un premier sens est simplement celui de phénomène rare, qui n'a qu'une probabilité faible de se produire, mais qui cependant reste dans le domaine des possibilités prévues par la loi de probabilité gouvernant le type de phénomène considéré. C'est en ce sens que le rapport du GIEC ([GIEC2007] The physical science basis, p.53) explique que l'élévation de la température moyenne entraîne une translation vers le haut de la loi de probabilité gouvernant la distribution des températures ce qui provoquera davantage de phases de températures élevées constituant des "phénomènes extrêmes". Mais dans un deuxième sens, souvent implicite en statistique, un phénomène extrême est un phénomène qui sort nettement du domaine des possibilités envisagées; par exemple pour une statistique qu'on suppose gouvernée par une loi normale ce serait une valeur qui s'écarterait de la moyenne de nettement plus que trois fois l'écart-type. Dans notre étude ne sont apparus que des phénomènes extrêmes au premier sens.

Par exemple durant l'année 2016 Sydney a connu 34 jours "chauds", le maximum depuis cent ans; cela correspond à une fluctuation d valant environ 15 alors que l'écart-type de la fluctuation vaut 6,2 (voir Fig. 6) donc cette valeur 15 reste comprise entre 2 et 3 fois l'écart-type ce qui, dans le cas normal, est peu probable mais cependant possible. Un autre exemple est celui du nombre de jours "froids" de 1950 valant 57 mais correspondant à une fluctuation de $-36,5$ alors que l'écart-type des fluctuations vaut 11,2. Sous l'hypothèse de normalité une telle valeur n'a que la probabilité 0,0007 de se produire; mais sur 95 tirages indépendants son occurrence a la probabilité $1 - (1 - 0,0007)^{95} \simeq 0,06$ donc reste possible.

Dans la présente étude aucun phénomène extrême au second sens n'a été rencontré. Le phénomène le plus étonnant observé c'est la baisse du nombre annuel de jours "froids" à Sydney qui s'opère rapidement et presque linéairement sans fluctuations importantes (voir Fig. 11)!

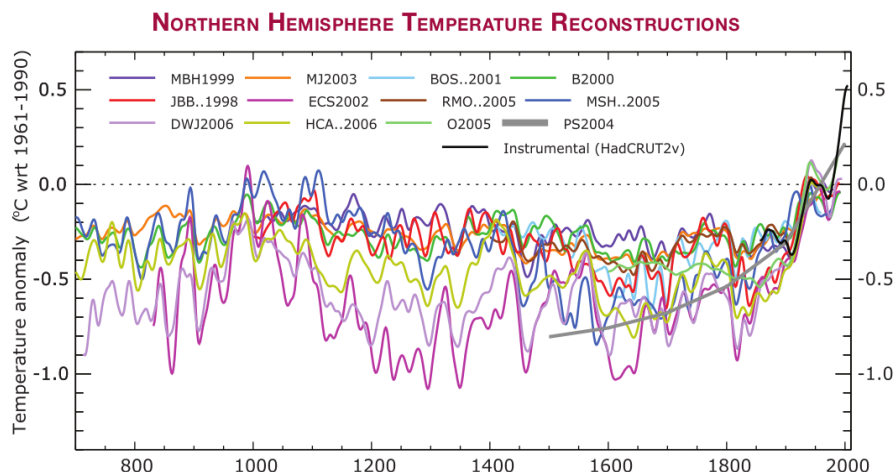
XI. Remarques méthodologiques.

A Analyse statistique et prévisions.

Savoir qu'un phénomène s'est déjà produit conduit bien sûr à croire qu'il pourra arriver à nouveau, mais au delà de cette évidence banale, pour qu'une analyse statistique puisse vraiment permettre des prévisions, ce que nous nous sommes gardés de faire ici, il est nécessaire qu'elle porte sur des données qui ne soient pas à relevé unique, qui puissent être reprises dans le cadre d'un plan d'expérience. Ce que nous avons appelé modèle ne peut servir, éventuellement, qu'à une meilleure compréhension de phénomènes passés, en amenant des questions et suggérant des hypothèses; d'ailleurs les paramètres de ces modèles sont extraits des mesures passées. La possibilité de répéter les expériences ou les raisonnements est une exigence de la science expérimentale. Mais le climat de la Terre et son évolution font partie de l'Histoire; les événements ne se produisent qu'une fois. La climatologie se trouve à mi-chemin entre la Science expérimentale et l'Histoire.

Cependant il faut essayer de prévoir. Les modèles physiques élaborés à but explicatif et prévisionnel, dont la mise en oeuvre demande des moyens informatiques puissants, ne peuvent être évalués que par comparaison avec les mesures effectives. Depuis quinze ans les spécialistes du climat ont publié bon nombre d'ouvrages exposant les principes de ces modèles, mais une certaine frustration peut être ressentie devant la part réduite que prennent dans ces ouvrages l'étude des mesures réalisées depuis des dizaines d'années. Le plus souvent la "mesure" qui sert de référence est la "température moyenne globale" à la surface de la terre. La mesure de celle-ci est l'apanage d'instituts puissants utilisant de nombreux satellites artificiels mesurant

depuis l'espace les températures au sol suivant des lois physiques sophistiquées: le Goddard institute, branche de la NASA, le Hadley Center UK Met Office, le NOAA... Ces instituts sont aussi souvent auteurs des modèles. Cette "température moyenne globale" est présentée dans des graphiques parfois sibyllins: par exemple voici ci-dessous le diagramme connu sous le nom de "diagramme des spaghettis" extrait du rapport [GIEC2007] (The physical science basis, p.55):



Cette figure est reprise dans [Ma] p.50 et dans [PoBrMa] p.182 avec la légende: "estimation des changements de température de surface au cours des derniers 1300 ans: 14 reconstructions différentes pour l'hémisphère nord...". Au paragraphe suivant on reviendra sur cette notion de moyenne globale.

Pourtant l'étude des mesures passées peut apprendre beaucoup. Au lieu des méthodes élémentaires de la présente étude, l'analyse multidimensionnelle des données et l'analyse des séries chronologiques pourraient être utilisées. Par exemple dans [PrFo] ou [OgLiJal] les auteurs utilisent les processus stochastiques autorégressifs afin d'isoler la part du hasard dans l'évolution climatique, suivant toujours un point de vue phénoménologique. Cela ne s'oppose pas à la recherche des causes physiques. La question du rapport entre hasard et causalité n'est-elle pas toujours ouverte depuis les origines du Calcul des Probabilités? Cette importante discussion a mobilisé de grands penseurs; nous renvoyons à deux ouvrages contemporains [D], [F].

B Sur la moyenne arithmétique.

Comme le rythme des saisons et la périodicité annuelle sont toujours dominants dans les séquences de températures relevées jour après jour sur une longue durée, il est nécessaire d'éliminer cette période annuelle pour ne garder que les variations d'une année à l'autre. Le plus souvent pour procéder à cette élimination on a recours à un calcul de moyenne sur une année; on parle d'année plus ou moins chaude "en moyenne". Ce calcul de moyenne est présenté comme une évidence, ne demandant pas à être explicité. Cependant cela interroge. Pour de nombreux sites météo la température moyenne d'un jour n'est rien d'autre que la demie somme de la température maximum et de la température minimum de ce jour; ensuite pour une durée d'un mois ou d'un an, la moyenne arithmétique de ces "températures moyennes" quotidiennes est calculée. Pour d'autres cette méthode apparaîtra sommaire; ils demanderont la moyenne par jour prise sur 24 mesures, une par heure, ou plus encore. Mais obtenu d'une façon ou d'une autre ce nombre moyen n'est plus vraiment une mesure: c'est un indice permettant des comparaisons. Pour le public qui s'informe des températures quotidiennes, maximum et minimum, cette moyenne n'est pas significative.

Ici la périodicité annuelle a été éliminée par le comptage par an de jours "chauds" ou "froids". Bien que cette dénomination repose sur une règle conventionnelle, on saisit aussitôt ce que ces nombres mesurent. Prendre ensuite la moyenne arithmétique des nombres de Rennes et de Sydney n'aurait aucun sens. Mais élargir le champ d'étude à d'autres lieux choisis de façon à

représenter la Terre est possible: il faudrait construire les histogrammes et détailler les comparaisons.

Pour représenter l'évolution du climat de la Terre on fait souvent référence à la "température globale moyenne". Celle-ci repose sur deux opérations successives: une moyenne sur l'année prise en un grand nombre de lieux répartis sur toute la Terre puis une moyenne par rapport à cette ensemble de lieux. Cette dernière moyenne est éventuellement pondérée par des superficies (dans [JD] il est expliqué que la température est relevé par des satellites dans 2592 cellules tracées abstraitement à la surface du globe en subdivisant les méridiens et les parallèles de 5° en 5° ; ainsi donc dans un disque centré au pôle de rayon 555 km on effectuerait 72 mesures, mais dans une zone de même superficie située à l'équateur on en ferait 3 ou 4!). Cette température "globale moyenne" est souvent donnée au centième près; il n'est pas rare de lire des phrases comme "la température moyenne a augmenté de $0,87^\circ$ ". Cela interroge: les mesures des stations météo sont au dixième de degré près; le calcul de la moyenne arithmétique ne peut pas changer la précision des mesures (sauf si l'on s'appuie sur la théorie des erreurs de Gauss mais celle-ci ne s'applique que dans des conditions bien précises). D'ailleurs cette précision semble peu signifiante; dans [GIEC1990] p.xii on lisait: "Global-mean-surface air temperature has increased by 0.3° to 0.6°C over the last 100 years".

Cette moyenne globale fournit un indice dont on peut noter les évolutions, mais peut-on la considérer comme une température relevant des lois de la thermodynamique: par exemple utilisable dans la loi de Stefan-Boltzmann [Ka] [PoBrMa] (pour cette loi de la thermodynamique voir [Br] chap.XVII) ou mesurant l'enthalpie de la Terre [Ma] (pour la notion d'enthalpie voir [Br] chap.XI)? Cet indice "température globale moyenne", utilisé comme référence principale, a de plus le désavantage de ne plus être accessible qu'au très restreint groupe de scientifiques disposant des systèmes satellitaires de mesure, dépendants des instituts déjà évoqués ci-dessus, et de faire tomber le problème de l'évolution du climat dans le domaine de ce qu'on appelle parfois "big science": à moins de disposer de moyens énormes les scientifiques sont réduits à la condition de spectateurs.

Le recours à la moyenne arithmétique dans l'étude des phénomènes climatiques est parfois présenté comme une nécessité imposée par la loi des grands nombres. En effet le système climatique est considéré comme un système chaotique ([PoBrMa] p.95) qui n'est prévisible qu'à très court terme mais dont l'avenir peut quand même être décrit en moyenne grâce à la loi des grands nombres (ou au "théorème ergodique" qui en est une forme générale [G]). Ce point mérite une petite digression. Un des plus simples systèmes chaotiques que l'on peut concevoir est celui qui consiste en une suite de tirages au sort d'une urne contenant des boules noires ou blanches; l'évolution du système est simplement la suite des couleurs observées successivement. Les tirages sont effectués avec remise de façon à vérifier l'indépendance stochastique et la stationnarité (homogénéité). A court terme rien n'est prévisible. Mais en effet, grâce à la loi des grands nombres, sur une longue série on peut prévoir que la proportion de boules blanches tirées sera voisine de la proportion réelle de boules blanches présentes dans l'urne ([F]). Cependant cette prévision suppose la connaissance *a priori* de cette proportion réelle; de plus le nombre de tirages nécessaires pour une prévision valable dépend aussi de celle-ci (on dit souvent qu'il faut 30 ans pour qu'une évolution observée du climat devienne vraiment significative; notre étude indique que même 30 ans peut-être insuffisant comme l'a montré la Figure 3). Expérimentalement c'est plutôt la réciproque de la loi des grands nombres qu'il faut utiliser: ne connaissant pas la composition de l'urne *a priori* on peut la déduire approximativement *a posteriori* de la fréquence de boules blanches observée dans une longue série de tirages, mais alors plus question de prévision. N'est-ce pas plutôt ainsi que se pose le problème de l'évolution climatique? En fait ce problème se présente de façon plus difficile encore: un "mauvais génie" caché pourrait introduire subrepticement des boules blanches ou noires dans l'urne et ainsi rendre le processus de tirages non stationnaire. Comment faire pour s'en apercevoir? La loi des grands nombres (ou le "théorème ergodique") qui suppose la stationnarité, ne suffit plus à nous éclairer!

Encore un mot sur la moyenne arithmétique. Les études des milieux désordonnés, par exemple celle de la diffusion en milieu non homogène, montrent que l'homogénéisation correcte

demande d'autres opérations que la moyenne arithmétique; elle peut demander par exemple d'utiliser la moyenne harmonique (règle des résistances électriques en parallèle). On sait aussi que les régimes apparaissant dans les diverses formes réelles du milieu peuvent être très différents du régime créé par la moyennisation du milieu ([JiKO]).

XII. Quelques conclusions.

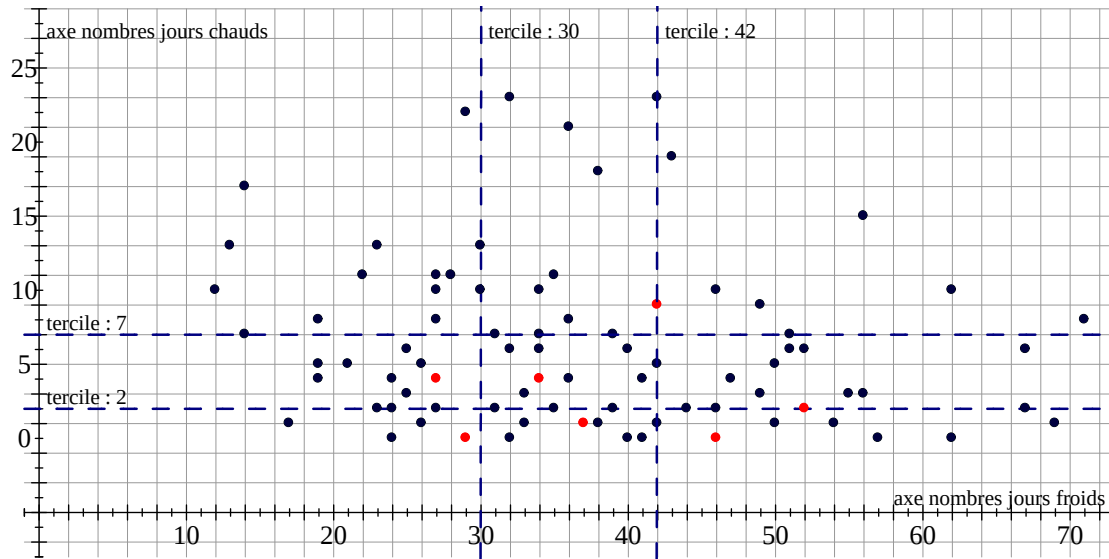
Le comptage des jours "chauds" ou "froids" par année nous a permis d'éliminer la périodicité saisonnière annuelle de l'évolution des températures en conservant la transparence des valeurs mesurées donc en permettant leur compréhension intuitive. Voici de façon résumée nos principales observations:

- Il n'apparaît pas de corrélation entre le nombre de jours "chauds" et celui des jours "froids" d'une même année.
- Les évolutions depuis 1925 de ces nombres, pour Rennes et pour Sydney, qui montrent une augmentation du nombre annuel de jours "chauds" et une baisse de celui des jours "froids", ne peuvent pas être attribuées au hasard: c'est flagrant pour les jours "froids" de Rennes ou Sydney mais moins net pour les jours "chauds" en particulier ceux de Sydney.
- Il suffit de retrancher à ces évolutions une fonction linéaire bien choisie, croissante pour les jours "chauds", décroissante pour les jours "froids", pour mettre en évidence des fluctuations que l'on peut considérer comme des suites stochastiquement indépendantes, stationnaires et qui obéissent à une loi statistique qu'il est plausible de considérer comme "normale". De façon assez surprenante on observe que la taille de ces fluctuations est en gros la même pour Rennes et Sydney, que ce soit pour les jours "chauds" d'une part ou les jours "froids" d'autre part, bien que les nombres de jours soient très différents en raison des conventions choisies pour définir les jours "chauds" ou "froids" et de la différence de climat entre Rennes et Sydney.
- La croissance du nombre annuel de jours "chauds" est restée modérée, en particulier à Sydney.
- La décroissance du nombre de jours "froids" est forte, surtout à Sydney où le phénomène de décroissance domine largement les fluctuations annuelles au hasard.

Tout se passe comme si le réchauffement en cours se produisait principalement par l'élévation des températures basses, plus précisément par la disparition des basses températures de nuit d'hiver. On peut conjecturer aussi que l'amplitude annuelle des températures diminue. Ces phénomènes sont-ils particuliers aux deux lieux choisis, Rennes et Sydney?

Appendice 1. Figures complémentaires.

1) La distribution croisée des nombres de jours "chauds" ou "froids" de Rennes subdivisée en 9 cases suivant les terciles:



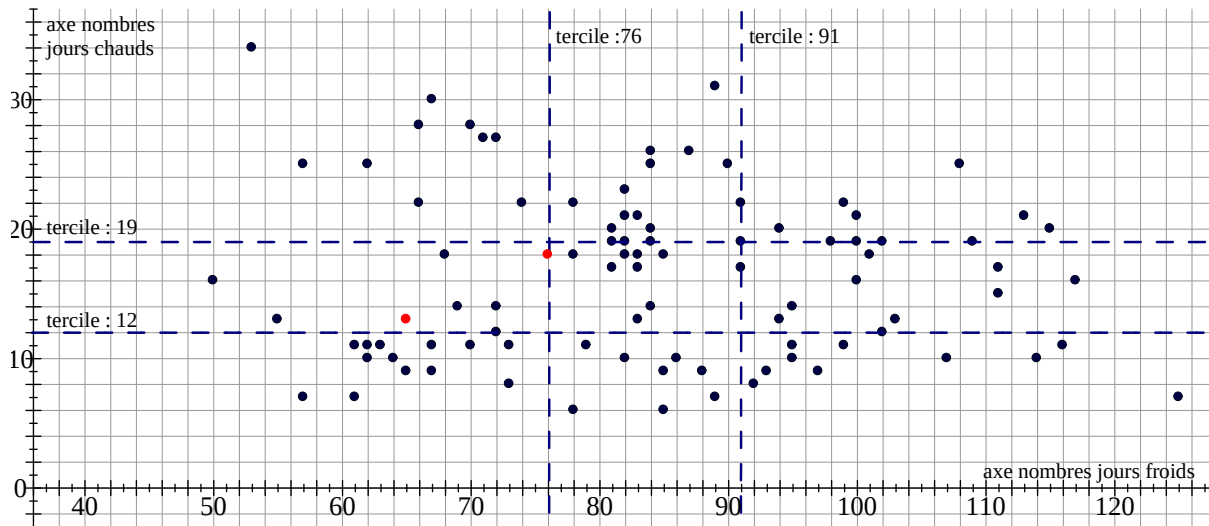
On observe la répartition (approximative) suivante:

| | | |
|------|------|------|
| 12,5 | 10 | 8 |
| 11 | 11,5 | 11,5 |
| 6,5 | 9 | 10 |

qui n'est pas

incompatible avec le hasard c'est à dire avec l'indépendance entre "chauds" et "froids" (on trouve la valeur $\chi^2 \simeq 2,9$ pour 4 degrés de liberté, très inférieure au seuil de rejet de l'hypothèse d'indépendance).

2) La distribution croisée des nombres de jours "chauds" et "froids" de Sydney subdivisée en 9 cases suivant les terciles:



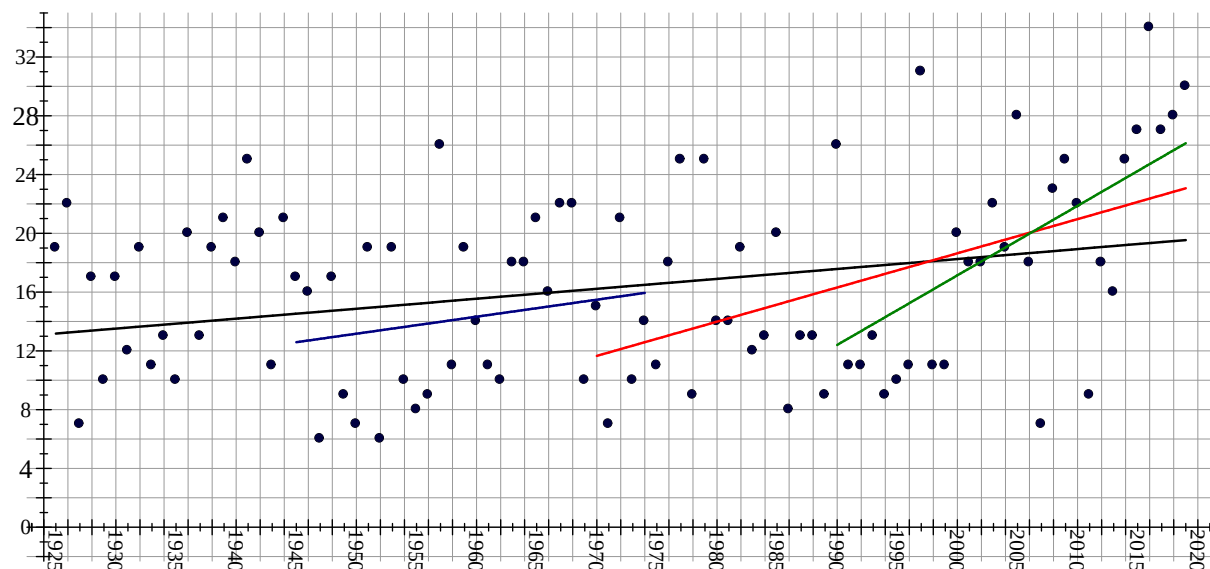
On observe la répartition (approximative) suivante:

| | | |
|------|-------|-------|
| 10 | 13,25 | 8,75 |
| 8,5 | 11,25 | 11,25 |
| 13,5 | 8 | 10,5 |

qui est loin d'être

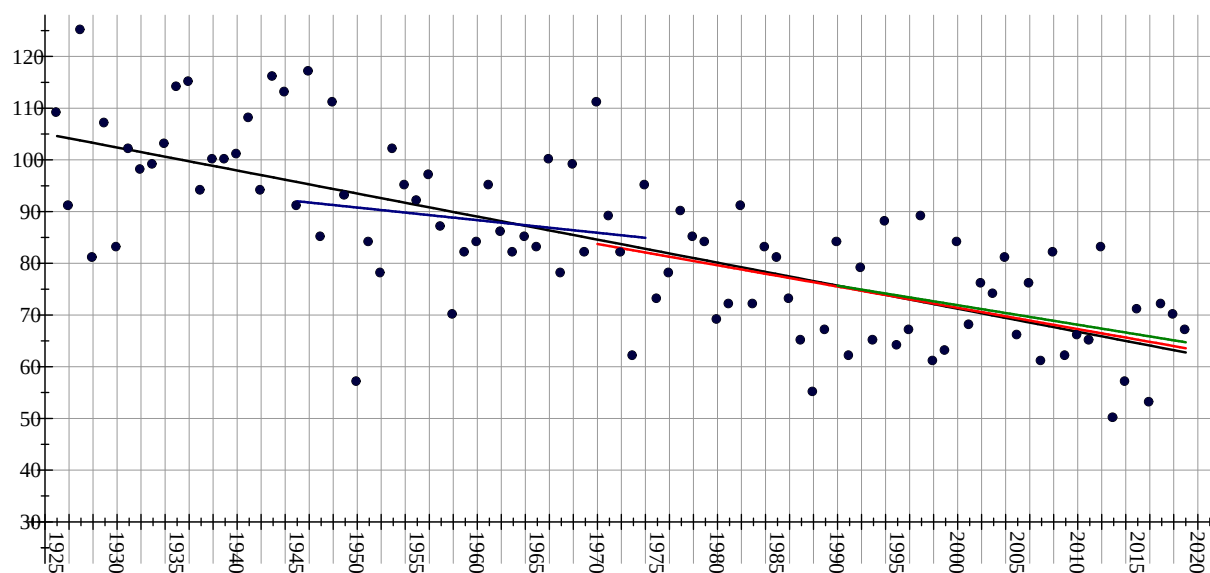
incompatible avec l'indépendance entre "chauds" et "froids" (on trouve la valeur $\chi^2 \simeq 3$ pour 4 degrés de liberté, très inférieure au seuil de rejet de l'hypothèse d'indépendance).

3) Les droites d'approximation au sens des "moindres carrés" des nombres de jours "chauds" à Sydney pour les intervalles de 30 ans 1945-1974 et 1990-2019: en bleu et en vert respectivement; les pentes sont 0,12 et 0,47 (la droite noire pour 1925-2019, et rouge pour 1970-2019, de pentes respectives 0,07 et 0,23, déjà présentées à la Fig.6):



La conclusion est différente de celle obtenue pour Rennes (Fig.3): ici on voit une accentuation régulière de la pente de croissance.

4) La figure analogue pour les jours "froids" à Sydney; on voit un processus presque constant depuis 95 ans:



Pour la période 1945-74 la droite d'approximation, tracée en bleu, a la pente $-0,24$, et pour 1990-2019, la droite tracée en vert a la pente $-0,38$ (les droites noire et rouge de pentes respectives $-0,44$ et $-0,41$ déjà présentées à la Fig. 7).

Appendice 2.

Cet appendice contient tous les détails des calculs et tests effectués. Il sera transmis à tout lecteur qui en fera la demande.

Références.

- [A] S. Aïvazian, "Etude statistique des dépendances", Ed. Mir, 1978.
- [Br] G. Bruhat, "Thermodynamique", Masson, 1942 (6ème ed. 1968).
- [Bo] A. Borovkov, "Statistique mathématique", Ed. Mir, 1987.
- [D] D. Dacuhna-Castelle, "Chemins de l'aléatoire", Champs Flammarion, 1996.
- [F] W. Feller, "An introduction to probability theory and its applications", Wiley, 1950 (3^{ème} éd. 1968).
- [G] E. Ghys, intervention orale Colloque Académie des Sciences "Face au changement climatique, le champ des possibles" accessible sur Youtube 28/1/2020.
- [JD] J. Jouzel, A. Debroise, "Le climat, jeu dangereux", Dunod, 2007.
- [JiKO] V.V. Jikov, S.M. Kozlov, O.A. Oleinik, "Homogenization of differential operators and integral functionals", Springer, 1994.
- [Ka] R. Kandel, "Le réchauffement climatique", Que sais-je?, P.U.F., 2019.
- [LJa] H. Le Treut, J.M. Jankovici, "L'effet de serre", Champs Sciences, Flammarion, 2004.
- [Ma] V. Masson-Delmotte, "Climat: le vrai et le faux", le Pommier, 2011.
- [Mo] A. Mood, Introduction to the theory of statistics, McGraw-Hill, 1950.
- [OgLiJal] M. Ogurstov, M. Lindholm, R. Jalkanen, *On the possible contribution of natural climate fluctuations to the global warming of the last 135 years*, Atmospheric and Climate Sciences, Vol. 7, pp. 256-262, 2017.
- [PoBrMa] J. Poitou, P. Braconnot, V. Masson-Delmotte, "Le climat: la terre et les hommes", edp sciences, 2014.
- [PrFo] V.E. Privalsky, M.I. Fortus, *On possible causes of global warming*, Theory Probab. Appl., Vol. 56, n°2, pp.313-317, 2012.
- [S] S. Siegel, "Non parametrics statistics", McGraw-Hill, 1956.
- [GIEC1990] Rapport GIEC 1990.
- [GIEC2007] Rapport GIEC 2007.
- [GIEC2013] Rapport GIEC 2013.
- [Site 1] www.infoclimat.fr/climatologie/rennes-st-jacques
- [Site 2] [bom.gov.au/climate/data-services/sydney\(observatory hill\)](http://bom.gov.au/climate/data-services/sydney(observatory hill))