

CAHIER DU LAMSADE

Laboratoire d'Analyse et Maîtrise des Systèmes pour l'Aide à la Décision
(Université Paris IX Dauphine)

Présentation synthétique de modèles de prévision à
très court terme de l'énergie journalière produite
par Electricité de France et de la température
moyenne journalière relevée à Paris-Montsouris

N°10-1977

J. ABADIE et F. MESLIER
octobre 1977



78 L 12 04

(544564)

PRESENTATION SYNTHETIQUE DE MODELES DE PREVISION A TRES COURT TERME
DE L'ENERGIE JOURNALIERE PRODUITE PAR ELECTRICITE DE FRANCE ET DE
LA TEMPERATURE MOYENNE JOURNALIERE RELEVÉE A PARIS-MONTSOURIS

J. ABADIE et F. MESLIER

Résumé

La qualité des prévisions à très court terme est une donnée fondamentale de la bonne tenue d'une entreprise productrice d'Electricité, en particulier en raison de capacités limitées de stockage de l'énergie électrique. Toute amélioration amènera à la fois une diminution des coûts pour le producteur et une augmentation de la qualité de service pour le consommateur.

L'expérience montre clairement que les méthodes utilisées actuellement par Electricité de France sont très satisfaisantes de ce point de vue. Cependant, pendant les trois dernières années, divers travaux ont été entrepris pour étudier les possibilités d'autres approches.

Les résultats que nous proposons ici s'inscrivent dans cette ligne générale, en insistant tout particulièrement sur la première utilisation élaborée des techniques développées par Box et Jenkins (Auto Regressive Moving Average Processes ou ARIMA).

ABSTRACT

The quality of short term load forecasts is essential for a real time control center and any greater accuracy will tend to decrease the costs for the producer and increase the quality of service for the consumer.

Experience shows clearly that present methods used by Electricité de France are very satisfactory in this respect. However, during the last three years, research has been undertaken to investigate the suitability of other approaches.

We summarize here the main results and conclusions of our first elaborate use of the techniques developed by Box and Jenkins (Auto Regressive Integrated Moving Average Processes or ARIMA).

TABLE DES MATIERES

	Page
1.- INTRODUCTION	4
2.- SERIE DES TEMPERATURES	7
2.1- Introduction	7
2.2- Mise au point du modèle ARIMA	7
2.3- Résumé des autres modélisations envisagées	9
2.3.1- Existe-t-il un réchauffement des températures relevées à Paris-Montsouris	9
2.3.2- Modélisation tenant compte des températures moyennes, journalières	9
2.3.3- Evaluation du motif de la périodicité par les moindres carrés ..	11
2.3.4- Extension du modèle précédent en tenant compte de tendances de températures jour par jour	11
2.3.5- Modélisation en termes d'écarts à la température normale ...	12
2.3.6- Utilisation de l'analyse harmonique	13
2.4- Simulation d'une utilisation en prévision du modèle ARIMA	14
2.5- Bilan de l'étude des températures	14
3.- SERIE DES ENERGIES JOURNALIERES	15
3.1- Introduction	15
3.2- Recherche d'un ou plusieurs modèles ARIMA possibles	15
3.3- Inventaire des jours de fête	17
3.4- Correction des fêtes mobiles	18
3.5- Correction des fêtes à date fixe	19
3.6- Prise en compte de la cassure du mois d'Août	20
3.7- Choix d'un modèle	21
3.8- Simulation de l'utilisation en temps réel du modèle ARIMA (numéro 1) ..	22
3.9- Bilan de l'étude des énergies journalières	24
4.- CONCLUSION ET PERSPECTIVES	25
BIBLIOGRAPHIE	28
ANNEXE I - Estimation des propriétés du premier et du deuxième ordre d'une série chronologique stationnaire	30
ANNEXE II - Estimation des modèles ARMA	31
ANNEXE III - Premier essai d'identification d'une fonction de transfert liant l'énergie journalière à la température journalière ..	34
ANNEXE IV - Exemples de coefficients correcteurs obtenus à l'aide du modèle (2) final	36

1.- INTRODUCTION

1.1- Nous proposons ici un résumé d'une analyse poussée de deux séries chronologiques qui revêtent une importance toute particulière pour Electricité de France, aussi bien pour des cycles de décision à très court terme que pour des études de gestion prévisionnelle et de planification. Le travail que nous avons effectué ([1] à [6]) s'est situé d'emblée dans la perspective de la mise au point de modèles de prévision à très court terme. Nous complétons nos précédentes publications en proposant ici les résultats d'utilisation de nos modèles dans des conditions réelles de prévision.

1.2- La première série est relative à la demande d'énergie électrique en France.

1.2.1- La figure 1 représente une forme possible de la demande horaire, sur deux semaines (du mardi matin au lundi soir) : cette courbe s'intitule dans le jargon des électriciens une courbe de charge. L'analyse de ce schéma de principe permet de mettre en évidence quelques caractéristiques simples de la demande à EDF :

- les niveaux et les formes des lundi, mardi, mercredi, jeudi et vendredi sont extrêmement semblables. Cependant, le lundi se distingue en général par un niveau moyen plus faible que celui des autres "jours ouvrables" (mardi, ..., vendredi)
- les samedis semblent bien différents, à la fois en forme et en niveau général. Il en est de même pour les dimanches.

1.2.2- Les méthodes actuelles de prévision sont basées sur l'hypothèse de la stabilité de telles caractéristiques au cours du temps. Ceci conduit à effectuer les prévisions à l'aide de divers facteurs de modulation. La procédure, résumée sur le tableau 1, comporte deux étapes clefs :

- prévision de l'énergie journalière à température normale *
- prévision d'une courbe de charge horaire et introduction d'une prévision de température journalière

../..

La température normale d'un jour t dans l'année se définit comme la moyenne statistique des réalisations des températures de ce jour sur un grand nombre d'années.

1.2.3- Comment peut-on situer notre contribution par rapport aux méthodes actuellement utilisées ?

- Nous nous sommes limités a priori à l'analyse des seules énergies journalières : la complexité de ce problème apparaîtra très nettement dans le texte. Des travaux complémentaires, sans doute très importants, permettraient d'aborder la prévision de la courbe de charge dans le même esprit. Il est, en l'état actuel des résultats obtenus, tout à fait envisageable d'utiliser, à l'aval des modèles obtenus pour les énergies journalières, des coefficients π (voir tableau I) pour définir des courbes de charge prévisionnelles.
- Les réalisations des énergies journalières constituent une série chronologique et c'est dans ce seul esprit que nous les avons envisagées
- Les différents types de jours envisagés de manière systématique dans les méthodes actuelles de prévision à court terme sont les lundi, mardi, ..., samedi, dimanche "normaux". Or il existe de très nombreuses réalisations particulières, correspondant principalement aux jours de fête, et nous en avons tenu compte de manière explicite dans notre modélisation.
- Les modèles envisagés pour expliciter le comportement de la série n'ont pu être ajustés qu'en faisant appel à des méthodes d'optimisation non linéaire assez générales, en raison surtout du nombre très important de réalisations prises en compte et du nombre élevé de paramètres définissant les modèles. L'annexe II donne les idées principales de cet aspect.

1.3- La seconde série abordée est celle des températures moyennes journalières de Paris-Montsouris. Les principales raisons de ce choix sont au nombre de trois :

- cette série est sans doute plus facile à analyser que celle des énergies journalières. Elle nous a en quelque sorte servi de banc d'essai de différents types de modélisation possibles
- cette série est représentative de l'influence de la température sur la consommation d'énergie électrique en France. Elle conviendrait donc à la mise au point d'une fonction de transfert liant l'énergie journalière à la température journalière. L'annexe III donne les idées de développement de nos présents travaux dans cette direction.
- les réalisations de cette série sont utilisées dans les procédures actuelles de prévision de la courbe de charge (voir tableau I), la prévision à un jour étant la réalisation de la veille. Il semble utile de voir si un meilleur modèle est possible.

1.4- Nous avons donc envisagé en premier lieu la série des températures journalières de Paris-Montsouris, de manière à aborder plusieurs types de modélisations possibles, de façon à être mieux armé pour la seconde série étudiée, celle des énergies journalières.

Nous nous sommes limités a priori à la classe des seuls modèles endogènes et ce pour diverses raisons :

- nous pensons qu'avant de tenter d'expliquer le comportement d'une série par celui d'autres dont elle est réputée dépendre, il importe de retirer au préalable toute la substance que la série apporte en elle-même : autrement dit, la mise au point d'un modèle endogène nous semble de toute manière être la première étape indispensable. La recherche d'éventuels éléments explicatifs extérieurs à la série elle-même ne doivent donc porter que sur ce que le modèle endogène ne permet pas de prendre en compte par lui-même
- compléter nos modèles endogènes par des modèles exogènes (mise au point de fonctions de transfert) ne constituerait donc qu'une extension de nos travaux. A cet égard, nous voudrions distinguer nos deux séries :

α) en ce qui concerne la série des températures, il est probable qu'adjoindre au modèle endogène retenu la prise en compte des prévisions d'évolution des champs de pression atmosphérique au-dessus de l'Europe Occidentale améliorerait nos résultats en permettant essentiellement de tenir compte des phénomènes de rupture (perturbations) évidemment non pris en compte par le modèle endogène.

β) en ce qui concerne la série des énergies journalières, la prise en compte des effets climatiques améliorerait certainement nos résultats (voir par exemple l'annexe III pour la température). Par contre, nous ne pensons pas que de grandeurs macroéconomiques aient une influence significative pour la prévision à très court terme, l'inertie étant beaucoup plus grande que pour les effets climatiques. Néanmoins, l'évolution du contexte économique peut provoquer des dérives plus ou moins lentes du modèle (modification de la structure de la consommation entre usages industriels et domestiques par exemple) ; à cet égard, un suivi systématique des qualités prévisionnelles du modèle permet de procéder, si nécessaire à une réactualisation des paramètres du modèle, voire du modèle lui-même.

1.5- Parmi les différents types de modélisations envisagés, les résultats obtenus nous ont amenés à utiliser surtout les processus ARIMA (Auto Regressive Moving Average). Pour une étude approfondie de ces processus, il conviendra de se reporter à l'ouvrage de référence [7]. Nous donnons en Annexe I les définitions des diverses estimations proposées dans le corps du texte.

1.6 Ce document est articulé en trois parties :

- analyse de la série des températures moyennes journalières de Paris-Montsouris
- analyse de la série des énergies journalières produites par EDF
- conclusions et perspectives.

2.- SERIE DES TEMPERATURES

2.1- Introduction

Nous disposons de la température moyenne journalière relevée à Paris-Montsouris de 1920 à 1974 inclus. Nous avons utilisé les 53 premières années, soient 19359 observations, pour estimer différents modèles. Les deux dernières années ont servi à réaliser une simulation de la prévision en temps réel.

Une première analyse, tant de la série observée que des estimations de ses principales propriétés montre, bien entendu, que le processus sous-jacent n'est pas stationnaire et présente une forte périodicité annuelle.

2.2- Mise au point du modèle ARIMA

2.2.1- L'existence de la périodicité annuelle conduit à filtrer la série des observations x_t par l'opérateur ∇_{365} (défini par $\nabla_{365} x_t = x_t - x_{t-365}$). La série ainsi transformée ne présente plus de non stationnarité évidente.

Nous pouvons donc tenter l'ajustement d'un modèle ARMA. L'analyse de l'autocorrélation (figure 2) et de l'autocorrélation partielle (figure 3) conduit à postuler le modèle :

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3) w_t = (1 - \theta_{365} B^{365}) a_t$$

$$\text{où } w_t = \nabla_{365} x_t$$

où a_t représente l'erreur de prévision à l'instant t et où B est l'opérateur défini par : $Bx_t = x_{t-1}$.

2.2.2- Le critère d'ajustement des quatre paramètres du modèle ARMA postulé a été

$$\begin{array}{ccc} \text{Min} & 19359 & a_t^2 \\ \phi_1, \phi_2, \phi_3, \theta_{365} & \Sigma & t=3654 \end{array}$$

ce qui conduit, pour la série observée, à :

$$\begin{aligned} \phi_1 &= 0,99121 \\ \phi_2 &= -0,28602 \\ \phi_3 &= 0,08418 \\ \theta_{365} &= 0,90179 \end{aligned}$$

La variance du résidu est estimée à 4,0890 soit un écart-type de prévision à un jour de l'ordre de 2 degrés.

2.2.3- L'analyse d'une première catégorie d'estimations permet de retenir pour le résidu a_t du précédent modèle une hypothèse de bruit blanc :

- autocorrélation : 3,6 % d'estimations sont en dehors de la bande de confiance à 5 % du bruit blanc normal
- autocorrélation partielle : 4 % d'estimations sont en dehors de la bande de confiance à 5 % du bruit blanc normal
- spectre : 1,6 % d'estimations sont en dehors de la bande de confiance à 5 % du bruit blanc normal.

2.2.4- L'analyse d'estimations plus fines (analyse dite modulo 365) montre en fait que le résidu n'est pas stationnaire : l'écart-type du résidu de la prévision à un jour présente en effet un aspect saisonnier assez net (variance forte en hiver et faible en été). A moins d'afficher un écart-type pour chaque jour de l'année, le modèle ARIMA n'est donc pas suffisant (d'autant que les écarts-types présentent de violentes fluctuations d'échantillonnage d'un jour sur l'autre). Nous avons esquissé la manière correcte de résoudre le problème ainsi posé en étudiant les conséquences, sur la série initiale d'une transformation non linéaire [8] ainsi définie :

$$z_t(\lambda_1, \lambda_2) \begin{cases} = \frac{(x_t + \lambda_2)^{\lambda_1 - 1}}{\lambda_1} & \text{si } \lambda_1 \neq 0 \\ = \log(x_t + \lambda_2) & \text{si } \lambda_1 = 0 \end{cases}$$

Rechercher les meilleurs λ_1, λ_2 c'est rechercher le maximum d'une fonction de vraisemblance. Quelques essais numériques simples nous ont montré qu'un couple de valeurs de λ_1 et λ_2 améliore sensiblement les qualités de stationnarité et de normalité du résidu a_t du modèle ARIMA, estimé cette fois-ci sur le z_t . Il resterait cependant à trouver le meilleur couple des λ pour savoir si cette transformation est convenable ou s'il convient d'en utiliser d'autres.

2.3- Résumé des autres modélisations envisagées

2.3.1- Existe-t-il un réchauffement des températures relevées à Paris Montsouris

Ce phénomène, connu, semble a priori peu sensible au niveau de la prévision de la température d'un jour sur l'autre. Nous avons cependant recherché si la série des observations présentait des non stationnarités de cet ordre.

Les principales conclusions de cette étude sont :

- l'existence d'un réchauffement significatif de 0,7 à 0,8 degré par siècle sur la température moyenne annuelle (ce réchauffement établi sur 53 ans se retrouve d'ailleurs sur des chroniques plus longues [9])
- ce réchauffement ne se manifeste pas uniformément dans l'année ainsi que l'a traduit une analyse des températures moyennes mensuelles.

Tenir compte de ces phénomènes sur la série de nos observations conduirait donc à introduire des tendances de température variables avec le jour.

2.3.2- Modélisation tenant compte des températures moyennes journalières

2.3.2.1- La température moyenne journalière d'un 1er Janvier peut se concevoir comme la moyenne des 53 réalisations des 1er Janvier 1920 à 1972. En fait, ce n'est pas tout à fait aussi simple que cela, en raison de la présence de 29 Février tous les 4 ans. Soit n le nombre d'années ; nous introduisons 3 variantes de calcul des moyennes journalières $\bar{x}_t$:

$$\text{variante 1 : } \overline{x}_t = \frac{1}{n} \sum_{i=1}^n x_{t+365(i-1)} \quad t = 1, 2, \dots, 365$$

variante 2 : comme ci-dessus, mais en ayant omis tous les 29 Février

$$\text{variante 3 : } \overline{x}_t = \frac{1}{m} \sum_{i=1}^m x_{t+1461(i-1)} \quad t = 1, \dots, 1461 \text{ et } m=1$$

Nous désignerons par série 1 la série avec les 29 Février et par série 2 la série sans les 29 Février.

2.3.2.2- Nous avons alors comparé les 3 variantes sur le modèle $x_t = \overline{x}_t + a_t$. Dans les trois cas, le résidu présente un aspect autorégressif d'ordre 3. Par contre, des différences importantes se manifestent au niveau de l'estimation du spectre du résidu :

- existence d'un maximum local très marqué pour la période 365 pour la variante 1
- pas de maximum de ce type pour les variantes 2 et 3.

Le modèle étudié ne possède donc pas de bonnes propriétés si les 29 Février sont présents dans la série et si l'on n'en tient pas compte de manière explicite.

2.3.2.2- En complétant le modèle par des termes autorégressifs, nous définissons alors trois variantes du même modèle, de la forme :

$$x_t = \lambda \overline{x}_t + \sum_{i=1}^3 \psi_i x_{t-i} + a_t$$

L'analyse spectrale fournit encore des renseignements intéressants

variante 1 : maximum très marqué pour la période $T = 365$

variantes 2 et 3 : légers maximum pour la période $T = 365$.

Nous retrouvons la conclusion du paragraphe précédent en ce qui concerne le problème posé par les 29 Février. Par contre, le résultat obtenu avec les variantes 2 et 3 pose une nouvelle difficulté, que nous levons dans le paragraphe suivant.

2.3.3- Evaluation du motif de la périodicité par les moindres carrés

2.3.3.1- Nous avons ajusté, en tenant compte de la structure particulière de la matrice du système linéaire correspondant, le modèle :

$$x_t = \lambda_{t \bmod 365} + \sum_{i=1}^3 \psi_i x_{t-i} + a_t$$

sur la série 1 et la série 2. Ce modèle peut se considérer comme une généralisation de celui du paragraphe 2.3.2 en ce sens que les $\lambda_{t \bmod 365}$ sont des approximations des $\lambda_{t \bmod 365}$.

2.3.3.2- L'analyse spectrale du résidu des deux variantes indique :

- série 1 : présence d'un maximum local pour $t = 365$
- série 2 : pas de maximum de ce type.

Les différentes conclusions se mettent maintenant bien en place :

- conserver les 29 Février sans en tenir compte explicitement conduit, pour cette classe de modèles, à un filtrage insuffisant de la périodicité annuelle
- les $\lambda_{t \bmod 365}$ sont une approximation des $\lambda_{t \bmod 365}$ et l'estimation du spectre du résidu permet d'en détecter les conséquences quant au filtrage de la période annuelle.

2.3.3.3- L'inconvénient de cette modélisation est de fournir 365 valeurs de $\lambda_{t \bmod 365}$ qui présentent de très fortes fluctuations d'échantillonnage, d'un jour sur l'autre ; bien que les résultats au niveau du spectre soient convenables avec la série 2, nous ne pouvons considérer que le motif de la périodicité est bien déterminé par cette approche : il se pose à leur égard un problème de lissage. Nous y reviendrons à l'occasion de l'analyse harmonique.

2.3.4- Extension du modèle précédent en tenant compte de tendances de température jour par jour

2.3.4.1- Ce modèle comporte 733 coefficients et s'écrit :

$$x_t = \lambda_{t \bmod 365} + \mu_{t \bmod 365} E\left(\frac{t}{365}\right) + \sum_{i=1}^3 \psi_i x_{t-i} + a_t$$

où E représente l'opérateur "partie entière de".

L'utilisation de méthodes numériques adaptées [10] a permis de résoudre très simplement le système linéaire auquel conduit ce modèle.

2.3.4.2- Les $\lambda_{tmod365}$ sont toujours très touchés par les fluctuations d'échantillonnage. Quant aux $\mu_{tmod365}$, leur dispersion est telle qu'il est pratiquement impossible de retrouver, au niveau journalier, les résultats relatifs aux moyennes mensuelles.

2.3.5- Modélisation en termes d'écarts à la température normale

2.3.5.1- Une autre manière de tenir compte de la périodicité annuelle et de l'aspect autorégressif d'ordre 3 consiste à déterminer des modèles de la forme :

$$y_t = x_t - \bar{x}_t$$

$$y_t = \sum_{i=1}^3 \psi_i y_{t-i} + a_t$$

Nous introduisons ici les trois variantes de calcul de $\bar{x}_t$.

2.3.5.2- L'analyse spectrale des résidus des 3 modèles correspondants indique :

- un maximum local pour $T = 365$ en ce qui concerne la variante 1
- pas de maximum local pour $T = 365$ en ce qui concerne les variantes 2 et 3

2.3.5.3- Par rapport à la modélisation du paragraphe 2, nous obtenons des meilleurs résultats puisque l'approximation $\lambda \bar{x}_t$ n'est pas nécessaire.

Par rapport à la modélisation du paragraphe 3, nous obtenons également de meilleurs résultats en ce sens que les $\bar{x}_t$ sont beaucoup plus lisses que les $\lambda_{tmod365}$.

Par contre, la difficulté que posent les 29 Février subsiste. (maximum local pour la période annuelle avec la variante 1).

2.3.6- Utilisation de l'analyse harmonique

2.3.6.1- Nous avons d'abord recherché le nombre d'harmoniques nécessaires pour expliquer la période de 365 jours, à la fois pour la série 1 (avec le 29 Février) et la série 2 (sans les 29 Février), en ajustant des modèles de la forme :

$$x_t = a_0 + \sum_{i=1}^h \left(a_i \cos \frac{2\pi i t}{365} + b_i \sin \frac{2\pi i t}{365} \right) + \varepsilon_t$$

2.3.6.2- L'analyse spectrale des différents résidus montre :

- pour $h=5$, le spectre du résidu de la série 1 présente encore deux extrêmes marqués pour $T = 365$ et $T = 182,5$
- la périodicité annuelle est totalement prise en compte avec la série 2 pour $h=2$.

Nous retrouvons donc les difficultés que posent les 29 Février au niveau de la détermination de la périodicité annuelle.

2.3.6.3- Nous avons terminé la modélisation sur la série dans laquelle les 29 Février sont supprimés. Sur le résidu ε_t du modèle à deux harmoniques, nous avons ajusté le modèle autorégressif :

$$\varepsilon_t = 0,9868 \varepsilon_{t-1} - 0,2866 \varepsilon_{t-2} + 0,0881 \varepsilon_{t-3} + a_t$$

La variance du résidu est de 3,814 soit un écart-type de 1,953.

2.3.6.4- L'analyse harmonique donne donc de bons résultats, en particulier dans le sens d'une détermination d'un motif très régulier pour la périodicité annuelle (figure 4). La prise en compte des 29 Février conduirait sans doute à introduire une ou plusieurs paires de sinus cosinus de période 146.

Le modèle ARIMA apparaît donc plus simple, d'autant que le spectre estimé sur le résidu correspondant ne présente pas de maximum pour $T=365$ même lorsque les 29 Février sont présents dans la série. Par contre, les performances en terme d'écart-type de la prévision à un jour semblent un peu moins bonnes (2 au lieu de 1,95).

2.4- Simulation d'une utilisation en prévision du modèle ARIMA

Nous avons simulé une prévision en temps réel de la température sur les 2 ans de réalisations non utilisées au niveau de l'estimation, en supposant a priori que les 4 paramètres du modèle ARIMA n'avaient pas besoin d'être réactualisés pendant cette période.

Si x_T est la dernière température observée, la prévision $\hat{x}_{T+1}$ de variance minimale est donnée par :

$$\hat{x}_{T+1} = x_{T-364} + \hat{w}_{T+1}$$

où :

$$\hat{w}_{T+1} = \phi_1 w_T + \phi_2 w_{T-1} + \phi_3 w_{T-3} - \theta_{365} a_{T-364}$$

et : $w_t = x_t - x_{t-365} \quad t \leq T$

L'erreur de prévision à l'instant $t+1$ est :

$$a_{T+1} = x_{T+1} - \hat{x}_{T+1}$$

L'écart-type des 730 erreurs de prévision est de 1,84 ; la fonction d'autocorrélation (figure 5), la fonction d'autocorrélation partielle (figure 6), le spectre (figure 7) ne présentent pas d'éléments significatifs.

La figure 8 propose une comparaison simple entre les prévisions du modèle et les réalisations (pour la 1ère année).

2.5- Bilan de l'étude des températures

2.5.1- Nous avons donc étudié sur cette série plusieurs types de modélisations, sans toutefois prétendre à l'exhaustivité. Les méthodes "classiques" que nous avons utilisées, distinguant les corrélations longues (la périodicité annuelle) et les corrélations courtes (l'aspect autorégressif d'ordre 3 se sont trouvées gênées par la présence des 29 Février. Il semblerait qu'il soit nécessaire, pour obtenir des résultats satisfaisants au niveau de l'analyse spectrale, d'introduire non une période de 365 jours, mais une période de 1461 jours.

2.5.2- En supprimant les 29 Février, les méthodes "classiques" donnent par contre de bons résultats, en particulier l'analyse harmonique conduit à un modèle extrêmement agréable.

.../...

2.5.3- En contrepartie, le modèle ARIMA n'a pas présenté de réaction particulière à la présence des 29 Février. C'est celui que nous avons préféré et retenu pour aborder notre deuxième série, celle des énergies journalières. D'autant plus que la suppression des 29 Février, envisageable pour les températures, poserait des problèmes pour cette deuxième série, en introduisant une cassure dans la périodicité hebdomadaire.

2.5.4- Les résultats obtenus par la simulation de la prévision en temps réel de 1 température à l'aide du modèle ARIMA, en confirment la validité.

3.- SERIE DES ENERGIES JOURNALIERES

3.1- Introduction

Nous disposons de statistiques journalières du 1er Janvier 1952 à 30 Juin 1975. Nous avons ici conservé la dernière année pour réaliser une expérience de simulation semblable à celle de la température. Au niveau de l'estimation, nous ne retenons donc que les 8217 premières réalisations, soit du 1.1.1952 au 30.6.74.

L'analyse de la série observée et des estimations de ses principales propriétés permet de dégager quelques premières caractéristiques :

- existence d'une périodicité hebdomadaire
- existence d'une tendance
- existence d'une périodicité annuelle
- amplitude des oscillations de la série qui croît avec le temps
- existence d'observations particulières (jours de fête, jours de grève).

Remarque :

Cette série étant sans doute moins connue que celle des températures, nous proposons sur la figure 9 un tracé graphique d'une partie des observations, ce qui permet de préciser quelques unes des caractéristiques précédentes.

3.2- Recherche d'un ou plusieurs modèles ARIMA possibles

3.2.1- En ce qui concerne l'accroissement de l'amplitude des oscillations de la série, la construction d'un diagramme étendue-moyenne (figure 10) conduit à effectuer une transformation logarithmique sur la série initiale [11]. La série transformée est alors beaucoup plus homogène, avec cependant une légère indication de transformation trop forte.

3.2.2- Pour pouvoir ajuster un modèle ARMA, il convient ensuite de filtrer les non stationnarités sur la moyenne par des opérateurs différences. Différents essais conduisent à postuler les trois modèles suivants :

$$(1) \quad (1 - \phi_{365} B^{365}) \nabla_1 \nabla_7 \log_{10} x_t = (1 - \theta_1 B) (1 - \theta_7 B^7) a_t$$

$$(2) \quad (1 - \phi_1 B) \nabla_7 \nabla_{365} \log_{10} x_t = (1 - \theta_7 B^7) (1 - \theta_{365} B^{365}) a_t$$

$$(3) \quad \nabla_1 \nabla_7 \nabla_{365} \log_{10} x_t = (1 - \theta_1 B) (1 - \theta_7 B^7) (1 - \theta_{365} B^{365}) a_t$$

où a_t représente -pour chacun des modèles- l'erreur de prévision à l'instant

3.2.3- Comme point de départ, nous avons retenu le modèle (2). Le critère d'ajustement retenu a été :

$$\begin{array}{rcl} \text{Min} & 8217 & \\ \phi_1, \theta_7, \theta_{365} & \Sigma & a_t^2 \\ & t=2252 & \end{array}$$

Les résultats de l'ajustement sont alors :

$$(1 - 0,2919B) \nabla_7 \nabla_{365} \log_{10} x_t = (1 - 0,9686B^7) (1 - 0,7068B^{365}) a_t$$

$$\text{avec :} \quad \text{var}(a_t) = 0,0006929$$

$$\text{soit} \quad \sigma(a_t) = 0,02632 \quad *$$

L'analyse graphique du résidu de ce modèle permet de détecter de très nombreuses réalisations anormales, certaines atteignant jusqu'à 9 écart types. Ces résidus particuliers traduisent l'impact des différentes fêtes légales sur la consommation journalière d'énergie électrique.

La figure 11 propose un extrait de la série des résidus. Nous y distinguons 6 résidus fortement négatifs, suivis d'ailleurs à chaque fois par un résidu positif plus faible, très certainement induit par le précédent.

Il est vraisemblable que le modèle ARIMA est fortement perturbé par ces situations particulières. Leur modélisation permettra :

- d'en comprendre les mécanismes
- d'obtenir un modèle ARIMA moins perturbé.

.../...

Remarque : Cet écart-type, après multiplication par $\log 10 \approx 2,30$, représente l'écart-type de l'énergie journalière en pourcent de la prévision moyenne

3.3- Inventaire des jours de fête

3.3.1- Cet inventaire porte sur les résidus du modèle précédent. Nous ne détectons à ce niveau que les effets les plus importants. Au fur et à mesure que le modèle s'est amélioré, d'autres effets moins importants apparaîtront.

3.3.2- A ce stade, nous décelons des perturbations pour :

- les lundi de Pâques
- les jeudi de l'Ascension
- les lundi de Pentecôte
- le 1er Mai
- les 8 Mai
- les 14 Juillet
- les 15 Août
- les 1er Novembre
- les 11 Novembre
- les 25 Décembre
- les 1er Janvier.

Les phénomènes qui se manifestent par la suite seront essentiellement des "effets de bord" (anticipation ou rémanence sur les jours voisins des effets principaux cités) et les conséquences sur la demande d'énergie électrique des vacances du mois d'Août.

3.3.3- Nous pouvons d'emblée distinguer trois types de fêtes :

- les fêtes mobiles, sans doute assez simples à modéliser car se présentant toujours le même jour de la semaine
- les fêtes à date fixe, sans doute assez difficiles à modéliser car ayant probablement des conséquences différentes suivant le jour de la semaine où elles se produisent
- les vacances d'Août très difficiles à prendre en compte, car intervenant sur un nombre important de jours.

Nous les avons envisagées dans cet ordre.

L'idée très générale de la prise en compte des différentes fêtes a été la suivante : ramener par des corrections adéquates l'énergie électrique fournie un jour de fête à celle qui aurait été produite si le jour avait été un jour normal.

3.3.4- Nous ne décrirons pas ici l'ensemble des essais réalisés car cela impliquerait d'analyser un nombre considérable de valeurs numériques. Nous indiquerons simplement la trame des travaux effectués et illustrerons l'amélioration du modèle à la fin de chaque grande étape.

3.4- Correction des fêtes mobiles

3.4.1- Les corrections concernent Pâques, la Pentecôte et l'Ascension.

3.4.2- Nous avons tenté une première méthode dite de correction multiplicative. Soit par exemple Γ l'ensemble des indices des lundi de Pâques que nous voulons corriger par un seul coefficient λ . Nous résolvons le problème suivant :

$$\left\{ \begin{array}{l} \text{Min}_{\phi_1, \theta_7, \theta_{365}, \lambda} \sum_{t=2252}^{8217} a_t^2 \\ (1-\phi_1 B) \nabla_7 \nabla_{365} w_t = (1-\theta_7 B^7) (1-\theta_{365} B^{365}) a_t \\ w_t = z_t \quad \text{si} \quad t \notin \Gamma \\ w_t = \lambda z_t \quad \text{si} \quad t \in \Gamma \\ z_t = \log_{10} x_t \end{array} \right.$$

3.4.3- Nous avons tenté en parallèle une seconde méthode dite de correction additive. Par rapport au modèle précédent, la seule différence est :

$$w_t = z_t + \lambda \quad \text{si} \quad t \in \Gamma$$

Cette seconde méthode nous a semblé préférable au vu des résultats

3.4.4- L'analyse des résidus des jours soumis à correction montre cependant que les corrections ne sont pas suffisantes = par exemple, les résidus des lundi de Pâques sont plutôt négatifs en début de série et plutôt positifs à la fin.

La première idée a été de résoudre le même problème, avec un nombre d'observations plus limité, soit précisément :

$$\begin{array}{ccc} \text{Min} & 8217 & \\ \phi_1, \theta_7, \theta_{365}, \lambda & \sum_{t=5115} & a_t^2 \end{array}$$

L'effet de "non stationnarité" de la correction n'est alors plus perceptible.

3.4.5- Raccourcir sensiblement le nombre des observations est préoccupant, en particulier dans la perspective de l'analyse future des fêtes à date fixe. Nous avons alors introduit des corrections additives linéaires, de la forme

$$w_t = z_t + \lambda + \mu t$$

Les corrections constantes constituent une sorte de correction de niveau moyen. Par contre, ce type de correction linéaire provoque également une correction de trend, indiquant que le poids relatif des perturbations induites par les jours de fête est décroissant (les paramètres μ seront toujours négatifs).

3.4.6- L'introduction de corrections de ce type nous conduit à un modèle à 14 paramètres et la variance du résidu devient 0,0004388 soit un écart-type de 0,0209. Le modèle ARIMA numéro 2 s'écrit alors :

$$(1-0,2534B)^7 \nabla_{365} w_t = (1-0,91571B^7) (1-0,52001B^{365}) a_t$$

La structure de ce modèle est donc déjà assez différente de celle du paragraphe 3.2.3.

L'analyse graphique du résidu de ce modèle montre encore -ce qui est très normal - des résidus anormalement élevés (de l'ordre de 9 écarts-types) : ils sont dus aux fêtes à date fixe non encore prises en compte.

3.5- Correction des fêtes à date fixe

3.5.1- Ces corrections concernent les 1er Mai, 8 Mai, 14 Juillet, 15 Août, 1er Novembre, 11 Novembre, 25 Décembre et 1er Janvier.

On pouvait espérer que l'opérateur ∇_{365} ait pris en compte ces effets en procédant à une autocorrection. En fait, il n'en est rien pour les deux raisons suivantes :

- l'existence des années bissextiles
- l'effet des fêtes à date fixe dépend du jour de la semaine où elles se produisent.

3.5.2- Aux deux types de corrections précédentes, nous avons dû ajouter des corrections paraboliques de la forme :

$$w_t = z_t + \lambda + \mu t + \nu t^2$$

3.5.3- Le dernier modèle que nous avons mis au point dans le contexte de ce paragraphe comporte 100 paramètres (3 pour le modèle ARIMA proprement dit et 97 pour les diverses corrections). La variance du résidu est de 0,00006274 soit un écart-type de 0,007921. Le modèle ARIMA s'écrit alors :

$$(1-0,72934B)\nabla_7\nabla_{365}w_t = (1-0,37305B^7)(1-0,82241B^{365})a_t$$

Cette fois-ci le changement de structure du modèle est extrêmement important.

3.6- Prise en compte de la cassure du mois d'Août

3.6.1- Ces perturbations se sont avérées impossibles à modéliser à l'aide d'un nombre réduit de paramètres, étant nettement plus complexes que les précédentes : un fort résidu positif apparaît pour un dimanche proche du 1er Août, suivi par un autre résidu important - négatif cette fois-ci - pour le lundi suivant : ce motif se retrouve en général à l'identique, bien que plus faible 7 jours après. Au voisinage de la fin du mois d'Août, un effet symétrique se manifeste : fort résidu négatif un dimanche et positif le lundi suivant avec existence du même motif, plus amorti, 7 jours plus tôt.

.../...

Le premier modèle ayant donné satisfaction nous a conduits à introduire 37 variables supplémentaires en distinguant deux configurations :

- 1er Août lundi, mardi, vendredi, samedi, dimanche
- 1er Août mercredi, jeudi.

3.6.2- La variance du résidu devient 0,000038086, soit un écart-type de 0,00617. Le modèle ARIMA numéro 2 devient :

$$(1-0,81765B)\nabla_7\nabla_{365} w_t = (1-0,78773B^7)(1-0,78855B^{365}) a_t$$

3.6.3- Ce modèle prend bien en compte la plupart des effets dus aux vacances d'Août. Il reste cependant des effets systématiques du second ordre qui, pour être modélisés demanderaient de l'ordre de 100 paramètres supplémentaires ! Nous y avons renoncé car :

- toutes les principales perturbations sont prises en compte et nous pensons que la méthode de correction retenue est pertinente
- les temps de calcul nécessaires à la résolution des différents modèles deviennent extrêmement importants
- les corrections sont assez fines pour aborder le problème de la sélection d'un des trois modèles postulés en 3.2.2.

3.6.4- Le dernier modèle estimé comporte 138 paramètres. Les trois premiers sont relatifs au modèle ARIMA et nous les avons donnés ci-dessus. L'Annexe propose quelques exemples caractéristiques de coefficients correcteurs.

3.7- Choix d'un modèle (voir paragraphe 3.2.2)

3.7.1- Les modèles (1) et (2) sont équivalents en terme de variance du résidu. Le modèle (3) semble un peu moins bon.

3.7.2- Le modèle (1) se présente comme le plus perturbé - au niveau du corrélogramme - par les effets des jours de fête qui subsistent : c'est donc sans doute celui sur lequel les progrès au niveau des corrections se traduiront par les progrès les plus importants.

3.7.3- Le modèle ARIMA retenu à ce stade est donc le modèle (1), qui s'écrit alors :

$$(1-0,06608B^{365})\nabla_1\nabla_7w_t = (1-0,09906B)(1-0,82560B^7)a_t$$

où $w_t = \log_{10}x_t$ pour les jours normaux

$w_t = \log_{10}x_t + \text{correction pour les jours soumis à correction.}$

La variance du résidu est estimée à 0,00003798, soit un écart-type de 0,00616.

L'analyse des estimations des divers coefficients correcteurs par rapport à ceux du modèle (2) montre que :

- les résultats sont pratiquement invariants pour les fêtes à date fixe et les fêtes mobiles ce qui est une bonne confirmation de la pertinence de la modélisation
- les résultats sont moins stables pour les vacances d'Août ce qui illustre un besoin d'affiner la modélisation à ce stade.

3.8- Simulation de l'utilisation en temps réel du modèle ARIMA numéro (1)

3.8.1- Nous avons simulé une prévision en temps réel, à l'aide du modèle 1, pour l'année de réalisations non prises en compte dans la phase d'estimation (du 1er Juillet 1974 au 30 Juin 1975). Nous avons supposé a priori qu'il n'était pas nécessaire de réestimer les divers paramètres sur toute cette période.

3.8.2- Soit T le nombre total des observations et soit x_t l'observation courante. Nous définissons :

$$\begin{aligned} z_t &= \log_{10}x_t & t \leq T \\ w_t &= z_t + \text{éventuelle correction} & t \leq T \\ \alpha_t &= w_t - w_{t-1} & t \leq T \\ d_t &= \alpha_t - \alpha_{t-7} & t \leq T \\ a_t & \text{ le résidu du modèle ARIMA} & t \leq T \end{aligned}$$

Pour prévoir $\hat{x}_{T+1}$, nous effectuons en premier lieu une prévision $\hat{z}_{T+1}$, de la manière suivante :

$$\hat{d}_{T+1} = \phi_{365} d_{T-364} - \theta_1 a_T - \theta_7 a_{T-6} + \theta_1 \theta_7 a_{T-7}$$

$$\hat{a}_{T+1} = \hat{d}_{T+1} + \alpha_{T-6}$$

$$\hat{w}_{T+1} = \alpha_{T+1} + w_T$$

$$\hat{z}_{T+1} = \hat{w}_{T+1} - \text{éventuelle correction}$$

Puis, bien que cela soit une approximation [12], nous calculons simplement :

$$\hat{x}_{T+1} = 10^{\hat{z}_{T+1}}$$

L'erreur de prévision à l'instant T+1 est alors :

$$a_{T+1} = \log_{10} x_{T+1} - \hat{z}_{T+1}$$

Avant de prévoir l'énergie du jour suivant, il convient de mettre à jour les vecteurs suivants :

$$z_{T+1} = \log_{10} x_{T+1}$$

$$w_{T+1} = z_{T+1} + \text{correction éventuelle}$$

$$\alpha_{T+1} = w_{T+1} - w_T$$

$$d_{T+1} = \alpha_{T+1} - \alpha_{T-6}$$

3.8.3- L'écart-type des 365 erreurs de prévision est 0,00244. Les résidus (figure 12) sont en général à l'intérieur ou au voisinage de la bande de confiance à 5 % du bruit blanc normal ; quatre d'entre eux apparaissent nettement en dehors de la bande de confiance :

- ① 1er Août 1974 (Dimanche) : un des effets du mois d'Août non pris en compte par le modèle
- ② 11 Novembre 1974 (Lundi) : déjà pris en compte par une correction linéaire (ce genre de résultats nous avait conduits pour Noël et le Nouvel An à introduire des corrections quadratiques)
- ③ 23 Décembre 1974 (Lundi) : un effet des fêtes de Noël non pris en compte par le modèle
- ④ 5 Mai 1975 (Lundi) : bien qu'il n'y ait pas de fête très voisine, cette situation demanderait une analyse approfondie permettant de savoir si cette erreur peut être considérée comme "due au hasard" ou non.

La figure 16 compare prévisions et réalisations en termes d'énergie proprement dite.

3.8.4- La fonction d'autocorrélation (figure 13), la fonction d'autocorrélation partielle (figure 14) ne montrent pas d'aspect particulier (3 à 4 des estimations en dehors de la bande de confiance à 5 %), sauf peut-être à l'ordre 1 pour lequel les estimations semblent significatives. Le spectre (figure 15) ne révèle rien d'évident.

3.9- Bilan de l'étude des énergies journalières

3.9.1- Ainsi que nous pouvions le pressentir a priori, la modélisation de la série des énergies journalières est loin d'être une chose aisée en raison de l'abondance des jours de fête qui ont une influence notable sur la demande d'énergie en France.

3.9.2- La mise au point d'une succession de modèles ARIMA, avec prise en compte progressive d'un nombre croissant de perturbations, a permis de réaliser un modèle dont les caractéristiques sur un plan statistique sont, sans être encore parfaites, déjà convenables.

3.9.3- Nous obtenons un dernier modèle dont le nombre de paramètres est très important, mais dont très peu s'occupent du modèle ARIMA proprement dit. Ce qui peut sembler être une prolifération de coefficients correcteurs mérite ici une explication. La correction des perturbations introduites par les très nombreux jours particuliers relevait de deux soucis :

- éliminer les perturbations pour que le modèle ARIMA s'adapte le mieux possible aux jours normaux. L'attitude extrême à cet égard aurait été d'introduire un coefficient chaque fois qu'un jour perturbé se serait présenté. Ceci aurait évidemment conduit à un nombre de paramètres extrêmement élevé
- essayer de comprendre les perturbations en se proposant d'en dégager des mécanismes généraux.

Il fallait donc trouver un compromis entre le désir d'obtenir un modèle ARIMA le plus pur possible et celui de mettre en évidence un certain nombre de mécanismes sous-jacents aux perturbations de la série. C'est ce souci qui nous a conduits à accroître, par étapes successives, le nombre de paramètres des modèles et c'est la seule complexité des phénomènes qui en a provoqué l'importante augmentation.

3.9.4- Nous pensons que les résultats de l'expérience de prévision en temps réel sont tout à fait encourageants et donnent une bonne validation de l'approche que nous avons développée.

4.- CONCLUSION ET PERSPECTIVES

4.1- En ce qui concerne la prévision à très court terme de la température, nous proposons un modèle qui peut être considéré comme définitif pour la prévision d'une valeur moyenne à un jour. La définition d'intervalles de confiance demanderait un léger travail complémentaire (soit lissage des écarts-types journaliers, soit recherche d'une transformation optimale).

Nous pensons que ce travail effectué sur les relevés de Paris-Montsouris pourrait s'étendre aisément aux relevés d'autres stations de la Météorologie Nationale.

4.2- En ce qui concerne la prévision à très court terme des énergies journalières, nous proposons un modèle qui possède déjà de bonnes propriétés. Des compléments du deuxième ordre pourraient encore être apportés pour les vacances du mois d'Août, pour quelques fêtes non prises en compte (Saint Eloi et Sainte Barbe par exemple). Un complément plus important serait de déterminer une transformation "un peu moins forte" que la transformation logarithmique utilisée.

De la même manière que pour les températures, on pourrait envisager d'analyser des statistiques de production journalières régionalisées.

4.3- Les deux séries que nous avons envisagées n'ont pas été choisies par hasard. Une troisième analyse, complémentaire des deux précédentes, amènerait à mettre au point une fonction de transfert entre température de Paris-Montsouris et production journalière d'énergie électrique en France. L'annexe III développe les premières idées dans cette direction.

4.4- La technique d'optimisation retenue, utilisant le code GRGA [17], pourrait également donner lieu à une analyse spécifique. Les gros modèles auxquels nous sommes arrivés demandaient des temps de calcul importants pour leur estimation. Il faut voir d'ailleurs dans ce fait que toutes les voies de recherche complémentaires définies ci-dessus n'ont pu être explorées aussi bien que nous l'aurions souhaité.

4.5- Une extension très intéressante des travaux entrepris serait d'envisager la mise au point de modèles de prévision à très court terme de la courbe de charge. Nous pensons facile de postuler un modèle pour les jours normaux et assez facile de l'estimer. Par contre, nous estimons qu'il serait déraisonnable d'envisager la prise en compte des phénomènes de fête de la manière retenue pour

les énergies journalières : les temps de calcul seraient nécessairement très élevés et les temps d'analyse des résultats se compteraient en jours par modèle testé. Est-ce à dire qu'une telle analyse est impossible ? Nous ne le pensons pas et nous voudrions donner une idée de prise en compte des perturbations des jours de fête sur courbe de charge.

Considérons l'exemple des lundi de Pâques, au niveau des énergies journalières. Ce modèle nous a conduits à introduire une correction additive linéaire sur le logarithme des énergies. Or si l'on trace un diagramme (figure 17) représentant l'évolution pour chaque année du rapport de l'énergie réalisée du lundi de Pâques à une estimation grossière de ce qu'aurait été cette énergie pour un jour normal, nous retrouvons exactement la même idée. Une méthode moins rigoureuse que celle que nous avons retenue serait alors d'estimer par les moindres carrés une droite sur le nuage de points de la figure 21, de corriger les lundi de Pâques à l'aide des résultats correspondants et d'estimer le modèle ARIMA ultérieurement. Il est plausible que, sur un plan pratique, cette méthode donne des résultats satisfaisants. Elle éviterait en tout cas, au niveau de l'analyse des courbes de charge, de donner des modèles avec un nombre de paramètres très élevé.

4.6- Pour terminer, nous voudrions revenir sur les temps de calcul cités plus haut. Il est certain qu'ils sont très importants, mais il convient de souligner qu'ils n'interviennent qu'au niveau de la phase d'estimation des paramètres du modèle retenu.

Au niveau de la prévision, les temps de calcul sont au contraire extrêmement faibles. La prévision à un jour est, dans le cadre du modèle retenu pour les énergies journalières, obtenue de la manière suivante :

- calcul de $\hat{w}_t$ suivant :

$$\hat{w}_t = w_{t-1} + w_{t-7} - w_{t-8} - 0,06608w_{t-365} - 0,06608w_{t-372} + 0,06608w_{t-373} - 0,09906a_{t-1} - 0,82560a_{t-7} + 0,08178a_{t-8}$$

c'est-à-dire quelques additions et multiplications.

- pour un jour soumis à correction, il convient ensuite d'apporter la modification convenable, soient quelques tests pour trouver les bons coefficients correcteurs, suivis de quelques additions et multiplications
- une exponentiation du résultat pour obtenir la prévision effective en MWh.

.../...

Les temps de calcul nécessaires à une prévision sont donc extrêmement faibles (de l'ordre de quelques millièmes de seconde). De plus, en ayant adopté une série assez longue, il est probable que la réactualisation des coefficients estimés n'interviendrait que peu souvent (une fois par mois dans le pire des cas) ; en tout état de cause, il serait très facile de provoquer un déclenchement quasi automatique d'une réestimation des paramètres : il suffit par exemple de suivre les résidus de prévision et d'en estimer diverses propriétés (le corrélogramme ou le spectre par exemple) et de déclencher la procédure d'ajustement dès qu'une hypothèse de bruit blanc, définie a priori, n'est plus satisfaite. Par ailleurs, le modèle aurait alors la même structure et on pourrait calculer exactement les dérivées utilisées par la méthode de gradient, ce qui conduirait à des temps d'estimation nettement plus faibles que ceux obtenus jusqu'alors.

B I B L I O G R A P H I E

- [1] F. MESLIER : Thèse de doctorat de 3ème cycle - Université Paris IX-1976
"Contribution à l'analyse des séries chronologiques et application à la mise au point de modèles de prévision à court terme relatifs à la demande journalière d'énergie électrique en France et à la température relevée à Paris-Montsouris"
- [2] J. ABADIE - F. MESLIER : Note interne EDF HR 31-0146 du 12/1/76 "Etude de 1 série chronologique des températures de Paris-Montsouris"
- [3] J. ABADIE - F. MESLIER : Note interne EDF HR 31-0147 du 12/1/76 "Etude de 1. série chronologique des températures de Paris-Montsouris - Annexes"
- [4] J. ABADIE - F. MESLIER : Note interne EDF HR 31-0148 du 12/1/76 "Simulation de la température de Paris-Montsouris"
- [5] J. ABADIE - F. MESLIER : Note interne EDF HR 31-0178 du 24/6/76 "Analyse de la série chronologique des températures moyennes journalières de Paris-Montsouris. Réponse aux commentaires faits après la publication des notes HR 31-0146 à 31-0148 étudiant la même série"
- [6] J. ABADIE - F. MESLIER : Note interne EDF HR 31-0179 du 24/6/76 "Analyse de la série chronologique des températures moyennes journalières de Paris Montsouris. Réponse aux commentaires faits après la publication des notes HR 31-0146 à 31-0148 étudiant la même série - Annexes".
- [7] BOX JENKINS : Holden Day - San Francisco 1970 "Time Series Analysis Forecasting and Control"
- [8] BOX AND COX - Journ. Roy. Stat. Soc. B26,211-1964 "An analysis of transformations"
- [9] Gaz de France - Service des Etudes Economiques - Note du 13/12/71 "Etude des températures quotidiennes de Paris-Montsouris sur la période 1896-1970"
- [10] F. BROUSSOLLE - Note interne EDF HR 10664/3 du 5/3/73 "Inversion des matrices creuses et calcul de courants de court-circuit dans les réseaux électriques"

- [11] BOX AND JENKINS : Journ. Roy. Stat. Soc. A136,3-1973 - "Some comments on a paper by Chatfield and Prothero and on a review by Kendall (with a reply)"
- [12] GRANGER NEWBOLD - Journ. Royal. Stat. Soc. n°2-1976 "Forecasting Transform Series"
- [13] DURBIN - Rev. Int. Inst. Stat 28, 233 - 1960 "The fitting of time series models"
- [14] GRANGER HATANAKA - Princeton 1964 "Spectral Analysis of Economic Time Series"
- [15] MARQUARDT - Journ. Soc. Ind. Appl. Math. 11,431-1963 "An algorithm for least squares estimation of non linear parameters".
- [16] J. ABADIE - J. CARPENTIER in "Optimization" (FLETCHER editeur) Academic Press - 1969 "Generalization of the Wolfe Reduced Gradient Method to the case of non linear constraints"
- [17] J. ABADIE - Note interne EDF HI1758 - Février 1975 "Méthode du gradient réduit généralisé : le code GRGA"
- [18] FLETCHER REEVES - Comput. Journ. 6, 194 "Function Minimization by conjugate gradients"
- [19] LAUGAG - Thèse de 3eme cycle - Université Paris VI 1976 "Etude d'algorithmes d'optimisation non linéaire (une variante du GRG)".

- ANNEXE I -

ESTIMATION DES PROPRIETES DU PREMIER ET DU DEUXIEME ORDRE D'UNE SERIE
CHRONOLOGIQUE STATIONNAIRE

Soient $x_1, x_2, \dots, x_n$ les diverses observations. Nous définissons :

- la moyenne
$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t$$

- la fonction d'autocovariance $C(i)$:

$$C(i) = \frac{1}{n-i+1} \left[\sum_{t=i}^n x_t x_{t-i+1} - \frac{1}{n-i+1} \sum_{t=i}^n x_t \sum_{t=1}^{n-i+1} x_t \right]$$

- la fonction d'autocorrélation $r(i)$:

$$r(i) = \frac{C(i)}{C(0)}$$

- la fonction d'autocorrélation partielle $P(i,i)$ (Méthode de Durbin [13])

$$\left\{ \begin{array}{l} P(i,i) = \begin{cases} r(1) & i = 1 \\ \frac{r(i) - \sum_{j=1}^{i-1} P(i-1,j)r(i-j)}{1 - \sum_{j=1}^{i-1} P(i-1,j)r(j)} & i = 2, 3, \dots \end{cases} \\ P(i,j) = P(i-1,j) - P(i,i)P(i-1,i-j) & j = 1, 2, \dots, i-1 \end{array} \right.$$

- la transformée de l'autocovariance $T(i)$ (Estimation au sens de Tukey Hanning [14])

$$T(i) = \sum_{j=1}^{m+1} C'(j) \cos \frac{(j-1)(i-1)}{m} \pi \quad i = 1, \dots, m+1$$

$$\text{où : } C'(j) = C(j) \text{ pour } j = 1, m+1$$

$$C'(j) = 2C(j) \text{ pour } j = 2, 3, \dots, m$$

- le spectre (Estimation au sens de Tukey Hanning [14]) :

$$S(1) = 0,5T(1) + 0,5T(2)$$

$$S(i) = 0,25T(i-1) + 0,5T(i) + 0,25T(i+1) \quad i = 2, \dots, m$$

$$S(m+1) = 0,5T(m) + 0,5T(m+1)$$

- ANNEXE II -

ESTIMATION DES MODELES ARMA

1.- Nous n'avons considéré que des estimations au sens des moindres carrés, l'hypothèse de normalité des résidus ayant toujours été acceptable. L'estimation des paramètres des modèles ARMA peut se définir (sur un ARMA(2,1) pris à titre d'exemple), comme étant la résolution du programme mathématique :

$$\left\{ \begin{array}{l} \text{Min} \quad \sum_{t=1}^T a_t^2 \\ \phi_1, \phi_2, \theta_1 \\ a_t = w_t - \phi_1 w_{t-1} - \phi_2 w_{t-2} + \theta_1 a_{t-1} \quad t = 1, \dots, T \end{array} \right. \quad (1)$$

plus un certain nombre de conditions assurant la stationnarité et l'inversibilité du processus, pour notre exemple :

$$\left\{ \begin{array}{l} \phi_2 + \phi_1 < 1 \\ \phi_2 - \phi_1 < 1 \\ \phi_2 \in]-1 ; +1[\\ \theta_1 \in]-1 ; +1[\end{array} \right.$$

Remarque :

La relation de récurrence (1) pose un problème d'initialisation. Pour $t=1$, elle fait appel à des quantités non disponibles. A cet égard, plusieurs solutions sont possibles :

- introduire les variables non connues comme paramètres à optimiser, c'est-à-dire résoudre :

$$\text{Min}_{\phi_1, \phi_2, \theta_1, a_0, w_0, w_{-1}} \sum_{t=1}^T a_t^2$$

Cette solution est envisageable pour des modèles non saisonniers. Par contre, le nombre de paramètres à ajouter devient très important pour des modèles saisonniers de période longue (environ 365 paramètres seraient nécessaires pour nos modèles de prévision de la température et de l'énergie journalière).

- utiliser la technique de backforecasting développée par Box et Jenkins [7], en particulier lorsque la série est courte. Cette solution complique l'implantation sur ordinateur des modèles envisagés

- lorsque la série est longue, introduire des valeurs arbitraires (0 par exemple) et prendre un anneau de garde, c'est-à-dire résoudre un problème de la forme

$$\text{Min}_{\phi_1, \phi_2, \theta_1} \sum_{t=T_1}^T a_t^2$$

C'est la solution que nous avons retenue, nos séries étant exceptionnellement longues.

L'estimation d'un modèle ARMA conduit donc en général à la résolution d'un programme non linéaire, sous des contraintes qui deviennent de plus en plus complexes à mesure que le degré du processus augmente. Heureusement, l'expérience numérique montre que si l'identification du processus est correcte, alors le résultat de l'estimation sera un processus stationnaire et inversible sans tenir compte explicitement des contraintes correspondantes.

2.- Box et Jenkins [7] recommandent un algorithme dérivé de celui de Marquardt [15]. Nous ne l'avons pas adopté pour les deux raisons principales suivantes :

- il est nécessaire de calculer les $\frac{\partial a_t}{\partial \beta_i}$ (les β_i étant les différents paramètres) pour nos plus gros modèles, le volume d'informations correspondant aurait été incompatible avec un programme informatique "in-core" (maximum de 2000 K sur l'ordinateur utilisé)
- il semble également nécessaire d'avoir de bonnes estimations initiales des paramètres pour obtenir une convergence acceptable. Les calculs préalables auraient été considérables, surtout pour la série des énergies.

La technique d'optimisation adoptée, utilisant le code GRGA [17], est plus générale et ne demande que le calcul des dérivées $\frac{\partial \sum a_t^2}{\partial \beta_i}$. Après plusieurs essais prenant en compte les contraintes (conditions de stationnarité et d'inversibilité), le code a été utilisé comme un code d'optimisation sans contraintes (méthode des gradients conjugués [18]).

3.- La technique d'optimisation retenue devrait cependant donner lieu à une analyse spécifique: GRGA permettait d'aborder sans difficulté l'estimation des divers modèles tant que le nombre de paramètres n'était pas trop important (avec 6000 a_t , quatre minutes avec 19 paramètres, trente minutes avec 40 paramètres).

Par contre, un nombre important de paramètres nous conduisait à des temps de calcul extrêmement importants (5 heures pour 100 paramètres, 10 heures pour 138 paramètres sur un IBM 370/168).

L'une des raisons de l'obtention de temps de calcul très élevés provient sans doute du mode de calcul des dérivées, par différences centrales, demandant pour 100 variables, 200 appels de fonction pour calculer le gradient, soient 200 explorations de la série temporelle : une formule exacte pour le gradient ne demanderait qu'une seule exploration. En contrepartie, la détermination analytique du gradient est extrêmement compliquée et tout est à remettre en cause dès qu'une modification est introduite dans le modèle. Une fois le modèle fixé (pour la recherche de la meilleure transformation par exemple), il est certainement rentable d'effectuer les calculs nécessaires ; réduction importante du temps de calcul peut en être attendue.

Une analyse plus spécifique pourrait porter sur la technique d'optimisation proprement dite. Quelques essais utilisant un autre code [19] ont été très encourageants pour un faible nombre de paramètres : temps divisé par 2 pour 19 paramètres. Par contre, un essai sur un plus grand modèle (100 paramètres) a donné de moins bons résultats que GRGA, au bout de 100 itérations. Une poursuite des travaux dans cette direction devrait également conduire à des réductions des temps d'estimation.

- ANNEXE III -

PREMIER ESSAI D'IDENTIFICATION D'UNE FONCTION DE TRANSFERT LIANT L'ENERGIE JOURNALIERE A LA
TEMPERATURE JOURNALIERE

Soient x_t, y_t les observations courantes de la température et de l'énergie et soit n le nombre total d'observations. Nous transformons en premier lieu les 2 séries pour induire leur stationnarité, c'est-à-dire calculons :

$$\mu_t = \sqrt[7]{7^{365}} x_t \quad t = 1, \dots,$$

$$v_t = \log_{10} y_t + \text{correction pour les jours de fête } t = 1, \dots,$$

$$w_t = \sqrt[7]{7^{365}} v_t \quad t = 1, \dots$$

Nous calculons la covariance des deux séries ainsi transformées :

$$CC(i) = \frac{1}{n-i+1} \left[\sum_{t=i}^n w_t \mu_{t-i+1} - \frac{1}{n-i+1} \sum_{t=i}^n w_t \sum_{t=1}^{n-i+1} \mu_t \right]$$

Ceci conduit à la fonction de corrélation entre les deux séries, représentée par la figure 18. L'analyse de cette estimation montre :

- une forte valeur à l'ordre 0, suivi de valeurs plus faibles aux ordres 1 et 2, le tout semblant décroître de manière exponentielle
- une forte valeur à l'ordre 7 et rien de significatif à l'ordre 14 (la valeur à l'ordre 8 est certainement induite par celles des ordres 1 et 7)
- une forte valeur à l'ordre 365 et rien de significatif à l'ordre 730 (la valeur à l'ordre 366 est certainement induite par celle des ordres 1 et 365).

Ceci conduit à postuler la fonction de transfert suivante :

$$(1 - \delta_1 B) v_t = w_0 (1 - w_7 B^7) (1 - w_{365} B^{365}) x_t + N_t$$

où N_t représente la "part" des v_t qui ne peut être expliquée par la température ; pour N_t nous adopterons bien entendu la structure du modèle endogène de la série des énergies, c'est-à-dire :

$$(1 - \phi_{365} B^{365}) \sqrt[7]{7^{365}} N_t = (1 - \theta_1 B) (1 - \theta_7 B^7) a_t$$

La combinaison de ces deux termes conduit au modèle, suivant :

$$(1-\delta_1 B) \nabla \nabla_7 v_t = w_0 (1-w_7 B^7) (1-w_{365} B^{365}) \nabla \nabla_7 x_t - \frac{(1-\theta_1 B)(1-\theta_7 B^7)}{1-\phi_{365} B^{365}} a_t$$

L'estimation du modèle conduirait à résoudre :

$$\text{MIN}_{\delta_1, w_0, w_7, w_{365}, \theta_1, \theta_7, \phi_{365}} \sum_t a_t^2$$

La technique utilisée pour les modèles endogènes est ici directement applicable.

Remarques :

α) Ce modèle est à considérer comme une première tentative de prise en compte de l'influence de la température sur la consommation. Les résultats de l'estimation, en particulier l'analyse des résidus, peut induire des modifications de cette première identification. Cependant, ce modèle comprend des mécanismes bien connus par ailleurs, essentiellement la variation de la sensibilité de la consommation à la température en fonction du jour dans la semaine et du jour dans l'année (termes $1-w_7 B^7$ et $1-w_{365} B^{365}$ respectivement). Le terme $(1-\delta_1 B)$ caractérise un phénomène non pris en compte dans les procédures actuelles, à savoir une certaine inertie de la consommation face aux dernières températures réalisées.

β) Pour une utilisation dans des conditions réelles de prévision, la température x_t n'est évidemment pas connue et il conviendrait de la remplacer par une prévision $\hat{x}_t$ (réalisée à l'aide de notre modèle ARIMA, ou en utilisant d'autres techniques de prévision). Il serait peut-être également intéressant de substituer aux x_t des $\hat{x}_t$ et ce dès la phase d'estimation du modèle. Seules diverses expériences de simulation en temps réel permettraient de se décider en connaissance de cause.

EXEMPLES DE COEFFICIENTS CORRECTEURS OBTENUS A L'AIDE DU MODELE (2) FINAL

—————> une fête mobile : Pâques

samedi	$\lambda_1 = 0,01774$	
dimanche	$\lambda_2 = 0,02631$	
lundis	$\lambda_3 = 0,22105$	$\mu_3 = -0,09576 \times 10^{-4} *$
mardis	$\lambda_4 = 0,01838$	

—————> une fête à date fixe : 11 Novembre

jours ouvrables	$\lambda_{42} = 0,21770$	$\mu_{42} = -0,13251 \times 10^{-4}$
samedis	$\lambda_{43} = 0,15234$	$\mu_{43} = -0,11541 \times 10^{-4}$
ponts lundi	$\lambda_{44} = 0,04273$	
ponts samedi	$\lambda_{45} = 0,05532$	
mardis après	$\lambda_{46} = 0,01362$	
mercredis après	$\lambda_{47} = 0,01278$	
jeudis après	$\lambda_{48} = 0,01730$	
vendredis après	$\lambda_{49} = 0,02767$	
dimanches après	$\lambda_{50} = 0,01591$	

—————> une autre fête à date fixe : 25 Décembre

jours ouvrables	$\lambda_{51} = 0,34729$	$\mu_{51} = -0,61609 \times 10^{-4}$	$\nu_{51} = 0,49755 \times 10^{-8}$
samedis	$\lambda_{52} = 0,12380$		
dimanches	$\lambda_{53} = 0,02681$		
ponts lundis	$\lambda_{54} = 0,10241$		
ponts samedis	$\lambda_{55} = 0,08740$		
mardis avant	$\lambda_{56} = 0,02345$		
mercredis avant	$\lambda_{57} = 0,01479$		
jeudis avant	$\lambda_{58} = 0,02126$		
vendredis avant	$\lambda_{59} = 0,04025$		
samedis avant	$\lambda_{60} = 0,03185$		
lundis après	$\lambda_{61} = 0,03319$		
mardis après	$\lambda_{62} = 0,03443$		
mercredis après	$\lambda_{63} = 0,02152$		
jeudis après	$\lambda_{64} = 0,03580$		
vendredis après	$\lambda_{65} = 0,05869$		
dimanches après	$\lambda_{66} = 0,03051$		

* Certains de ces paramètres peuvent sembler extrêmement faibles. Cependant, le tableau 2 montre par exemple que le coefficient μ_3 a un rôle très important sur les corrections du lundi de Pâques.

- T A B L E A U 1 -

LES DIFFERENTS FACTEURS DE MODULATION

	ETAPE	FACTEUR DE MODELISATION
Niveau de l'énergie journalière ←	1- Conversion de la valeur de l'énergie annuelle en niveaux d'énergie par jour ouvrable (moyenne annuelle)	Nombre de jours ouvrables équivalents = N
	2- Consommation par jour ouvrable moyen, pour les divers intervalles (semaine ou mois)	Coefficient saisonnier = k_s
	3- Effet de l'augmentation de la consommation au cours de la période annuelle couverte	Coefficient de trend = r
	4- Conversion de l'énergie du jour ouvrable moyen d'un intervalle donné, en énergies pour chaque type de jour (samedi, dimanche ..)	Facteurs de pondération des divers types de jours = ρ
Courbe de charge horaire ←	5- Ventilation de l'énergie journalière en 24 valeurs de puissances horaires	Courbes de charges réduites = π
	6- Prise en compte de l'effet climatique	Gradient = ρ

- T A B L E A U 2 -

COMPARAISON ENTRE UNE CORRECTION CONSTANTE ET UNE CORRECTION LINEAIRE POUR LES LUNDI DE PAQUES

Années	Numéros des observations	Correction constante	Correction linéaire
1958	2289	0,1601	0,1991
1959	2646	"	0,1957
1960	3031	"	0,1920
1961	3381	"	0,1886
1962	3766	"	0,1849
1963	4123	"	0,1815
1964	4473	"	0,1782
1965	4858	"	0,1745
1966	5215	"	0,1711
1967	5565	"	0,1677
1968	5950	"	0,1640
1969	6307	"	0,1606
1970	6664	"	0,1572
1971	7042	"	0,1536
1972	7399	"	0,1501
1973	7734	"	0,1465
1974	8141	"	0,1430

.../...

FIGURE 1

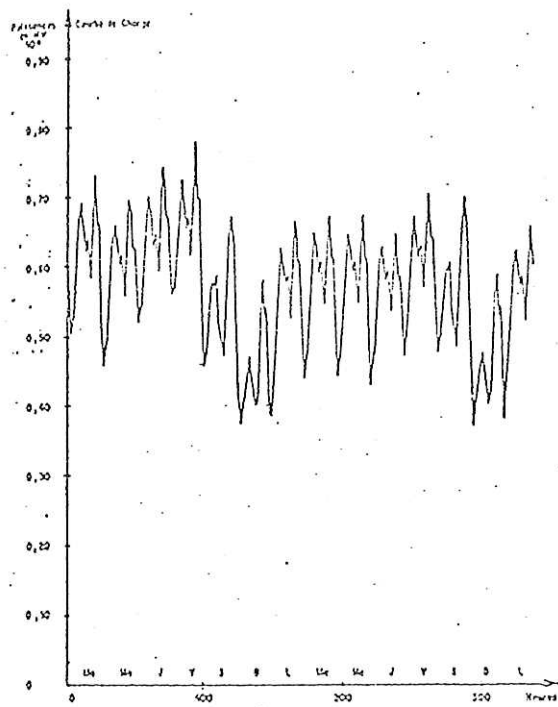


FIGURE 4

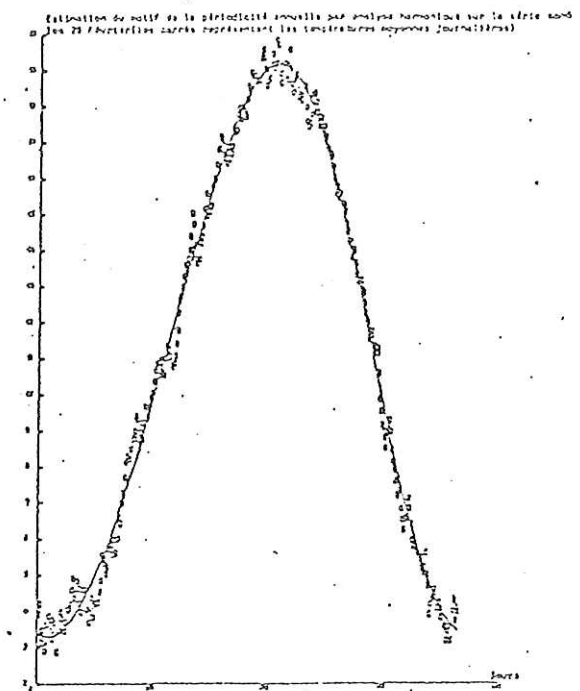


FIGURE 2

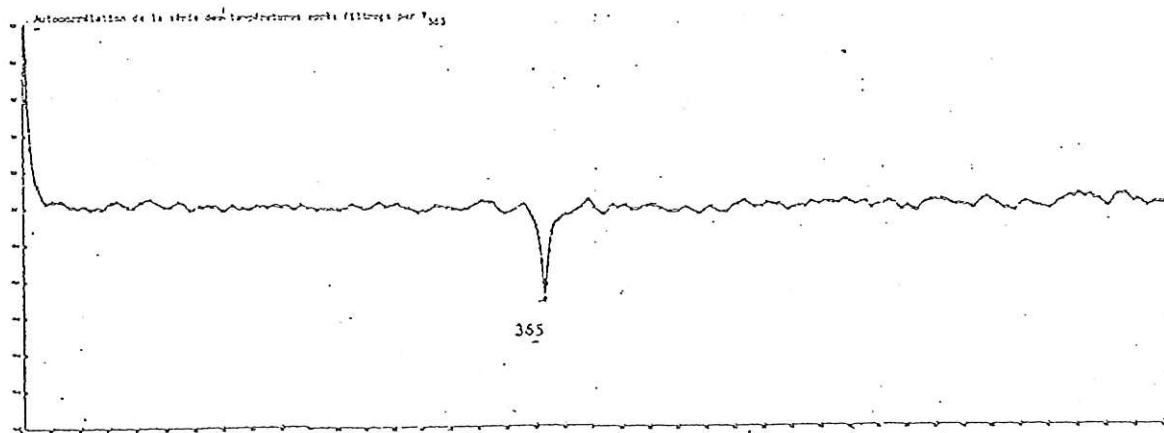


FIGURE 3

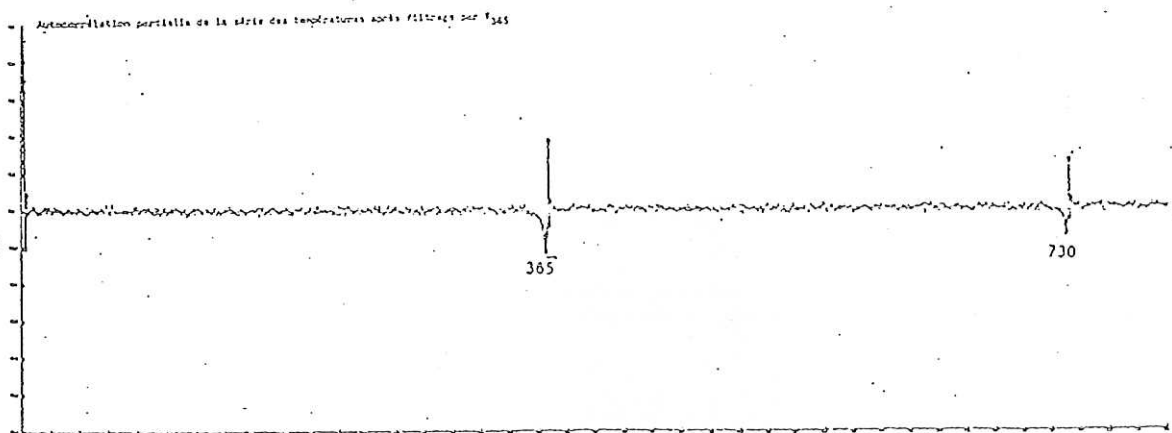


FIGURE 5
FONCTION D'ANTICORRELATION DES 750 ÉLÉMENTS DE PRÉVISION
DE LA SÉRIE DES TEMPÉRATURES

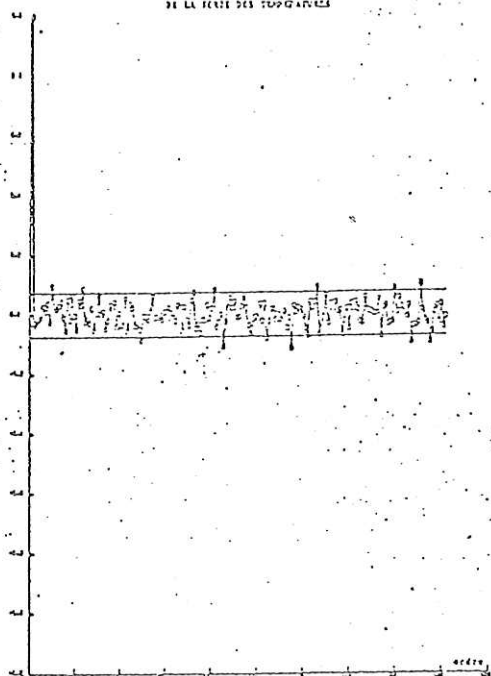


FIGURE 6
FONCTION D'ANTICORRELATION PARTIELLE DES 750 ÉLÉMENTS DE PRÉVISION
DE LA SÉRIE DES TEMPÉRATURES

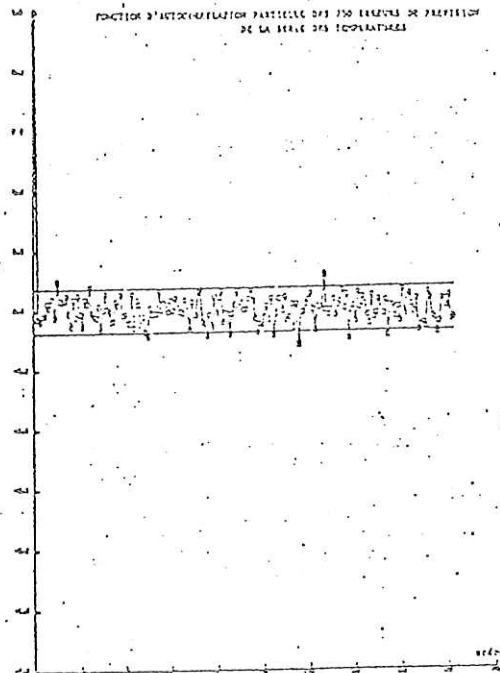


FIGURE 7
MATRIÈRE DES 750 ÉLÉMENTS DE PRÉVISION DE LA SÉRIE DES TEMPÉRATURES

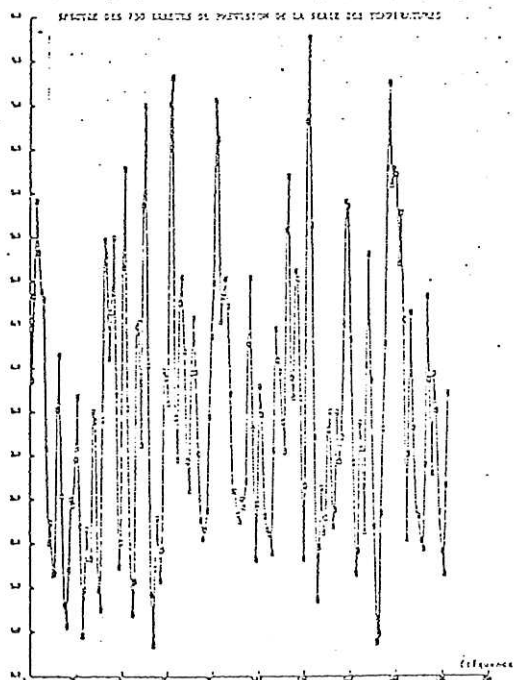
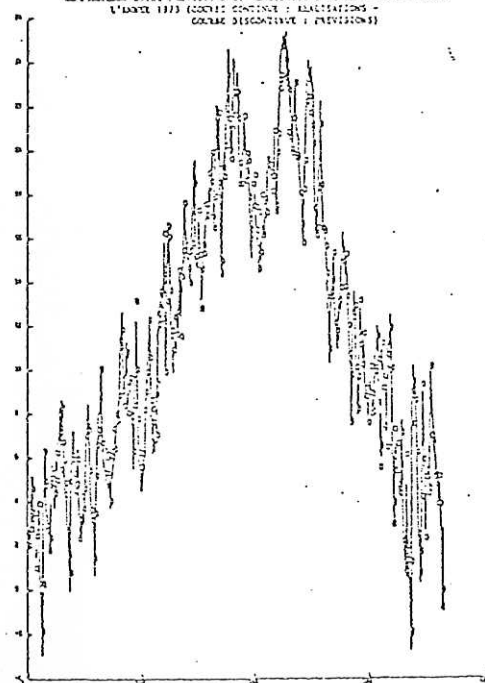


FIGURE 8
COMPARAISON ENTRE PRÉVISIONS ET RÉALISATIONS DE TEMPÉRATURES POUR
L'ANNÉE 1973 (COURBE CONTINUE : RÉALISATIONS -
COURBE DISCONTINUE : PRÉVISIONS)



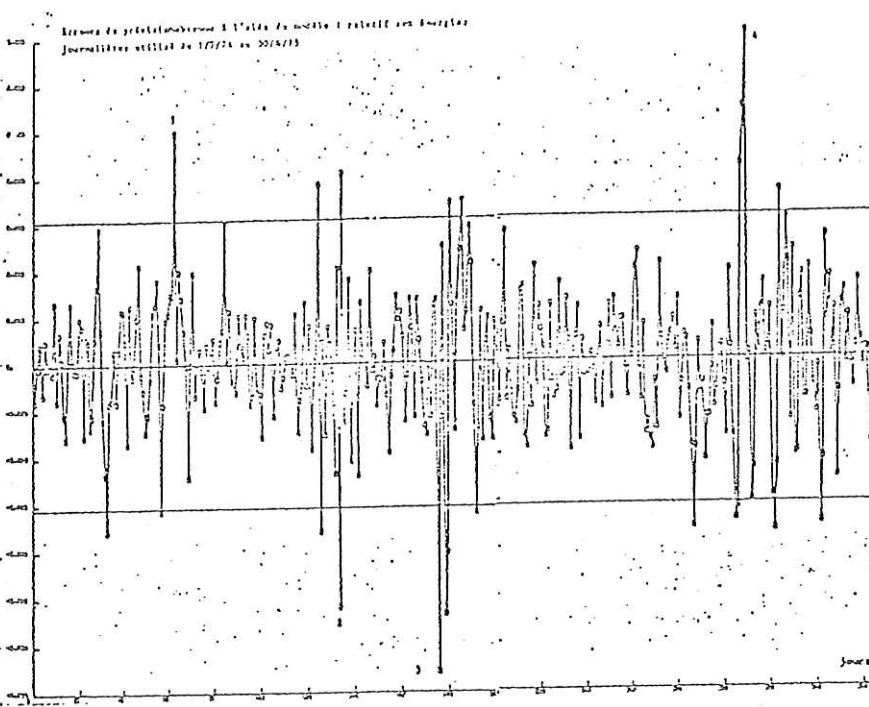


Figure 14

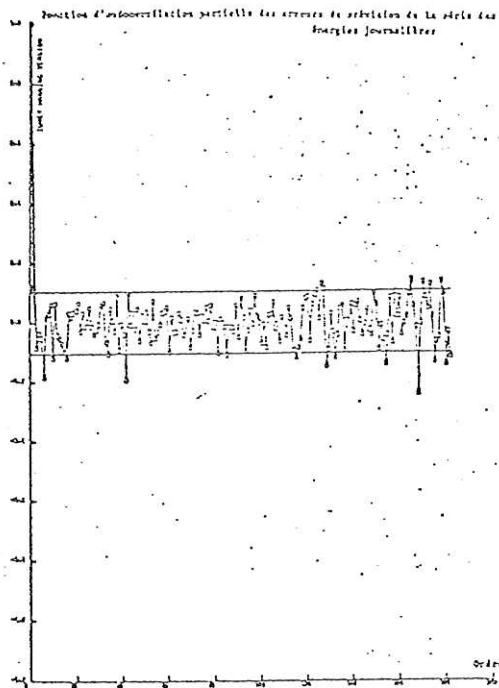


Figure 13

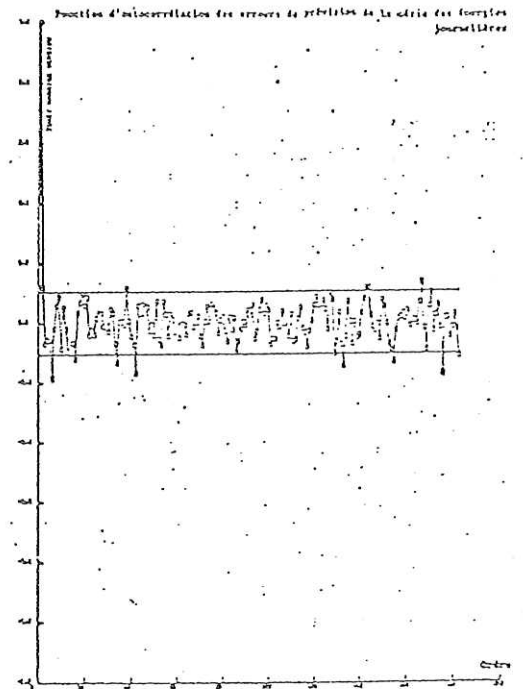


Figure 15

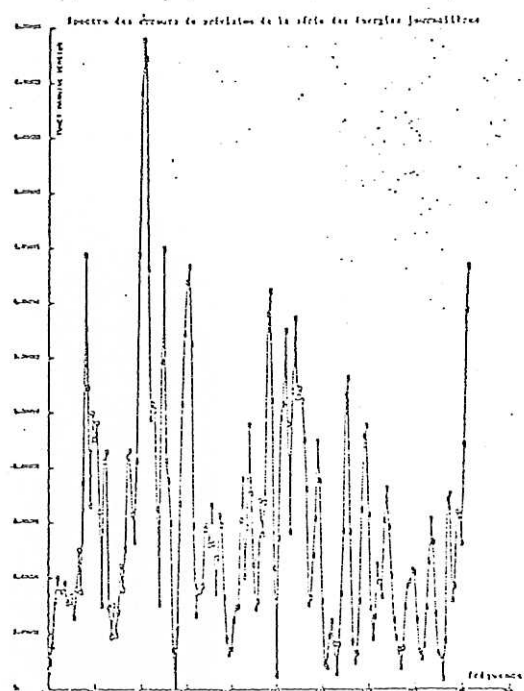


FIGURE 12

Erreurs de prévision de la série de temps à court terme
Journalières utilisées de 1972 à 1973

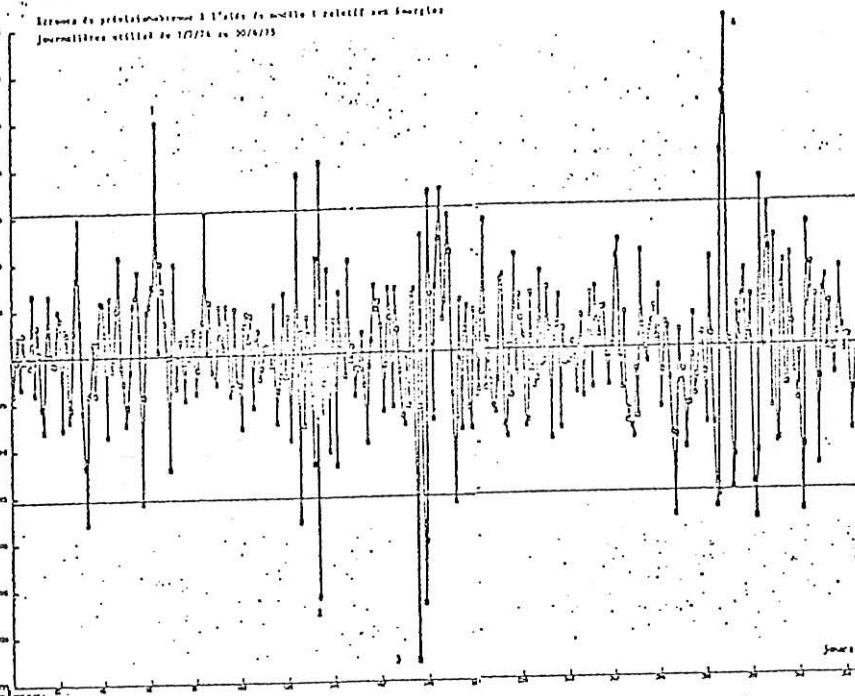


FIGURE 13

Fonction d'autocorrélation des erreurs de prévision de la série des énergies journalières

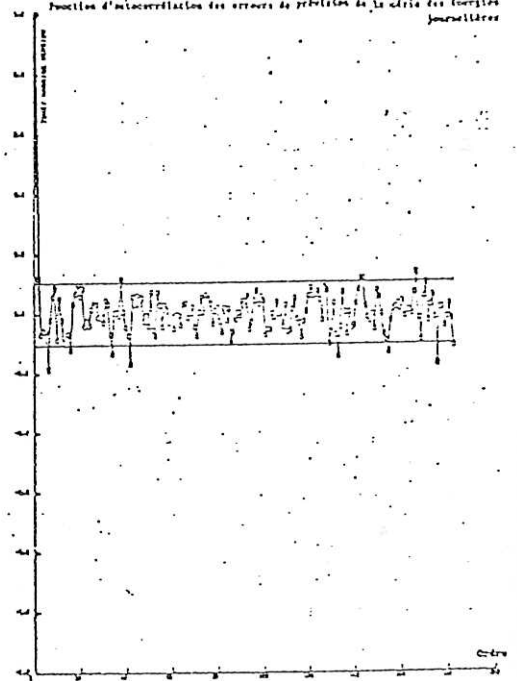


FIGURE 14

Fonction d'autocorrélation partielle des erreurs de la série des énergies journalières

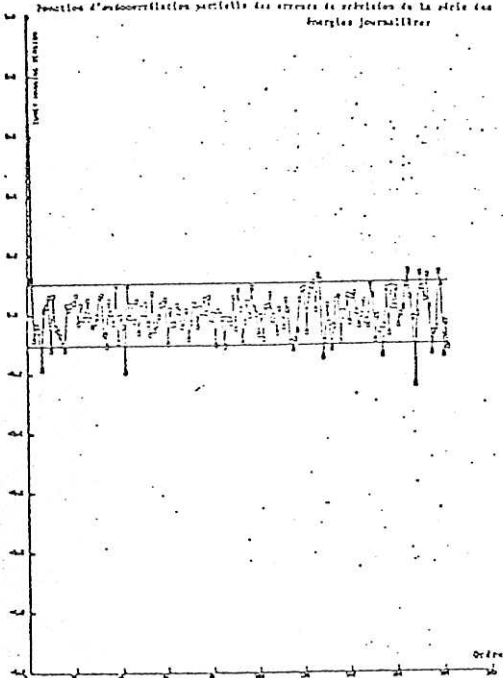
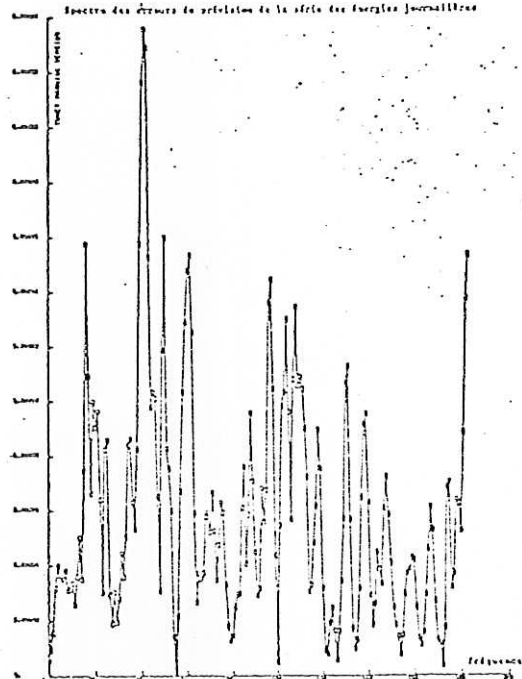
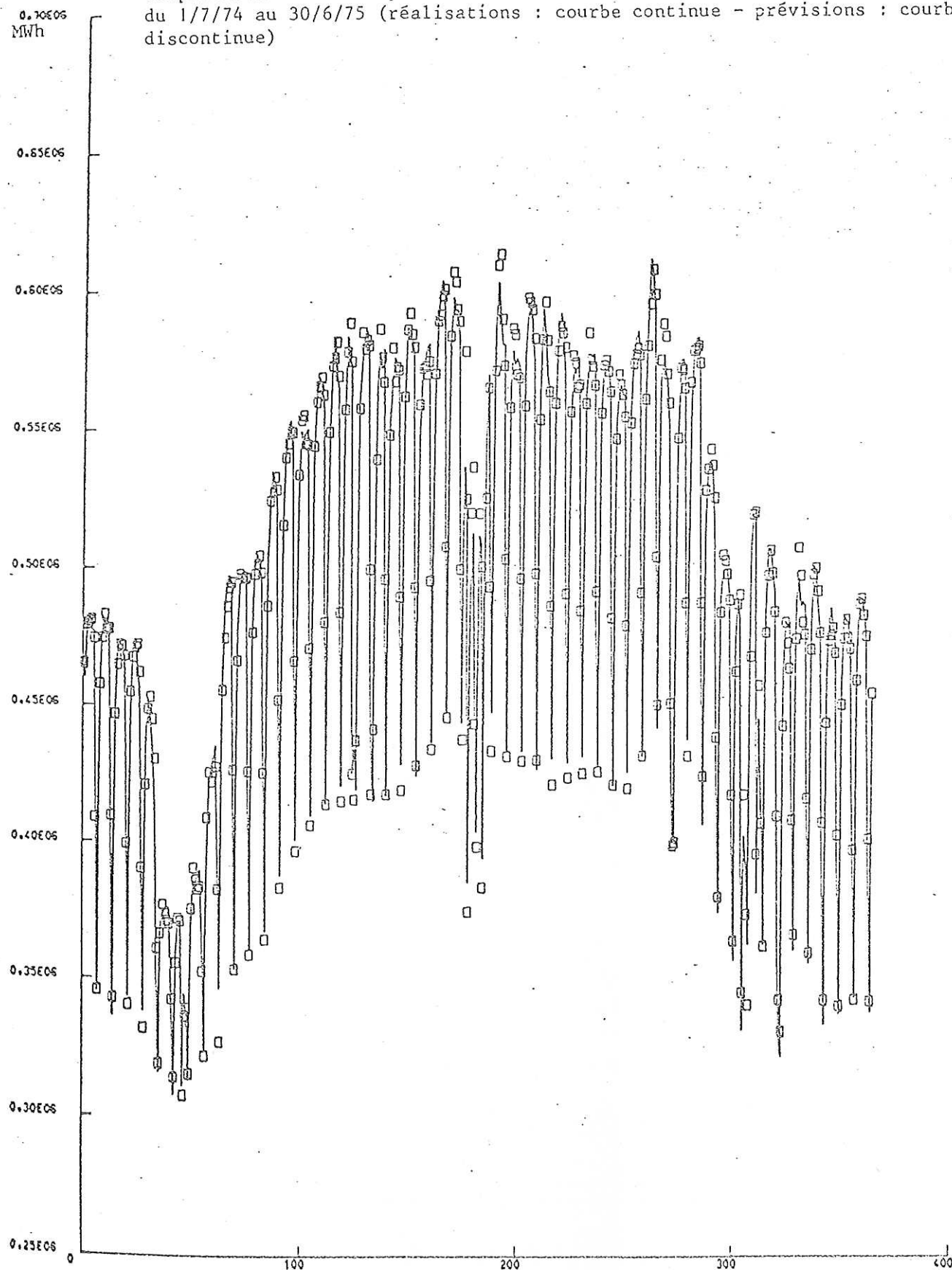


FIGURE 15

Spectre des erreurs de prévision de la série des énergies journalières



Comparaison entre les prévisions et les réalisations des énergies journalières du 1/7/74 au 30/6/75 (réalisations : courbe continue - prévisions : courbes discontinues)



Coefficients correcteurs des longueurs d'ondes

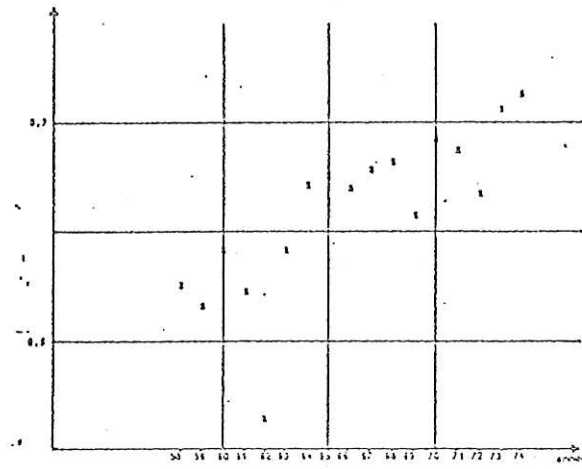


FIGURE 12

