

J. DE TIBEIRO

**Consommation d'électricité sous un climat
extrême : estimation en fonction de la date
et de la température**

Les cahiers de l'analyse des données, tome 22, n° 2 (1997),
p. 199-210

http://www.numdam.org/item?id=CAD_1997__22_2_199_0

© Les cahiers de l'analyse des données, Dunod, 1997, tous droits réservés.
L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

CONSOMMATION D'ÉLECTRICITÉ SOUS UN CLIMAT EXTRÊME: ESTIMATION EN FONCTION DE LA DATE ET DE LA TEMPÉRATURE

[CONSOMM. ÉLEC.]

J. de TIBEIRO*

1 Structure des données et examen de leurs variations

Les données dont nous disposons couvrent une période de près de 10 ans. le nombre total des enregistrements est de 3621. Le tableau initial, donne

électricité en climat extrême

4000009

num	jour	mois	an	fér	T14	T20	max1	sum1
1	1	1	1 1986	0	1694	1148	2150	40822
2	2	2	1 1986	0	1735	1430	2233	42076
3	3	3	1 1986	1	1472	1158	2062	40439
4	4	4	1 1986	2	1599	1112	1762	32013
.....
3620	3620	29	11 1995	0	1511	1054	4722	87741
3621	3621	30	11 1995	0	1560	1044	4638	87261

explicitement la date, par trois nombres {jour, mois, an}; et signale, dans la colonne: fér, le Samedi (1), le Dimanche (2) et d'éventuels jours fériés. De la date, nous ne retiendrons que le numéro d'enregistrement: num, qui servira de coordonnée temporelle (et aussi de sigle à chaque enregistrement); en revanche, la distinction des jours ouvrables (fér=0) sera conservée comme intéressant la consommation.

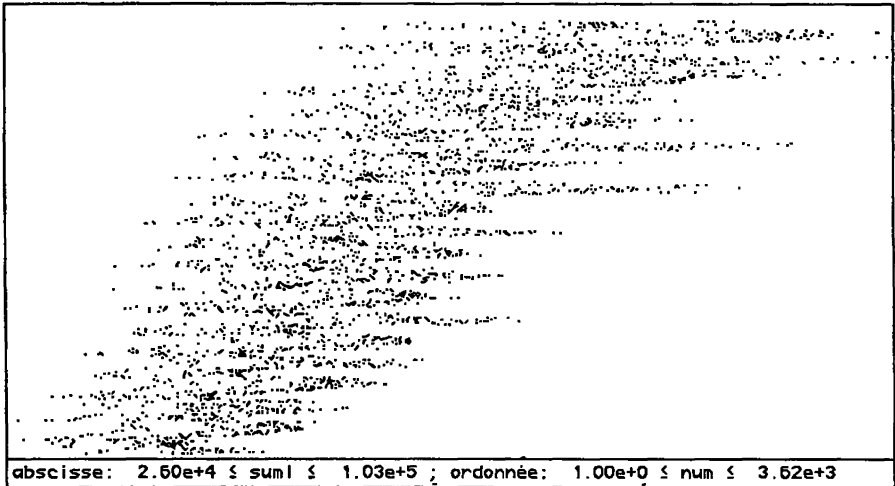
Il y a deux variables climatiques: T14, température à 14 heures; et T20, température à 20 heures. Dans une étude approfondie, destinée à être immédiatement appliquée, il conviendrait de tenir compte des deux variables. Mais ici, ne visant qu'à proposer une méthode, nous conservons seulement T14; variable qui offre l'intérêt de signaler au mieux les poussées thermiques, très importantes dans la région d'où proviennent les données.

(*) Professeur à l'Université de Moncton à Shippagan; Secteur des Sciences;

C.P. 2000 ; Shippagan, Nouveau-Brunswick; CANADA

L'auteur exprime sa gratitude au Pr. Ayala COHEN et au Dr. Lyme HYANS pour leur collaboration à la présente recherche.

La consommation est, elle aussi, caractérisée par deux nombres: $maxl$, maximum de la consommation horaire; et: $suml$, consommation totale sur 24 heures. Seule cette dernière variable fera l'objet d'une estimation; laquelle prendra en compte: num et $T14$ (température exprimée en centièmes de degrés centigrades, et toujours supérieure à zéro C.).

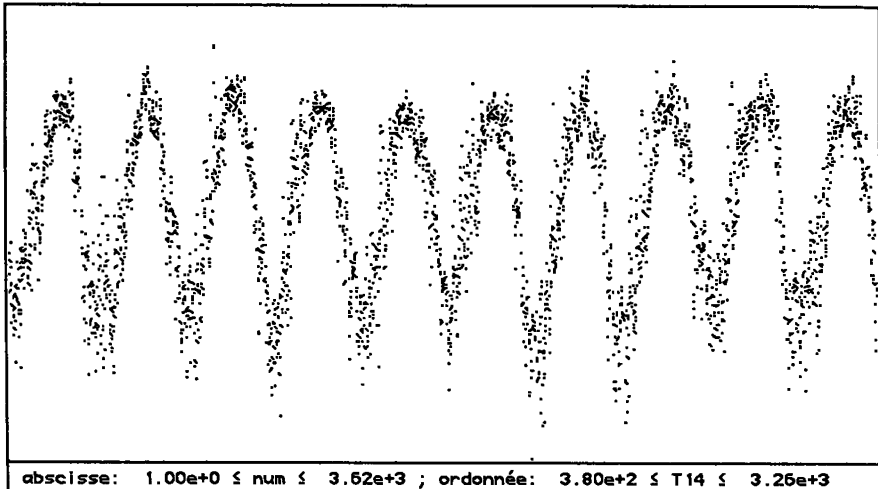


Le croisement de la coordonnée: $num \approx t$, avec la consommation totale: $suml$, montre immédiatement que celle-ci augmente régulièrement avec le temps. Mais à cette tendance, qui, à première vue, semble linéaire, se superposent de grandes variations saisonnières de la consommation. Assurément, bien plus que du rythme annuel du travail et des vacances, il s'agit ici de variations de la température.

De façon précise, le croisement de: $num \approx t$, avec: $T14$, montre que la température a, chaque année, un minimum au début de l'Hiver; et un maximum au début de l'Été. La variation d'ensemble paraît quasi sinusoïdale; s'il n'y avait des valeurs excentriques qui donnent aux extrema l'apparence d'une pointe.

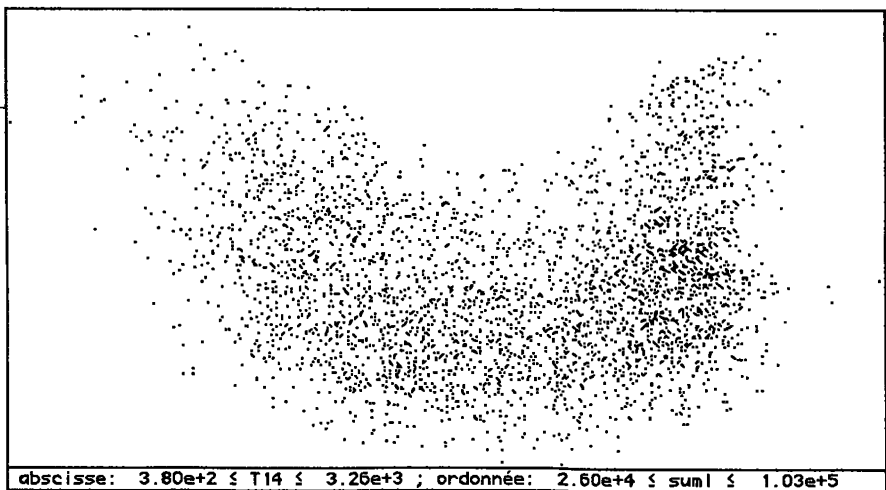
Mais en considérant attentivement le graphique, certes poudreux, de la consommation, on voit que la périodicité de celle-ci, ou, du moins, la distribution approximative de ses extrema, offre comme un harmonique de rang deux des variations de la température: l'hypothèse se présente donc que la consommation d'électricité s'accroît aussi bien en saison froide qu'en saison chaude.

Pour la chaleur, avec des maxima dépassant 30° , on admettra que la conservation des denrées et la climatisation mette à contribution l'électricité.



Pour le froid, avec des minima toujours positifs et des températures hivernales qui ne sont que rarement inférieures à 10°, la nécessité du chauffage n'apparaît pas manifeste à un résident du Nouveau Brunswick... Mais l'examen attentif du tableau des données confirme que la consommation hivernale répond, jour après jour, aux fluctuations du thermomètre; en sorte qu'il ne peut s'agir d'activité saisonnière de l'industrie, du tourisme ou de l'agriculture; mais bien d'une influence directe du froid, confirmée par nos calculs ultérieurs.

Le croisement de T14 avec la consommation suggère une sorte de



relation quadratique entre température et consommation (celle-ci étant minima pour les températures moyennes); mais, la relation est brouillée par la croissance générale de suml en fonction de $t \approx \text{num}$. (T20 s'accorde avec T14; mais, comme nous l'attendions, T14 semble avoir, avec suml, un lien un peu plus fort que ne l'a T20).

Cependant, la tendance générale à la croissance donne, à la sorte de courbe de suml en fonction de T14, une largeur qui (comptée dans la direction verticale) ne peut être inférieure à la variation de la consommation journalière moyenne de la première à la dernière année de notre période. En effet, même si, à une étape donnée du développement économique, la température déterminait rigoureusement la consommation, il resterait à ajouter à la valeur suml(T14) un terme de dérive annuelle: $\Delta \text{suml}(t)$.

```

électricité en climat extrême
il n'y a pas de colonne de pondération
corrélacion entre col      1: num et col      9: suml
corr( num, suml) = 8.000e-1
suml - 5.866e+4 ≈ 1.096e+1 * ( num - 1.811e+3)
num - 1.811e+3 ≈ 5.838e-2 * (suml - 5.866e+4)

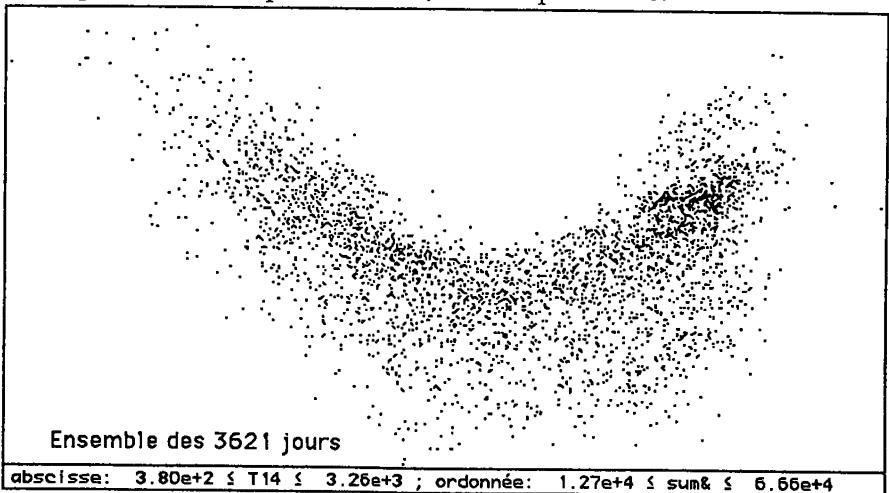
corr( T14, T20) = 9.839e-1 ; corr(maxl, suml) = 9.891e-1 ;
corr( num, maxl) = 7.837e-1

```

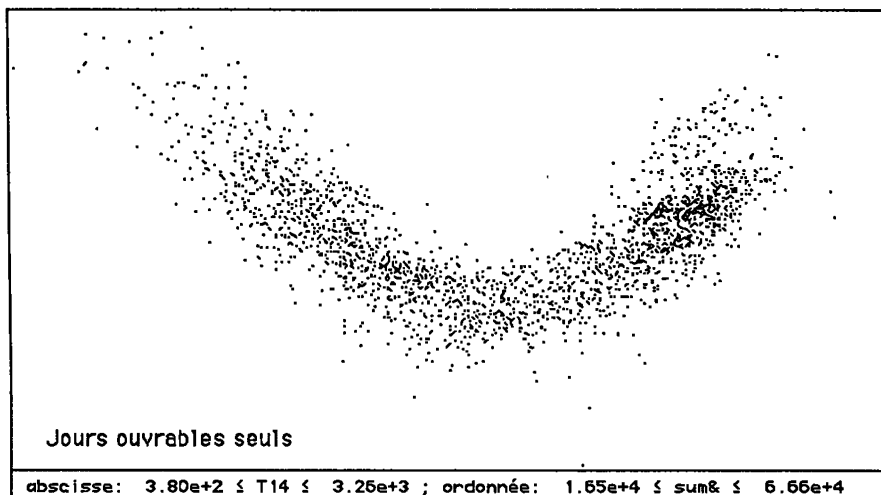
Un calcul de corrélation donne pour meilleure approximation linéaire de suml, en fonction de num, la combinaison:

$$\text{suml} \approx (10,96 \times (\text{num} - 1811)) + 58660 ;$$

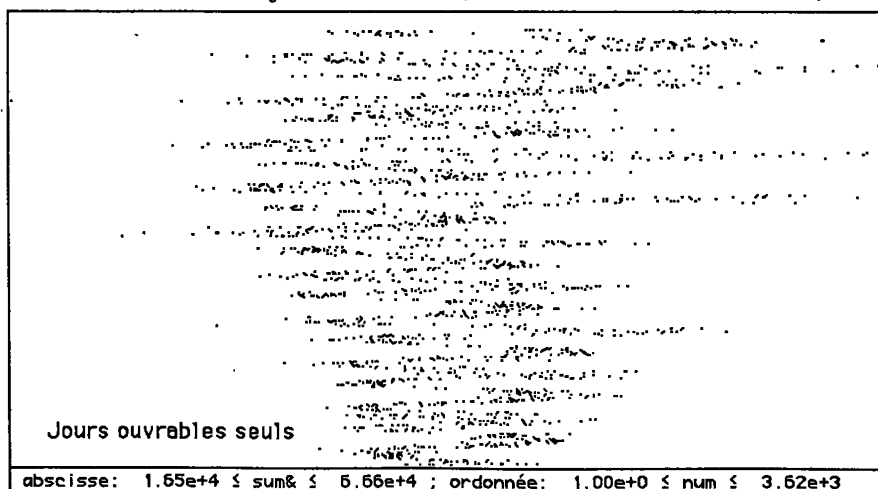
en d'autres termes, la combinaison $(\text{suml} - 11.\text{num})$ n'est pas corrélée avec $\text{num} \approx t$. On construit donc cette combinaison, notée: sum&: et la relation quadratique, avec la température: T14, est bien plus claire.



[Sans considérer toutes les corrélations, on notera, sur l'extrait de listage publié, que les couples {T14, T20} et {maxl, suml} sont étroitement liés; mais que l'approximation linéaire de la consommation horaire maxima: maxl, en fonction du temps, assimilé à num, est un peu moins bonne que celle de la consommation des 24 heures: suml.]



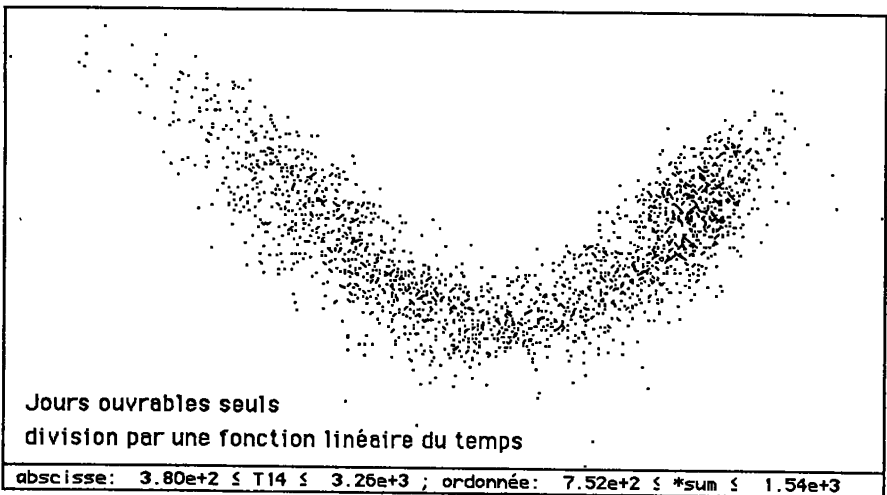
Mais les jours fériés mettent certainement du flou dans la relation. On reprend donc le croisement, mais pour le seul sous-ensemble des jours ouvrables. La relation quadratique est plus nette; car a disparu un essaim de points de faible consommation, situés en-dessous du graphe en croissant. Sur ce même ensemble des jours ouvrables, le croisement de: $\text{num} \approx t$, avec:



sum& (en bref, la consommation corrigée de la croissance linéaire à long terme), montre que la moyenne ne glisse guère sur la période étudiée; mais l'amplitude des variations liées à la température croît avec les années.

```
corr(num,suml) = 8.7724851e-1
suml - 6.2387848e+4 ≈ 1.1553333e+1 * (num - 1.8107802e+3)
suml - 62388 ≈ 11.55 * (num - 1811)
suml ≈ 41468 + (11.55 * num) ≈ 11.55 * (3590 + num)
*sum = 100 * suml / (3590 + num)
```

D'où un nouvel essai: au lieu de corriger suml en retranchant une fonction linéaire de: $\text{num} \approx t$, corriger en divisant par une fonction linéaire de num; exactement, par la meilleure approximation linéaire de suml (ou une fonction qui lui soit proportionnelle). Pour plus d'exactitude, les corrélations sont calculées ici sur l'ensemble des seuls jours ouvrables. La fonction ainsi obtenue, en éliminant la tendance à long terme de suml, est notée: *sum.



On calcule: *sum, pour les jours ouvrables. Le lien entre consommation et température, est au plus clair; et l'amplitude des fluctuations climatiques de *sum varie moins avec $t \approx \text{num}$ que ne le faisait celle trouvée pour sum&.

2 Recherche d'une formule de régression pour la consommation

Le croisement de T14 et de *sum, montre qu'on peut construire pour *sum une formule d'approximation quadratique en T14; plus précisément, une fonction linéaire du carré de la différence: ΔT , de T14 avec la valeur: 1930 (centièmes de degrés), correspondant au sommet du graphe parabolique. D'où, pour suml, une formule telle que:

$$\text{suml} \approx (3590 + \text{num}) \times (a \cdot \Delta T^2 + b) ;$$

avec deux constantes: a et b.

2.1 Analyse, après codage barycentrique, d'un rectangle de BURT croisant variable à expliquer et variables explicatives

électricité en climat extrême

Dcodx: bornes pour le découpage des variables

le nombre des variables est 3

num a 2 modalités dont les sigles et valeurs pivot sont

num< num> 1 3621

T14 a 7 modalités dont les sigles et valeurs pivot sont

Ta Tb Tc Td Te Tf Tg 883 1486 1853 2176 2456 2639 2876

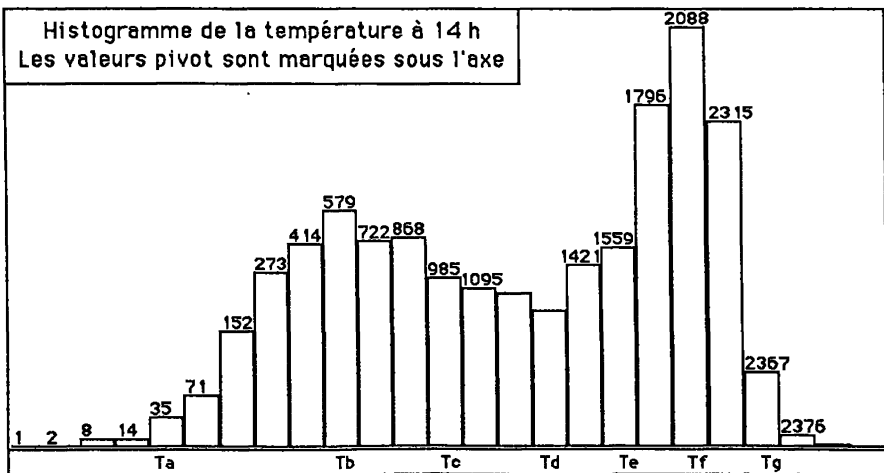
sum1 a 2 modalités dont les sigles et valeurs pivot sont

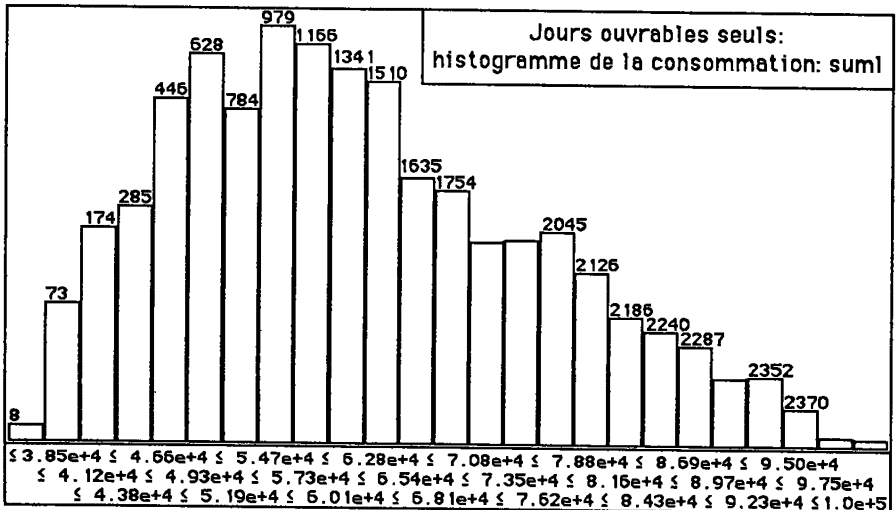
s< s> 38690 94905

Nous croyons plus instructif d'explorer ici une autre voie: soumettre les 3 variables {num, T14, sum1} à un codage barycentrique; d'où un tableau de BURT généralisé; et un (ou plusieurs) facteurs pouvant rentrer dans une formule d'approximation. En effet, l'analyse factorielle peut donner une formule simple; même dans des cas où l'examen des données ne suggère pas les termes à retenir.

Le modèle fonctionnel auquel on pense d'abord est celui des correspondances continues: en découpant les variables explicatives en un grand nombre de modalités, on aurait pour 1-er facteur la fonction, de forme générale, la mieux corrélée à la quantité à expliquer.

Mais ici, la coordonnée temporelle: num, a, sur la consommation, un double effet: d'une part, croissance à l'échelle de la décennie étudiée; d'autre part, chaque année, effets saisonniers, avec deux maxima, l'un en Hiver, l'autre en Été. Pour ce dernier effet, nous pensons que l'explication la plus directe est donnée par la température. De plus, si l'on cherchait, dans la séquence des cas, à capter cet effet saisonnier, il faudrait, pour découper num, au moins autant de pivots qu'il y a de saisons en dix ans: et rien n'assure que les calculs seraient alors protégés contre les fluctuations climatiques ou autres.



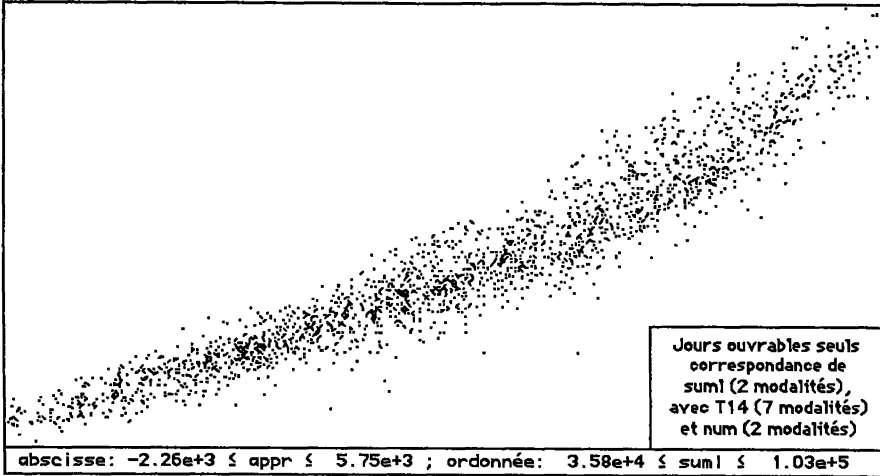


C'est pourquoi, tandis qu'afin de suivre dans ses détails la distribution bimodale de la température: T14, cette variable a été découpée avec 7 pivots, on n'a pris que deux modalités extrêmes pour num et suml. Ceci implique que, dans l'analyse, {num, suml} n'interviendront que par des termes linéaires. Mais alors, on sort du modèle des correspondances continues. Et le succès de la régression est affaire d'empirisme. Cependant, la formule obtenue dans le présent §2.1, sera retrouvée, au §2.3, dans l'analyse du rectangle de BURT croisant la variable à expliquer avec les seules modalités de T14; calcul qu'il semble légitime d'étendre à d'autres cas où il y aurait plusieurs variables explicatives minutieusement découpées.

Avec deux modalités seulement pour la variable à expliquer: suml, le rectangle de BURT, (2+7)×2, ne fournit qu'un seul facteur non trivial. Ce facteur (noté: axe1, dans nos formules) peut être calculé pour les jours, adjoints en colonnes supplémentaires à l'analyse du rectangle.

$$\begin{aligned} \text{corr}(\text{suml}, \text{axe1}) &= -8.7852487e-1 \\ \text{suml} - 6.2387848e+4 &\approx -2.3845654e+4 * (\text{axe1}) \\ \text{corr}(\text{suml}, \text{num}) &= 8.7724851e-1 \\ \text{suml} - 6.2387848e+4 &\approx 1.1553333e+1 * (\text{num} - 1.8107802e+3) \\ \text{corr}(\text{num}, \text{axe1}) &= -6.8275647e-1 \\ \text{axe1} - 2.0203272e-5 &\approx -3.3128005e-4 * (\text{num} - 1.8107802e+3) \\ \text{num} - 1.8107802e+3 &\approx -1.4071369e+3 * (\text{axe1} - 2.0203272e-5) \end{aligned}$$

La corrélation entre suml et axe1 n'est pas satisfaisante. Mais on remarque que (-axe1) et num ont, avec suml, des corrélations quasi rigoureusement égales à: 0,878. Suivant un raisonnement qui sera donné, au §2.3, sous sa forme la plus générale, plaçons-nous dans l'espace des fonctions sur l'ensemble I des jours (ouvrables). Dans le plan engendré par les formules de régression à une variable trouvées, pour suml, en fonction de num et axe1, la

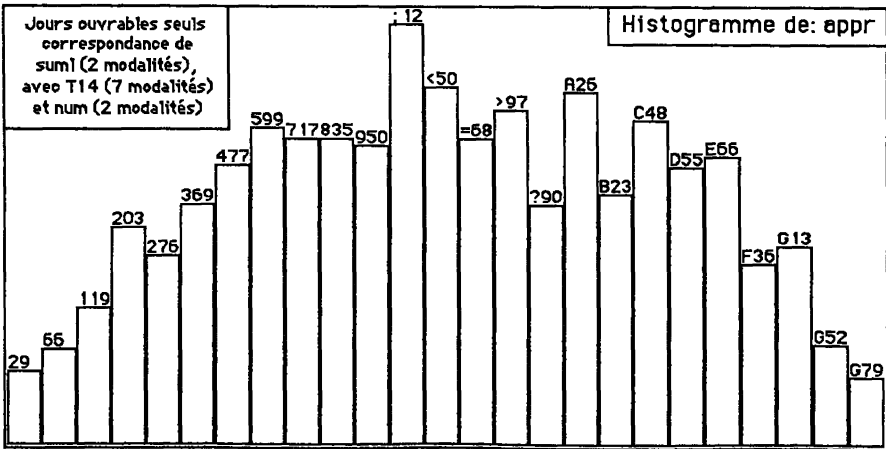


somme de ces formules donne la meilleure approximation de suml. Plutôt que cette somme, on considère la fonction: appr, qui lui est liée linéairement.

```

appr = num - (2064*axe1)
corr(suml,appr) = 9.5706892e-1
suml - 6.2387848e+4 ≈ 6.8656631e+0 * (appr - 1.8107384e+3)
suml ≈ 62388 + (6,865 * (appr - 1811)) = 6,865 * (appr + 7277)
    
```

La corrélation ici obtenue: .957, est satisfaisante; mais dans le nuage, quasi linéaire, du croisement de: suml, avec: appr, on note, pour les plus fortes valeurs, une pente supérieure à celle du reste du graphe. Ce qui offrirait matière à de nouveaux calculs, destinés à perfectionner, s'il était besoin, notre formule. D'ailleurs, l'histogramme de: suml, présente, du côté des plus fortes valeurs, une pointe qu'on ne trouve pas pour: appr.



2.2 Estimation de la consommation des Dimanches d'après la formule trouvée pour les jours ouvrables

À l'analyse du rectangle de BURT, calculé pour les jours ouvrables, les Dimanches peuvent être adjoints en colonne supplémentaire, comme l'ont été les jours ouvrables. Ainsi, On reprend sur l'ensemble: ID, des Dimanches, les calculs du §2.1.

```
consommation en électricité pour les Dimanches
corr(suml,axel) = -8.3257478e-1
suml - 4.7706887e+4 ≈ -1.8555626e+4 * (axel - 4.1530091e-2)
corr(suml,num) = 8.5818847e-1
suml - 4.7706887e+4 ≈ 9.3355026e+0 * (num - 1.8048296e+3)
corr(num,axel) = -6.9346329e-1
```

Une fonction: appr, adaptée aux corrélations trouvées ici, n'aurait pas les mêmes coefficients que celle adoptée pour les jours ouvrables... Mais l'échantillon des Dimanches étant restreint, on conserve exactement la fonction: appr, déjà définie comme combinaison de: num, et de: axel1.

```
consommation en électricité pour les Dimanches
corr(suml,appr) = 9.1866042e-1
appr - 1.7191115e+3 ≈ 1.5599569e-1 * (suml - 4.7706887e+4)
suml - 4.7706887e+4 ≈ 5.4100020e+0 * (appr - 1.7191115e+3)
suml ≈ 5,41 * (appr + (8818-1719))
      ≈ 5,41 * (appr + 7099) = (5,41 * (appr + 7277)) - 963
```

Pour l'ensemble: ID, des Dimanches, la corrélation de: appr, avec: suml, est quelque peu inférieure à celle trouvée sur l'ensemble I, des jours ouvrables. De plus, le terme constant n'est pas exactement le même: 7099 au lieu de 7277; soit, pour la consommation, un décalage de l'ordre de 1000. Mais il reste utile de conserver la formule des jours ouvrables simplement multipliée par un coefficient de réduction:

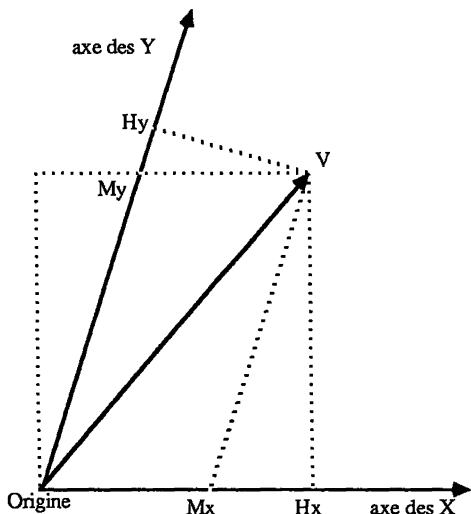
$$\text{red} = 5,41 / 6,865 = 0,788 .$$

2.3 Estimation de la consommation des jours ouvrables d'après l'analyse du tableau croisant les modalités de T14 et de suml

On procède comme au §2.; à ceci près que le rectangle de BURT considéré n'a pour dimension que 7×2. Mais dans le cas présent, on ne retrouve pas l'égalité, fortuite, de la corrélation de: suml, avec d'une part: num, et, d'autre part, le facteur (noté maintenant: axx1).

```
corr(suml, num) = .87725
suml - 62388 ≈ 11.5533 * (num - 1811)
corr(suml, axx1) = .42289
suml - 62388 ≈ 6577 * axx1
corr(num, axx1) = .046355
```

Il faut donc calculer, par une formule générale, les coefficients de la régression linéaire de: suml, en fonction des variables explicatives: {num, axx1}. C'est, au fond, le problème général du passage, dans le plan, des coordonnées rectangulaires aux coordonnées obliques.



Notons C et S le cos et le sin de l'angle (OX, OY) formé par les axes. Soit: OHx, OHy les projections d'un vecteur OV sur les axes. Et soit: OMx, OMy les projections obliques, parallèlement aux axes. Notons : (x, y) les abscisses de Hx et Hy (sur leurs axes respectifs); et, de même: (x', y') les abscisses de Mx et My. Alors, ces coordonnées sont liées par les formules:

$$\begin{aligned}
 x' &= (1 + (C^2/S^2)) \cdot x && - (C/S^2) \cdot y \\
 y' &= && - (C/S^2) \cdot x + (1 + (C^2/S^2)) \cdot y
 \end{aligned}$$

formules qu'il suffit de vérifier sur le cas particulier: x=1, y=0 ; et de compléter par symétrie et linéarité.

Le listage de corrélation et régression dont nous disposons, donne, d'une part, les corrélations, x et y, de la variable à estimer, suml, avec les deux variables explicatives, num et axx1; et, d'autre part, la formule d'approximation linéaire de suml par une seule variable, num ou axx1. On peut calculer la formule d'approximation linéaire à deux variables comme si les variables {suml, num} étaient de moyenne nulle, comme l'est véritablement axx1: il restera seulement à ajouter une constante à la formule. Il faut d'abord passer de (x, y) à (x', y'), par la formule déjà rappelée. Ensuite, il conviendrait de diviser chacune des variables explicatives par son écart type, sig, afin de lui donner variance 1. Ainsi on a:

$$\text{suml} \approx \text{constante} + (x' / \text{sig}(\text{num})) \cdot \text{num} + (y' / \text{sig}(\text{axx1})) \cdot \text{axx1} ;$$

en fait, les quotients (x / sig(num)) et (y / sig(axx1)) nous sont déjà connus, car ils figurent comme coefficients dans les formules d'approximation linéaire à une variable. Il nous suffit donc de multiplier ces coefficients par (x'/x) et (y'/y) respectivement. Plus précisément, on a construit, pour la meilleure approximation de suml, une combinaison: αππr, où num figure avec le coefficient 1. La corrélation finale trouvée est, comme il se doit, la même qu'au §2.1: soit 957‰; la corrélation entre appr et αππr est: 999.999‰

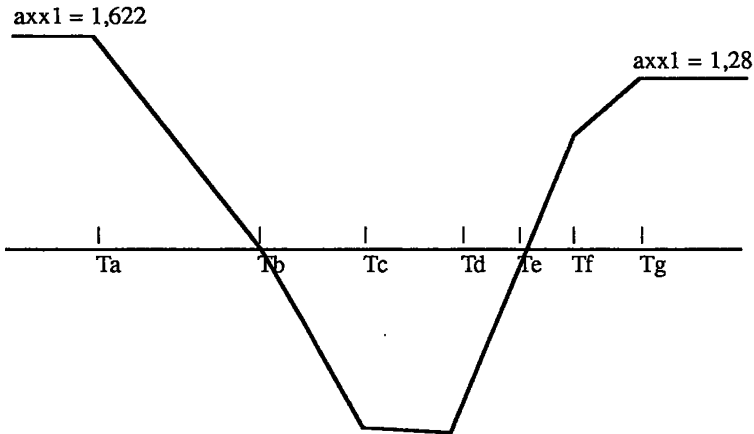
Avec les notations posées dans le rappel de la formule, on a:

$$\begin{aligned} x &= .87725 & ; & & y &= .42289 & ; & & c &= .046355 & ; \\ x' &= .86 & ; & & y' &= .383 & ; \\ \text{coeff de num} &= 11.5533 * (0,86 / 0,877) & = & 11.33 \\ \text{coeff de axx1} &= 6577 * (0,383 / 0,4229) & = & 5956.47 \end{aligned}$$

et, comme au §2.1, l'approximation utilise une combinaison linéaire: $a\pi\pi_r$, où: num, a pour coefficient: 1.

$$\begin{aligned} a\pi\pi_r &= \text{num} + (5956/11.33) \cdot \text{axx1} \approx \text{num} + 525 \cdot \text{axx1} \\ \text{corr}(\text{suml}, a\pi\pi_r) &= .957 \\ \text{suml} - 62388 &\approx 11.32 * (a\pi\pi_r - 1811) \end{aligned}$$

D'ailleurs, il résulte de la formule de transition que le facteur: axx1 , est, pour tout individu, principal ou supplémentaire, une fonction linéaire par morceaux de la température: T_{14} ; les points anguleux ayant pour abscisses les valeurs pivot.



3 Conclusion: l'art de la régression

Pour le problème étudié dans le présent article, une régression linéaire ne pouvait suffire; et il en est ainsi dans de nombreux cas. En principe, la régression polynomiale a des formules générales universelles. Mais si la taille de l'échantillon disponible n'offre pas de base stable au calcul de très nombreux coefficients, cette généralité est illusoire. Il faut alors prendre en compte la structure particulière des données considérées.

Référence bibliographique

A. El OUADRANI : "Généralisation du tableau de BURT et de l'analyse de ses sous-tableaux dans le cas d'un codage barycentrique", [BURT COD. BARY.]; in CAD, Vol.XIX, n°2, pp. 229-246; (1994);