



**HAL**  
open science

# Analyse des séries temporelles intervalles et prévision en économie de l'énergie

Hervé Tchouamani, Jules Sadefo-Kamdem

► **To cite this version:**

Hervé Tchouamani, Jules Sadefo-Kamdem. Analyse des séries temporelles intervalles et prévision en économie de l'énergie. SSRN Electronic Journal, 2014, 10.2139/ssrn.3523231 . hal-02939136

**HAL Id: hal-02939136**

**<https://hal.science/hal-02939136>**

Submitted on 15 Sep 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.

# Analyse des séries temporelles intervalles et prévision en économie de l'énergie

Hervé TCHOUMANI

FACULTÉ D'ÉCONOMIE

Université Montpellier

Jules SADEFO KAMDEM \*

UNIVERSITÉ DE LA GUYANE (DFR SJE)

Université de Montpellier (LAMETA CNRS UMR 5474)

## Résumé

Ce document de travail présente les méthodes d'estimation des paramètres d'un modèle de régression lorsque les variables sont observées sous un format intervalle. Il illustre cette procédure d'estimation sur les données de consommation d'électricité en France, afin d'effectuer une comparaison entre les valeurs prévisionnelles issues des données intervalles et l'intervalle de confiance de la prévision issu des données ponctuelles. Dans un premier temps, nous montrons que l'utilisation des valeurs uniques pour caractériser une variable économique entraîne une perte d'informations. Dans un deuxième temps, ces méthodes d'estimation ont été appliquées aux données intervalles de consommation d'électricité en France dans une optique de prévision. L'examen de l'évolution de nos données via des tests nous a permis de déceler un effet saisonnier, mais pas d'extra-saisonnier. Ainsi, les procédures d'estimation sont mises en sur les données désaisonnalisées par la méthode Census X12, et la prévision prendra en compte les coefficients saisonniers obtenus après désaisonnalisation. Pour mesurer l'intérêt de l'analyse sur les données intervalles, nous avons comparé la prévision sur les données intervalles, à l'intervalle de confiance de la prévision sur les données ponctuelles. Dans la dernière section, l'observation du profil des intervalles de confiance prévisionnels et celui des prévisions sur les données intervalles a révélé l'inclusion de l'intervalle de confiance de la prévision sur les données classiques, dans les bornes prévues par les données intervalles. Cette inclusion persiste avec l'augmentation du niveau de confiance accordé à l'IC de la prévision. Nous avons ainsi conclu sur l'impossibilité des données classiques à couvrir l'intervalle de la vraie valeur de la variable étudiée.

---

\*Corresponding author : LAMETA Université de Montpellier I - UFR d'Économie - Site de Richter, Avenue Raymond Dugrand CS 79606, 34960 Montpellier Cedex 2 France, email : [sadefo@lameta.univ-montp1.fr](mailto:sadefo@lameta.univ-montp1.fr), [jsadefo@gmail.com](mailto:jsadefo@gmail.com)

**Keywords :** Series Temporelles Intervalles, Commmation énergétique ;  
Prévision Economique.

**JEL Codes :** C12; C58; D52; G13; Q11; Q14

L'analyse des séries temporelles suppose le plus souvent que les variables mise en évidence sont représentées par **une série de réalisations uniques** ( $\{y_i\}$  pour  $i = 1, \dots, n$  ou  $\{y_t\}$  pour  $t = 1, \dots, T$ ). Or, ces données sont sous l'emprise de l'incertitude. C'est pourquoi, l'étude de la dynamique d'une variable par des données ponctuelles génère une perte d'information. L'idée de cette perte d'information est corroborée par la notion d'épaisseur du temps. De nos jours, avec l'avènement des systèmes d'informations sophistiqués, la collecte des données est devenue moins coûteuse. C'est ainsi que l'incertitude dans les données ponctuelles peut être atténuée par la prise en compte des données sous un format intervalle ( $[y_{min}, y_{max}]$ ). Elles seront représentées par la valeur minimale et maximale de la variable. C'est le cas en météorologie pour les données de température, en finance de marché pour les prix des actifs ou des indices boursiers, en médecine, lors de la mesure de la pression systolique et diastolique quotidienne, en énergie pour les données de consommation ou de production d'électricité,...etc. Dans la littérature, nous retrouvons plusieurs méthodes d'estimation des paramètres d'un modèle de régression lorsque les données sont sous un format intervalle.

Les données à valeurs intervalles sont aussi considérées comme des données symboliques<sup>1</sup>. Dans cette approche symbolique initiée par Billard et Diday (2003, 2006), il existe plusieurs propositions nécessaires pour estimer un modèle de régression sur les données intervalles dont celle de Arroyo, Gonzalez-Rivera, et Maté (2010). L'approche la plus simple de Billard et Diday (2000) permet d'estimer un modèle de régression sur le centre de l'intervalle des variables dépendantes et des variables explicatives. D'autres approches considèrent deux régressions séparées, une sur la borne inférieure de l'intervalle et l'autre sur la borne supérieure. Billard et Diday (2002) effectuent cette régression sans contraintes sur les coefficients et Brito (2007) impose aux deux régressions de partager les mêmes coefficients de régression. Lima Neto et De Carvalho (2008) proposent d'effectuer deux régressions différentes, une sur le centre de l'intervalle et l'autre sur son rayon sans aucune contrainte sur les paramètres. Parmi toutes ces approches, aucune ne garantit que les valeurs estimées des variables dépendantes satisfassent l'ordre naturel de l'intervalle c'est-à-dire,  $\hat{y}_{min} \leq \hat{y}_{max}$  pour toutes les observations. Lima Neto et De Carvalho (2010) imposent des contraintes de non-négativité des coefficients du modèle relatif au rayon de l'intervalle et résolvent par là, un problème de programmation quadratique pour trouver une solution pour les moindres carrés. Cependant, le fait d'imposer cette contrainte sur le modèle représenté par le rayon de l'intervalle, oblige la variable dépendante d'être positive ou nulle. Ce signe est nécessaire pour garantir l'ordre de l'intervalle mais **réduit l'utilité du modèle**. C'est pourquoi Gonzales-Riveira et Wei Lin (2013) vont publier un article dans

---

1. Les données sont symboliques quand dans chaque case du tableau de données symboliques, on ne trouve pas nécessairement une seule valeur qu'elle soit quantitative ou qualitative.

JBES<sup>2</sup> qui estime les paramètres d'un système bi-varié construit sur les bornes de l'intervalle en garantissant l'ordre naturel de l'intervalle ( $\hat{y}_{min} \leq \hat{y}_{max}$ ). L'utilisation de ces différentes techniques d'estimation ont permis plusieurs applications dont celles du domaine de l'énergie.

Les recherches menées pour améliorer les techniques de prévision se sont très souvent basées sur les données classiques (valeurs uniques). Les études étaient centrées sur les modèles statistiques utilisant par exemple les modèles ARIMA (Abdel-Aal et Al-Garni, 1997; Saab et al, 2001). Mais, on enregistrait de très fortes erreurs de prévision et ces erreurs étaient d'autant plus prononcées que les données étaient par essence recueillies sous un format intervalle, à l'instar de celles sur le marché de l'électricité. C'est la raison pour laquelle, dans le cadre des séries temporelles à valeurs intervalles, différentes techniques de prévision ont été développées ces dernières années. Arroyo et al. (2007) développent trois méthodes de lissage exponentiel pour les séries temporelles à valeurs intervalles. Maia et al. (2008) présentent une approche de prévision des données intervalles basée sur les modèles AR, ARMA et les réseaux de neurones artificiels. Cheung (2007) propose un modèle empirique pour les valeurs minimales et maximales pour trois indices américains utilisant un VECM<sup>3</sup>. Garcia C. et Maté C. (2010) présentent une comparaison entre une nouvelle approche de prévision considérant un modèle vecteur auto régressif (VAR) appliqué aux séries temporelles à valeurs intervalles et un modèle de perceptron multicouches adapté aux données intervalles.

Notre objectif dans ce papier est montré que dans le domaine de l'énergie, les prévisions de la consommation d'électricité sur les données intervalles seraient meilleures que celles sur les données ponctuelles. Autrement dit, nous voulons effectuer une comparaison entre les prévisions sur données intervalles et l'intervalle de confiance de la prévision sur données ponctuelles à différents niveau de confiance.

Pour cela, nous allons repartir notre analyse en quatre sections; la première section sera consacré à la présentation de l'intérêt des données intervalles en économie, la seconde section à l'estimation des paramètres sur valeurs intervalles, la troisième section va considérer les valeurs ponctuelles. On va effectuer la prévision à l'aide des séries de Fourier et enfin, la dernière section nous permettra d'effectuer la comparaison des résultats des différentes prévisions.

---

2. Journal of Business and Economic Statistics

3. vector error correction model

# 1 Intérêt des données intervalles en économie

Dans les sciences sociales, et plus particulièrement en économie, les phénomènes étudiés concernent le plus souvent des comportements (comportement de consommation des ménages, d'investissement et de production des entreprises, de dépense de l'État etc.) afin de mieux appréhender la nature et le fonctionnement des systèmes économiques. A cet effet, les économistes et/ou les économètres effectuent leurs analyses sur des données recueillies le plus souvent sous forme de séries de valeurs uniques. Par exemple, ils vont récolter une suite d'observations sur le revenu des ménages selon leurs catégories socio-professionnelles, sur leur niveau de consommation, sur l'investissement des entreprises, sur le taux de change, sur le prix des biens ou des actifs, etc. Or, la nature humaine rend les comportements humains difficilement compréhensibles. c'est pourquoi, il est limité de saisir par exemple le comportement de consommation des ménages par l'observation de leur niveau de consommation par une valeur unique. Ainsi, avec l'avènement des systèmes d'information sophistiqués, la collecte des données est devenue moins coûteuse. Nous pouvons nous rapprocher davantage de la réalité en étudiant certaines variables économiques sur des données intervalles. L'analyse sur les valeurs intervalles suppose que les observations et les estimations d'une variable économique sur des valeurs uniques soient généralement incomplètes ou incertaines, voire entachées d'erreurs. Comme corollaire à cette hypothèse, elles ne représentent pas précisément les données réelles. Si nous recherchons la précision dans notre analyse, les données doivent être représentées comme des intervalles renfermant quasiment toutes les réalisations d'une série chronologique. Cependant, nous savons que la macroéconomie est la branche de l'économie qui étudie les agrégats macroéconomiques. Ainsi, on ne s'intéresse pas dans cette branche à l'étude des unités individuelles mais plutôt des unités collectives, et comme les données collectives sont obtenues par agrégation des unités individuelles, il est évident qu'il y ait une perte d'informations dans ces données agrégées. C'est pourquoi, il n'est pas intéressant pour le statisticien économiste d'analyser les données constituées de valeurs ponctuelles au détriment des observations au format intervalle. Plusieurs institutions collectent et publient ces masses d'informations sous forme agrégée.

Le bureau du recensement aux États-Unis est un organisme du système statistique fédérale américain, responsable de la production des informations socio-économiques sur tous les individus sur le territoire. Ces informations de masse ont tendance à être publiées sous une forme agrégée, soit pour des raisons de confidentialité, ou parce que l'axe d'orientation de l'étude ne concerne pas les unités individuelles; mais les unités collectives. Dans ce cas, il est intéressant de disposer des données intervalles pour représenter au mieux ces informations. En France, l'INSEE (l'institut nationale de la statistique et des études économiques) recueille et publie les informations sur l'économie Française. Elle publie par exemple l'indice des prix des logements neufs ou anciens mais, comme les prix de l'immobilier varient fortement d'une région à l'autre, et même au sein d'une

région, d'un département à l'autre, cette forte fluctuation du prix de l'immobilier nécessiterait la prise en compte du prix minimal et maximal pour effectuer des prévisions de l'évolution des prix futurs. De même, elle publie également le taux de chômage par région et aussi pour chaque département. Cette instabilité du taux de chômage conduit au fait qu'on gagnerait à considérer le taux de chômage minimal et maximal pour prévoir l'évolution du taux de chômage. On trouverait ainsi autant d'exemple dans lequel il serait nécessaire de considérer les données intervalles.

Le modèle d'équilibre général de Walras est un modèle statique. Les notions de dynamique et d'incertitude ont été ajoutées pour avoir le modèle de Walras-Arrow-Debreu (WAD) qui repose sur les échanges de contrats à terme (Ferrari, 2006). Dans ce modèle, les agents s'engagent à exécuter le contrat si et seulement si l'événement pour lequel il a été conclu se réalise. L'incertitude est représentée par l'absence de connaissance à la date initiale, sur la réalisation de l'événement possible de la date finale. La réalisation de l'événement à la date finale conditionne la réalisation de la variable économique qui est mise en évidence (le prix, les rendements financiers ...). La quantité d'information dont dispose les individus sur la réalisation des états de la nature augmente au fur et à mesure qu'ils se rapprochent de la date finale. or, si on modifie la psychologie des individus en supposant qu'ils raisonnent non plus sur la réalisation d'un état terminal mais plutôt sur un ensemble d'états terminaux, ils disposeraient dès la date initiale de plus d'informations sur la réalisation de la variable économique qu'ils n'en disposeraient en raisonnant sur un vecteur d'états uniques. Donc, en raisonnant sur les intervalles, on arrive à réduire l'incertitude probabilisable qui est une problématique primordiale du modèle WAD et ainsi, améliorer la prise de décision des agents.

la raison la plus importante de l'intérêt que nous portons dans l'estimation et la prévision avec les données à valeurs intervalles s'explique par le fait que **le seul format disponible pour certaines bases de données est le format intervalle**. Les bases de données financières sont une exception. Ce format est le seul mis à la disposition des autorités pour recueillir les données de consommation, de production, de prix de l'électricité. En France, le Réseau de Transport d'électricité (RTE) publie les informations sur les valeurs minimales et maximales de la consommation d'électricité toutes les trente minutes. Aux États-Unis, l'administration américaine de l'information pour l'énergie (EIA<sup>4</sup>) rassemble le prix de l'électricité pour chaque État. Comme le prix de l'électricité est influencé par plusieurs facteurs, on enregistre des variations importantes du prix à travers les États et à travers les localités dans les mêmes États. L'EIA fournit les prix au niveau de chaque État dans un format intervalle, c'est-à-dire le prix minimal et maximal, ce qui permet de recueillir le maximum d'information sur les réalités de ce marché. Le département de l'agriculture aux États-Unis fournit les prix du bétail également en format intervalle. Le centre d'information de la commerciali-

---

4. Energy Information Administration

sation du bétail(LMIC<sup>5</sup>) déclare l'intervalle des prix de plusieurs animaux, par exemple, le minimum et le maximum du prix du bœuf par jour. En gestion des stocks, la description d'une situation de rupture de stock se fait de la façon suivante : Niveau de stock = [100, 150], Quantité en cours de commande = [50, 100], Durée de livraison = [30, 45]. Eu égard à ces observations, nous pouvons souligner que c'est **l'incertitude dans la conjoncture économique qui conduit les entreprises à considérer le format intervalle dans le processus de prise de décision**. Ainsi, lorsqu'on recueille les données sous un format intervalle et qu'on transforme ces données en valeurs uniques, on crée inéluctablement quelque chose dans ces données transformées qui se répercute immédiatement dans les résultats des analyses issus de la nouvelle base de données. Ces exemples montrent que la représentation des observations d'une variable économique sous un format intervalle fournit des informations supplémentaires au delà d'une mesure moyenne et dans certains cas, c'est le seul format disponible pour les chercheurs. Nous pouvons illustrer nos propos en considérant les représentations graphiques ci-dessous.

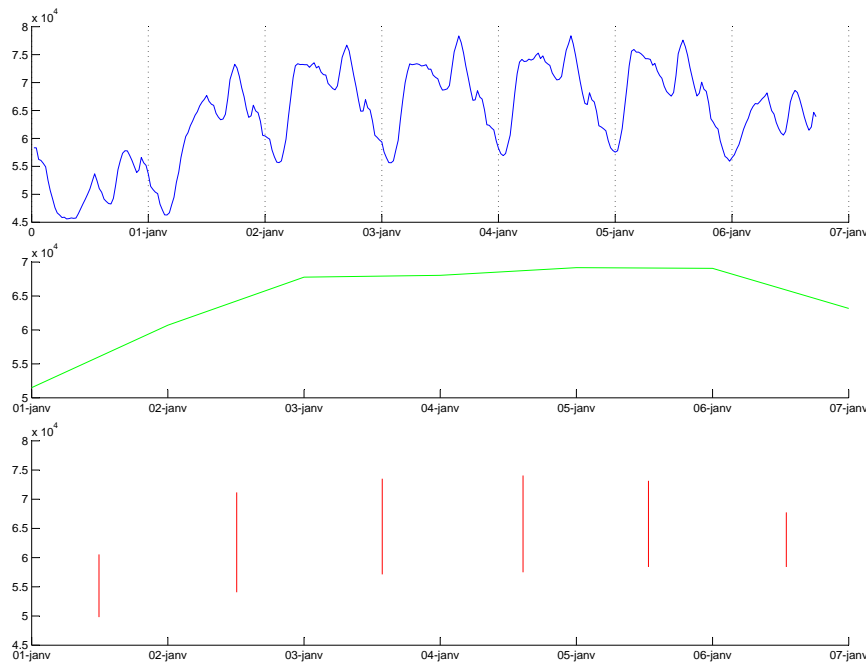


FIGURE 1: Évolution intra-journalière de la consommation intérieure d'électricité en France(01/01/2012 - 07/01/2012), Série temporelle de la consommation moyenne et enfin, amplitude des intervalles de consommation journalière.

Le premier graphique représente la consommation journalière d'électricité toutes les 30 minutes. Le seconde retrace le profil de la consommation moyenne et le dernier nous indique l'amplitude de la consommation journalière.

En économétrie classique, les analyses s'effectueront en considérant les va-

5. Livestock Marketing Information Center

leurs moyennes de la consommation or, l'observation du premier graphique laisse apparaître de fortes variabilités de la consommation d'électricité au sein d'une même journée. Nous remarquons aussi que le profile de consommation semble être presque identique d'un jour à l'autre. Ce comportement indique l'existence d'un effet de mémoire dans les habitudes de consommation. En plus, nous observons un pic de consommation entre 18 et 21 heures, révélateur d'un effet saisonnier dans la consommation journalière. Le dernier graphique montre qu'au sein d'une même journée, l'écart entre la consommation maximale et minimale est très important. L'observation des valeurs uniques (valeurs moyennes) ne peut en aucun cas révéler ces informations sur l'évolution de la consommation d'électricité alors qu'elles sont indispensables, primordiales, vitales lors de la prise de décision.

## 2 Analyse des données

La mise en œuvre des techniques d'estimation se fera sur les données mensuelles de consommation d'électricité en France publiées par le Réseau de Transport d'électricité (RTE) entre 01/2008 et 12/2013<sup>6</sup>. La recherche des différentes composantes contenues dans nos données sera indispensable lors de l'estimation des paramètres.

### Analyse graphique et de la variance des séries de consommation

Dans le graphique suivant, nous représentons les valeurs maximales et minimales de la consommation intérieure d'électricité en France.

L'observation de l'évolution du premier graphique montre que les données intervalles de la consommation d'électricité sont affectées d'une saisonnalité à priori **rigide**. Ce graphique laisse apparaître un pic de consommation sur le trimestre constitué des mois de Décembre, Janvier et Février. Cette forte croissance de la consommation s'explique tout simplement par la présence de la période hivernale. Nous pouvons noter également un creux dans la consommation relatif à la période estivale.

Le second graphique représente les amplitudes des intervalles de consommation. L'écart de consommation à l'approche et pendant la période hivernale est très marqué. Cependant, La période estivale est caractérisée par de faibles écarts. En Février 2012, nous observons un grand pic et un écart de consommation assez élevés; C'est l'illustration de la période de grand froid qui a incité Météo-France à activer sa carte de vigilance au niveau orange pour la première fois depuis la mise en place de ce dispositif en Novembre 2004. Toutefois, la seule observation de ces deux graphiques ne nous permet pas de nous prononcer sur l'existence d'un extra-saisonnier. L'analyse de la variance et le test de Fisher nous permettra d'analyser plus finement ces données.

---

6. Tous les calculs sont effectués sur 5 ans, l'année 2013 représente notre échantillon de prévision



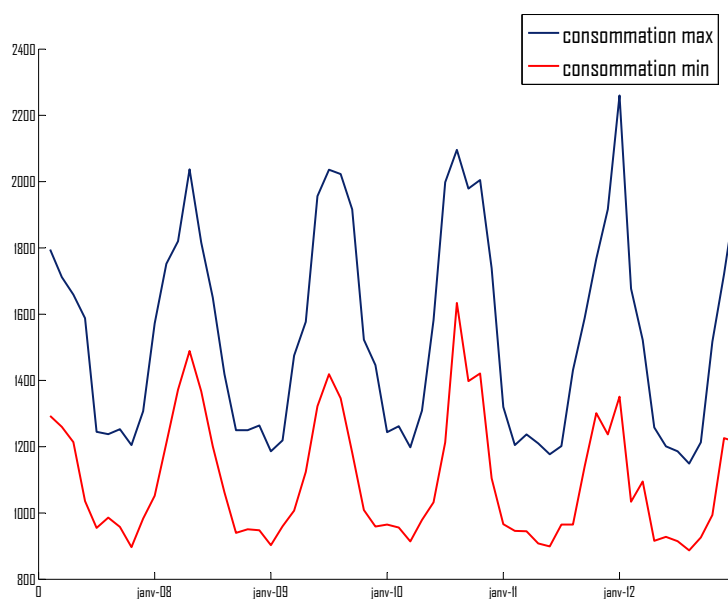


FIGURE 2: Évolution de la consommation intérieure maximale et minimale d'électricité en France

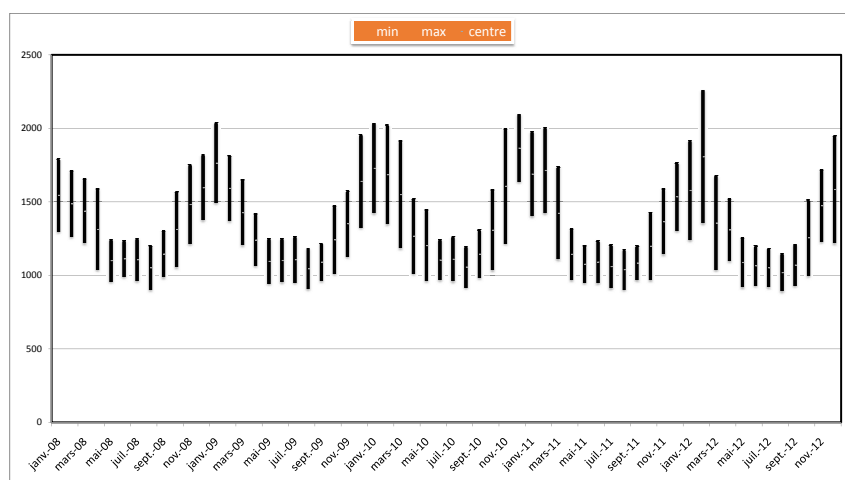


FIGURE 3: Évolution de l'amplitude des intervalles de consommation d'électricité en France

A partir du tableau de buys ballot, nous avons établi le tableau d'analyse de la variance permettant d'effectuer le test de Fisher. Pour le test d'influence du facteur colonne, la valeur calculée du Fisher est de 3.58 pour le min et de 3.60 pour le max. Ces valeurs sont supérieures à la valeur lue dans la table qui est de 2. Ainsi, les séries de consommation maximale et minimale sont affectées d'une

### saisonnalité.

Il ressort du test d'influence du facteur année que nos séries ne sont pas affectées d'un extra-saisonnier car ; le Fisher empirique est de 0.2 pour le min et de 0.32 pour le max tant disque le Fisher lu est de 2.61, alors nous acceptons l'hypothèse nulle d'absence d'influence du facteur année.

### Spectres des chroniques

L'analyse spectrale nous permet de décomposer la structure de la série chronologique dans le domaine des fréquences. Le spectre d'un signal ou d'une chronique est la représentation en fonction de la fréquence des amplitudes des différentes composantes présentes dans le signal. Les séries temporelles de consommation pour les deux bornes de l'intervalle évoluent de manière cyclique, nous pouvons ainsi les décomposer comme une somme de fonction trigonométrique. Cette modélisation nous permettra de construire le spectre de chaque signal sur lequel nous détecterons les composantes de la chronique. Les spectres des valeurs maximales et minimales sont représentés sur les graphiques suivants :

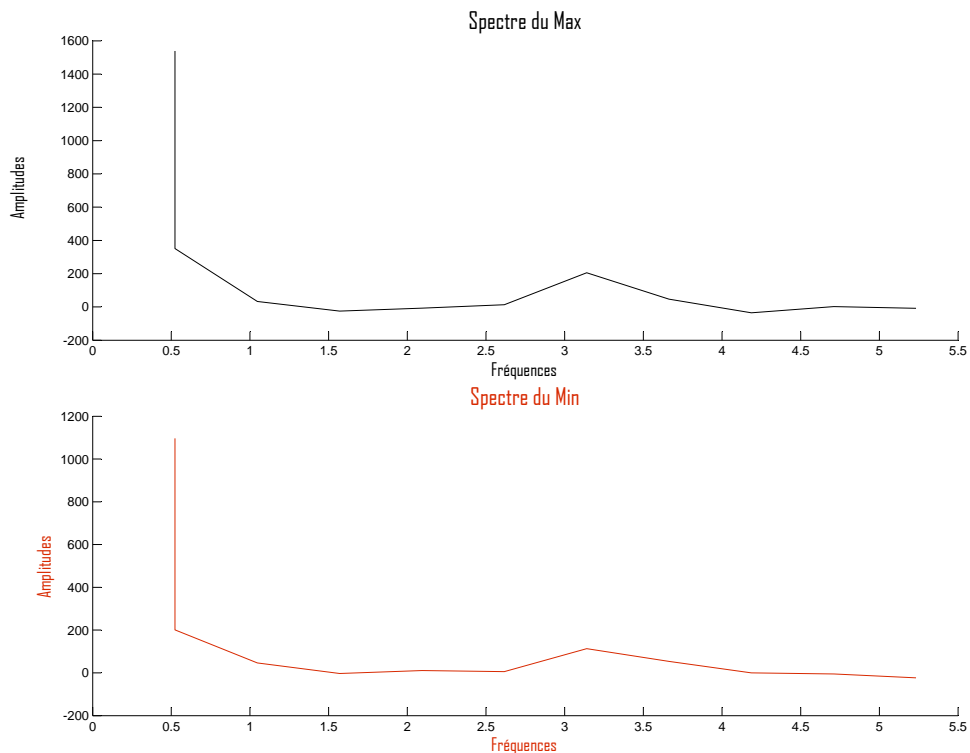


FIGURE 4: Représentation des spectres des valeurs minimales et maximales de la consommation

Ces deux spectres sont identiques compte tenu de la forte corrélation révélée par l'examen visuel de nos séries sur la figure (3). La lecture du spectre laisse apparaître un pic à la fréquence fondamentale qui se suit d'une décroissance aux multiples de cette fréquence. Nous observons un second pic à la fréquence  $\pi$ . Ces

deux spectres mettent donc en évidence une double saisonnalité en dents de scie. Une première conclusion nous indique que nos données intervalles sont affectées d'un effet saisonnier et pas d'une tendance. Toutefois, le test de Dicker-Fuller Augmentés nous permettra de nous prononcer sur l'existence ou non d'une racine unitaire.

### Corrélogramme des séries

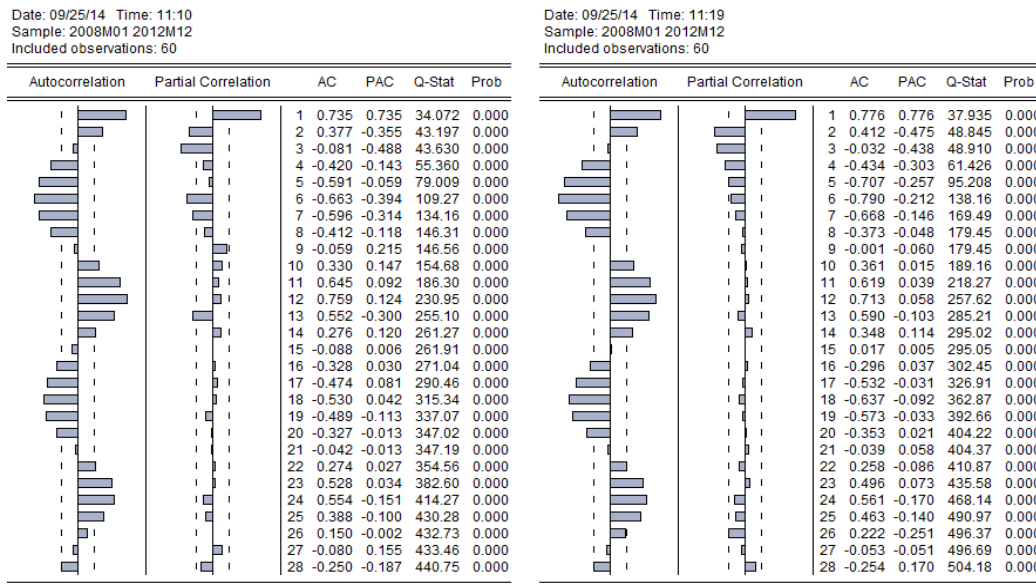


FIGURE 5: Corrélogramme de la consommation minimale et maximale respectivement

Les termes de la fonction d'autocorrélation s'alternent. Ce qui permet de caractériser l'existence d'une saisonnalité. Les probabilités critiques de la statistique de Ljung-Box sont tous nulles. Alors, il existe au moins un coefficient d'autocorrélation significativement différent de zéro.

### Test de Racine unitaire de Dicker-Fuller Augmentés(ADF)

Pour chaque borne de l'intervalle, nous réalisons le test ADF avec le critère de Schwarz comme critère d'information. La statistique ADF à 1, 5 et 10% est inférieure aux différentes valeurs de Mac Kinnon. Par conséquent, nous rejetons l'hypothèse nulle de présence d'une racine unitaire.

Avant l'application des méthodes d'estimation sur les données intervalles de consommation, la désaisonnalisation des chroniques est nécessaire afin de récu-

pérer les coefficients saisonniers indispensables à la prévision.

La désaisonnalisation des séries se fera à l'aide de la méthode Census X12. L'application de cette méthode s'est faite à l'aide du logiciel EVIEWS selon un schéma additif<sup>7</sup>. Les séries de consommation d'électricité corrigées des variations saisonnières sont présentées sur le graphique suivant :

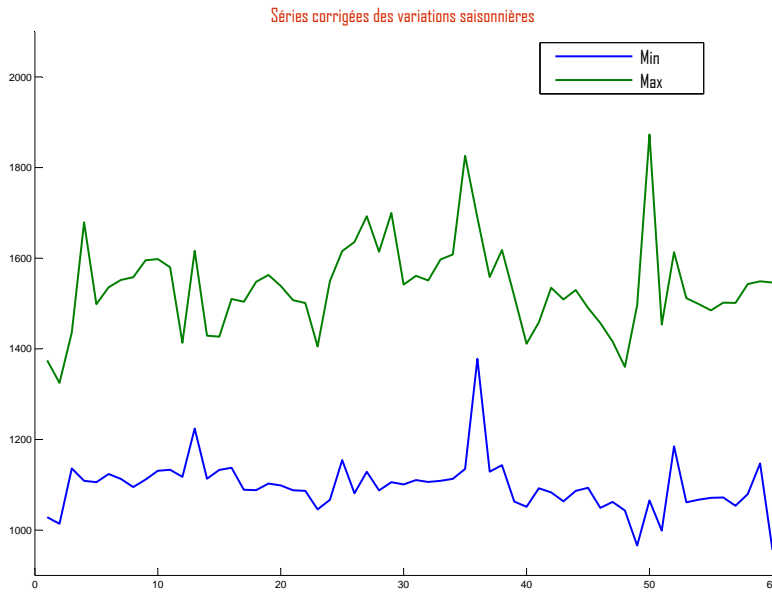


FIGURE 6: Séries de consommation d'électricité corrigées des variations saisonnières

Nous remarquons que ces deux séries semblent stationnaires. Concernant la limite supérieure de l'intervalle, nous enregistrons une valeur moyenne de 1538 avec un écart type de 99. La limite inférieure possède une valeur moyenne de 1096 avec une dispersion autour de cette moyenne de 58. Eu égard à ces résultats, nous pouvons conclure sur une faible volatilité des séries de consommation corrigées des variations saisonnières. Ce constat est très important pour la suite en ce sens que, la probabilité que les valeurs estimées des bornes de l'intervalle de consommation se croisent est très faible. Ainsi, **la restriction imposée sur les observations ne sera pas obligatoire**, la procédure d'estimation va en conséquence prendre en compte cette information.

### 3 Les Différentes étapes de la procédure d'estimation

Cette procédure a été développée par Gonzalez Riveira et Wei Lin (2013). Elle consiste à construire un système bi-varié sur les limites supérieures et inférieures

7. Le tableau de Buys Ballot nous a permis d'effectuer le test de schéma sous Excel

de l'intervalle. Le modèle auto régressif à valeur intervalle IAR<sup>8</sup> à p retards peut s'écrire de la manière suivante :

$$\begin{bmatrix} C_{bt} \\ C_{ht} \end{bmatrix} = \begin{bmatrix} \beta_{bc} \\ \beta_{hc} \end{bmatrix} + \sum_{j=1}^p \begin{bmatrix} \beta_{11}^{(j)} & \beta_{12}^{(j)} \\ \beta_{21}^{(j)} & \beta_{22}^{(j)} \end{bmatrix} \begin{bmatrix} C_{b,t-j} \\ C_{h,t-j} \end{bmatrix} + \begin{bmatrix} \varepsilon_{bt} \\ \varepsilon_{ht} \end{bmatrix} \text{ tel que } C_{b,t} \leq C_{h,t}$$

Nous partons de ce système IAR(p) sans restriction et nous sélectionnons le nombre de retard optimal par minimisation du critère de Schwarz. Nous obtenons deux retards. Ce qui nous conduit à décrire la dynamique de  $C_{b,t}$  et  $C_{h,t}$  de la façon suivante :

$$\begin{aligned} C_{b,t} &= \beta_{b,c} + \beta_{11}^1 C_{b,t-1} + \beta_{12}^1 C_{h,t-1} + \beta_{11}^2 C_{b,t-2} + \beta_{12}^2 C_{h,t-2} + \varepsilon_{b,t} \\ C_{h,t} &= \beta_{h,c} + \beta_{21}^1 C_{b,t-1} + \beta_{22}^1 C_{h,t-1} + \beta_{21}^2 C_{b,t-2} + \beta_{22}^2 C_{h,t-2} + \varepsilon_{h,t} \\ &\text{tel que } C_{b,t} \leq C_{h,t} \end{aligned} \quad (1)$$

### Première étape

Nous allons modéliser le rayon de l'intervalle  $\Delta C_t = C_{h,t} - C_{b,t}$  du système (1) et sachant que l'erreur suit une densité normale tronquée, nous pouvons maximiser la fonction log-vraisemblance qui se présente de la façon suivante :

$$T^{-1}L(\Delta C; \Delta\beta, \sigma_m) = \frac{1}{T} \sum_{t=1}^T \log f(\Delta C_t | \Delta C_t \geq 0, C^{t-1}; \Delta\beta, \sigma_m) \quad (2)$$

La résolution de ce programme d'optimisation permet de trouver les valeurs de  $\Delta\beta$  et de  $\sigma_m$ . Ces résultats nous permettront de déterminer l'inverse du ratio de Mills.

$$\lambda_{t-1} = \frac{\phi(\Delta(C^{t-1}, \Delta\beta)/\sigma_m)}{1 - \Phi(\Delta(C^{t-1}, \Delta\beta)/\sigma_m)} \quad (3)$$

où  $\phi$  et  $\Phi$  représentent respectivement la fonction densité et de répartition de la loi normale.

### Deuxième étape

Nous allons utiliser l'inverse du ratio de Mills dans l'équation de régression (1) pour corriger le biais de sélection imposé par la restriction  $C_{b,t} \leq C_{h,t}$ . Le modèle suivant sera estimé par MCO.

$$\begin{aligned} C_{b,t} &= \beta_{b,c} + \beta_{11}^1 C_{b,t-1} + \beta_{12}^1 C_{h,t-1} + \beta_{11}^2 C_{b,t-2} + \beta_{12}^2 C_{h,t-2} + M_b \hat{\lambda}_{t-1} + u_{b,t} \\ C_{h,t} &= \beta_{h,c} + \beta_{21}^1 C_{b,t-1} + \beta_{22}^1 C_{h,t-1} + \beta_{21}^2 C_{b,t-2} + \beta_{22}^2 C_{h,t-2} + M_h \hat{\lambda}_{t-1} + u_{h,t} \end{aligned} \quad (4)$$

Cette étape n'est nécessaire que si la restriction sur les variables dépendantes sont obligatoire. En cas de multi-colinéarité, notamment lorsque la restriction

---

8. interval autoregressive model. Dans ce modèle b signifie bas et h signifie haut

n'est pas obligatoire,  $\hat{\lambda}_{t-1}$  tend vers zéro et il n'est possible d'identifier les paramètres  $M_b$  et  $M_h$ . D'où la deuxième étape modifiée.

### Deuxième étape modifiée

l'identification de  $M_b$  et  $M_h$  s'effectuent grâce à l'**estimateur des distances minimums**. Le programme d'optimisation s'écrit :

$$(\tilde{M}_b, \tilde{M}_h) = \arg \min_{(M_b, M_h)} \{M_h + M_b - [\sigma_h^2(M_h) - \sigma_b^2(M_b)] / \hat{\sigma}_m\}^2, \quad (5)$$

tel que  $M_h - M_b = \hat{\sigma}_m$

La mise en œuvre de ce programme suit une procédure se déroulant en cinq étapes. Les valeurs estimées de  $M_b$  et  $M_h$  vont conduire à l'estimation des paramètres  $\beta$  du modèle (4) par l'application des MCO modifiées de la façon suivante :

$$\tilde{\beta}_b = \beta_b(\tilde{M}_b) = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{C}_b - \tilde{M}_b\hat{\Lambda})$$

$$\tilde{\beta}_h = \beta_h(\tilde{M}_h) = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{C}_h - \tilde{M}_h\hat{\Lambda})$$

Pour plus de détail sur cette procédure d'estimation, voir Gonzalez R. et Lin W. (2013).

## 4 La procédure d'estimation des paramètres

La maximisation de la fonction log-vraisemblance basée sur la densité normale tronquée nous permet d'estimer  $\hat{\theta}^* = \frac{\Delta\beta}{\sigma_m}$ . Les résultats sont consignés dans le tableau suivant :

TABLE 1: Première étape d'estimation

Régression normale tronquée	
Variables	Coefficients( $\Delta\beta^*$ )
C	-7.2475
$C_{min,t-1}$	0.0039
$C_{max,t-1}$	-0.0024
$C_{min,t-2}$	0.0011
$C_{max,t-2}$	0.0002
$\sigma_m$	86.9942
Période : 01/2008 12/2012	
Nombre d'observations : 60	

Nous déterminons ainsi l'inverse du ratio de Mills qui caractérise la sévérité de la restriction sur les observations. Les valeurs de  $\hat{\lambda}_t$  se situent entre 0 et 0.000038 ; avec une moyenne de 0.0000019 et un écart type de 0.0000051. Ces statistiques indiquent que  $\hat{\lambda}_t$  est pratiquement égale à zéro. Alors, **la restriction sur les observations n'est pas obligatoire**. C'est la raison pour laquelle, la deuxième étape d'estimation n'est pas nécessaire. Les résultats de la deuxième étape modifiée ainsi que ceux des MCO sont répertoriés dans le tableau ci-dessous :

TABLE 2: Les résultats des différentes étapes de la régression sur la borne supérieure

Variables	MCO		Deux Étapes Modifiées	
	Coefficients	écarts types	Coefficients	écarts types
Constante	1222.8	(321.81)	1222.8	(321.81)
$C_{min,t-1}$	-0.3232	(0.2491)	-0.3232	(0.2491)
$C_{max,t-1}$	0.2982	(0.1421)	0.2982	(0.1421)
$C_{min,t-2}$	0.0852	(0.2492)	0.0852	(0.2492)
$C_{max,t-2}$	0.0810	(0.1421)	0.0810	(0.1421)
$\hat{\lambda}_t$			86.99	
Degré liberté	55		55	
écart type erreur	90.1		90.1	
$R^2$ Ajusté	1		1	
Fischer Stat	3 292.2		3292.2	

TABLE 3: Les résultats des différentes étapes de la régression sur la borne inférieure

Variables	MCO		Deux Étapes Modifiées	
	Coefficients	écarts types	Coefficients	écarts types
Constante	592.30	(192.92)	592.30	(192.92)
$C_{min,t-1}$	0.0124	(0.1493)	0.0124	(0.1493)
$C_{max,t-1}$	0.0878	(0.0852)	0.0878	(0.0852)
$C_{min,t-2}$	0.1834	(0.1494)	0.1834	(0.1494)
$C_{max,t-2}$	0.1016	(0.0852)	0.1016	(0.0852)
$\hat{\lambda}_t$			0.0000	
Degré liberté	55		55	
écart type erreur	54		54	
$R^2$ Ajusté	1		1	
Fischer Stat	4 634.5		4 634.5	

Ces résultats sont identiques parce que la probabilité que les bornes de l'intervalle se croisent est très faible. Cependant, le seul estimateur des MCO suffit pour garantir la consistance des estimateurs. La prévision de la consommation est effectuée à l'aide du modèle estimé par MCO en rajoutant les coefficients saisonniers. Les valeurs prévisionnelles pour l'année 2013 sont présentées dans le tableau ci-après :

TABLE 4: Les valeurs observées et prévisionnelles de la consommation de l'année 2013 en GWh

Mois	valeurs observées	valeurs estimées(min,max)
Janvier	[1237-1917]	[1335-1916]
Février	[1351-2260]	[1354-1943]
MARS	[1034-1677]	[1131-1863]
avril	[1095-1522]	[1026-1485]
Mai	[916-1258]	[935-1270]
Juin	[928-1201]	[980-1264]
Juillet	[915-1186]	[929-1239]
Août	[887-1149]	[980-1179]
Septembre	[926-1213]	[958-1248]
Octobre	[994-1517]	[1000-1517]
Novembre	[1226-1721]	[1166-1717]
Décembre	[1215-1953]	[1357-1938]

GWh : million de kilowatt heure

#### 4.1 Mesure de performance de la prévision

Arroyo et Maté (2006) proposent des mesures basées sur le calcul des distances pour les données intervalles telle que la distance de Ichino-Yaguchi et celle de Hausdorff. La distance de Hausdorff a aussi été proposée par Chavent et Saracco (2008).

Appelons  $[x_t]$  les valeurs observées en intervalle et  $[\hat{x}_t]$  la prévision de cet intervalle avec  $t = 1, 2, \dots, n$ . La moyenne des distances des erreurs dans la logique de calcul de la distance de Hausdorff est définie par :

$$MDE_H = \frac{1}{n} \sum_{t=1}^n [|x_{ct} - \hat{x}_{ct}| + |x_{rt} - \hat{x}_{rt}|] \quad (6)$$

Celle basée sur la distance de Ichino-Yaguchi est donnée par :

$$MDE_{IY} = \frac{1}{n} \sum_{t=1}^n 0.5 [|x_{lt} - \hat{x}_{lt}| + |x_{ut} - \hat{x}_{ut}|] \quad (7)$$

Cependant, le fait que ces méthodes se concentrent sur le centre et le rayon ou sur la borne inférieure et supérieure de l'intervalle les rendent sensibles aux effets d'échelle qui sont susceptibles d'être rencontrés dans les données. Ainsi, afin de pouvoir analyser les performances prédictives du modèle pour les deux systèmes bi-variés, il faut définir une nouvelle méthode robuste à ces effets d'échelle.



Hyndman et Koehler (2006) proposent des mesures de l'erreur robustes à ces effets d'échelle. Ces mesures consistent à normaliser l'erreur basée sur l'erreur absolue moyenne (MAE<sup>9</sup>) de la méthode de prévision naïve (qui suppose un modèle de marche aléatoire).

Une erreur normalisée en t est définie comme :

$$q_t = \frac{x_t - \hat{x}_t}{(1/n - 1) \sum_{i=2}^n |x_i - x_{i-1}|} \quad (8)$$

Cette erreur est plus petite que l'unité lorsque le résultat de la prévision est meilleur. Elle est plus grande que un dans le cas d'une mauvaise prévision. L'erreur absolue moyenne normalisée proposée par Hyndman et Koehler (2006) est donnée par :

$$MASE = moyenne(|q_t|)$$

De façon analogue, la racine carrée du carré de l'erreur moyenne normalisée (RMSSE<sup>10</sup>) peut être définie. Comme la normalisation de l'erreur dépend des caractéristiques de la série, notre chronique sera normalisée en prenant en compte la saisonnalité et fera abstraction de l'extra-saisonnier. Ainsi, le carré de l'erreur normalisée au temps t est défini comme suit :

$$S_t = \frac{(x_t - \hat{x}_t)^2}{(1/n - s) \sum_{i=s}^n (x_{i-s})^2}$$

s représente la période de la saisonnalité. La prévision est d'autant meilleure que  $S_t$  tend vers zéro c'est-à-dire lorsque la valeur estimée est de plus en plus proche de la valeur observée. La RMSSE peut être définie de la façon suivante :

$$RMSSE = \sqrt{moyenne(S_t)}$$

Le tableau suivant nous donne les valeurs des RMSSE calculées pour chacune des méthodes d'estimation.

TABLE 5: Les mesures de performance des méthodes d'estimation

$RMSSE_b$	$RMSSE_h$
0.73	0.50

Les valeurs des RMSSE sont toutes inférieures à l'unité. Ces résultats supposent que les valeurs prédites de la consommation d'électricité sont proches des vraies valeurs. Par conséquent, nous pouvons conclure sur une meilleure qualité de la prévision. Cette prévision de la consommation intérieure d'électricité sur les données intervalles sera confrontée à celle sur les valeurs uniques afin de

9. Mean Absolute Error

10. Root Mean Square Scaled Error

mettre en évidence l'importance de la prise en compte des données sous un format intervalle. Pour cela, nous effectuerons d'abord une prévision sur les données ponctuelles par régression sur séries de Fourier.

## 5 Régression sur séries de Fourier et prévision

L'analyse traditionnelle d'une série chronologique repose sur une série de valeurs uniques. Pour constituer cette série de valeurs uniques à partir des intervalles de valeurs, il faut déterminer une classe d'équivalence pour ces intervalles. Cette détermination de la classe d'équivalence nécessite le choix d'une relation d'équivalence. Ainsi, les résultats issus de ces méthodes traditionnelles d'analyse des séries chronologiques seront indexés au choix de cette relation d'équivalence. Dans le cadre de notre application, nous choisirons comme relation d'équivalence la moyenne arithmétique simple ainsi, le représentant de chaque intervalle sera la valeur centrale de l'intervalle.

### 5.1 Désaisonnalisation par régression sur séries de Fourier

La représentation graphique de la série de valeurs centrales est donnée sur le graphique ci-dessous.

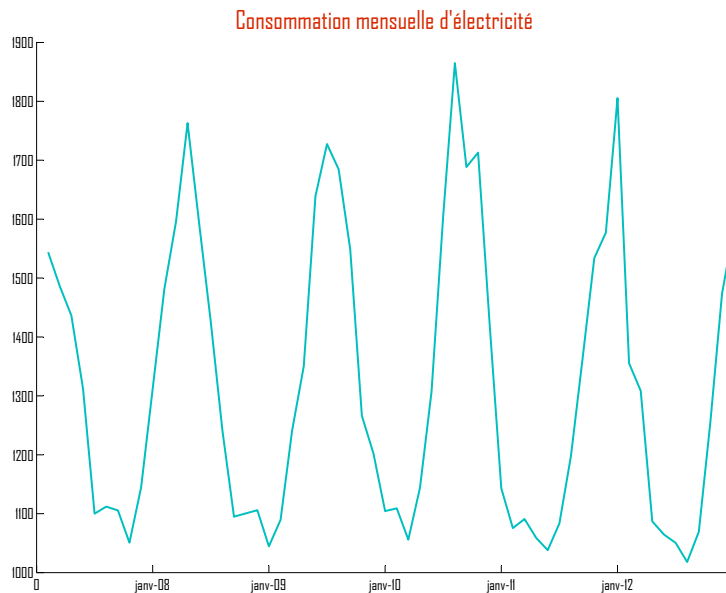


FIGURE 7: Évolution de la consommation mensuelle d'électricité

Une inspection visuelle de ce graphique nous indique que la consommation d'électricité varie de manière périodique. La période d'oscillation semble être annuelle (12 mois). La nature périodique de nos données suggère que le modèle approprié pour décrire cette dynamique peut s'écrire de la façon suivante :

$$C_t = \alpha_0 + \sum_{i=1}^m \beta_i \cos \frac{2\pi}{12} it + \sum_{i=1}^m \gamma_i \sin \frac{2\pi}{12} it + \varepsilon_t$$

avec  $x_t$  représentant la série de consommation,  $m = \frac{T}{2} - 1 = 5$  le nombre d'harmoniques,  $T$  est la périodicité,  $t = 1, 2, \dots, n$ . Les coefficients  $\alpha_0, \beta_i$  et  $\gamma_i$  représentent les coefficients de Fourier. Alors, développer une chronique en séries de Fourier revient à déterminer ses coefficients. Ils sont calculés de la manière suivante :

$$\alpha_0 = \frac{1}{T} \int_0^T C(t) dt$$

$$\beta_i = \frac{2}{T} \int_0^T C(t) \cos(i\omega t) dt$$

$$\gamma_i = \frac{2}{T} \int_0^T C(t) \sin(i\omega t) dt$$

Le paramètre  $\alpha_0$  représente la valeur moyenne de la série et les  $\beta_i$  et  $\gamma_i$  sont les amplitudes correspondant aux différents multiples de la fréquence fondamentale. L'estimation du modèle est donnée par :

$$\hat{C}_t = \hat{\alpha}_0 + \sum_{i=1}^5 \hat{\beta}_i \cos \frac{2\pi}{12} it + \sum_{i=1}^5 \hat{\gamma}_i \sin \frac{2\pi}{12} it$$

Nous soulignons que la série de consommation d'électricité est approchée par une somme de termes trigonométriques modélisant la saisonnalité. L'absence d'extra-saisonnier sur l'évolution du graphique et dans les tests de détection de cette composante n'est pas modélisée dans l'équation précédente. Les résultats d'estimation sont présentés dans le tableau ci-dessous.

TABLE 6: Résultats de la régression sur séries de Fourier

Variables	Coefficients	écarts types
C	1 317.5	(9.82)
CS1	275.82	(13.89)
CS2	39.96	(13.89)
CS3	-14.10	(13.89)
CS4	1.66	(13.88)
CS5	9.56	(13.87)
SN1	159.32	(13.88)
SN2	50.36	(13.88)
SN3	-17.60	(13.89)
SN4	-1.89	(13.89)
SN5	-15.71	(13.90)

Nous écrivons la série désaisonnalisée de la façon suivante :

$$C_t^{CVS} = C_t - \left( \hat{\alpha}_0 + \sum_{i=1}^5 \hat{\beta}_i \cos \frac{2\pi}{12} it + \sum_{i=1}^5 \hat{\gamma}_i \sin \frac{2\pi}{12} it \right)$$

$C_t^{CVS}$  est la série corrigée des variations saisonnières. Sa représentation est donnée par le graphique (8).

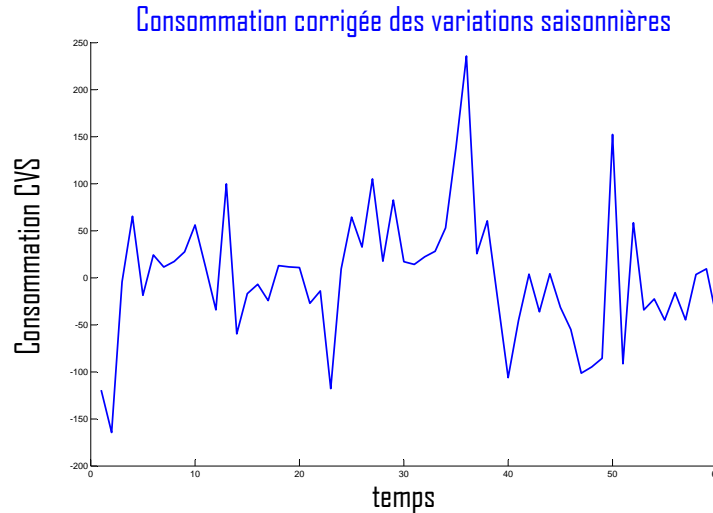


FIGURE 8: Série de consommation mensuelle d'électricité corrigée des variations saisonnières(CVS)

## 5.2 Prédiction par régression sur fonctions trigonométriques

Les coefficients saisonniers sont calculés à partir de la série corrigée des variations saisonnières. Ils représentent le profil saisonnier moyen de notre chronique. Comme la saisonnalité se reproduit presque à l'identique, ces coefficients seront utiles dans la prédiction des valeurs de cette composante. La prédiction consistera à rajouter à chaque valeur prévue de la consommation, le coefficient saisonnier correspondant. Les valeurs prévisionnelles sont répertoriées sur la table ci-dessous.

Pour mesurer la performance, nous avons calculé la RMSSE prenant en compte les effets d'échelle comme explicitée plus haut. Le résultat est présenté sur la table ci-dessous.

Nous remarquons que la RMSSE est plus faible que celle calculée sur la prédiction des données intervalles. Cela suppose que la modélisation par la série de Fourier est appropriée. Ainsi, nous pouvons construire l'intervalle de confiance(IC) de la prédiction afin d'analyser sa capacité à suivre les valeurs prévisionnelles calculées sur les bornes de l'intervalle. L'intervalle de confiance de la prédiction avec un niveau de confiance de 5% est donné par :

$$IC = \hat{C}_t \pm 1.96 \sqrt{Var(e_{n,n+h})}$$

TABLE 7: Pr evision par r egression sur fonction trigonom etrique

mois	Pr�evisions	Vraies valeurs
Janvier	1665.67	1698.93
F�evrier	1654.30	1690.20
Mars	1443.12	1566.47
Avril	1255.08	1346.36
Mai	1113.82	1172.04
Juin	1093.25	1080.75
Juillet	1086.18	1024.53
Ao�ut	1041.72	967.41
Septembre	1102.24	1012.76
Octobre	1260.87	1157.08
Novembre	1446.98	1393.52
D�ecembre	1648.20	1650.74

TABLE 8: mesure de performance de la pr evision sur s eries de Fourier

M�ethode	<i>RMSSE</i>
Fourier	0.37

$Var(e_{n,n+h})$  repr esente la variance des erreurs de pr evision et  $h$  l'horizon de pr evision qui vaut dans notre contexte 12.

L'int er et de la construction de cet intervalle de confiance de la pr evision est d'effectuer une  valuation des deux conceptions d'analyse des s eries chronologiques   savoir ; l'analyse par une s erie de valeurs uniques et celle prenant en compte les donn ees intervalles.

## 6  valuation des deux conceptions d'analyse des s eries temporelles

Nous allons confronter   l'aide des repr esentations graphiques suivantes, les valeurs pr edites pour chaque borne de l'intervalle<sup>11</sup> et les intervalles de confiance de la pr evision calcul es   partir du format de donn ees classiques.

Sur le premier graphique (9), pr evision bas et pr evision haut repr esentent respectivement les valeurs pr evues des bornes de l'intervalle. IC min et IC max sont les bornes inf erieures et sup erieures de l'intervalle de confiance. En confrontant ces diff erentes bornes, nous remarquons que l'intervalle de confiance   5% se retrouve le plus souvent entre les valeurs pr evisionnelles du minimum et du maximum de l'intervalle. Ce r esultat nous indique par cons equent qu'il existe des r egions dans lesquelles la vraie valeur pourrait se retrouver et dont il se-

11. Les valeurs pr evisionnelles utilis ees sont celles de la m ethode bas ee sur les bornes de l'intervalle car cette m ethode garantit l'ordre de l'intervalle sans une restriction qui peut concourir   r eduire l'efficacit e du mod ele.

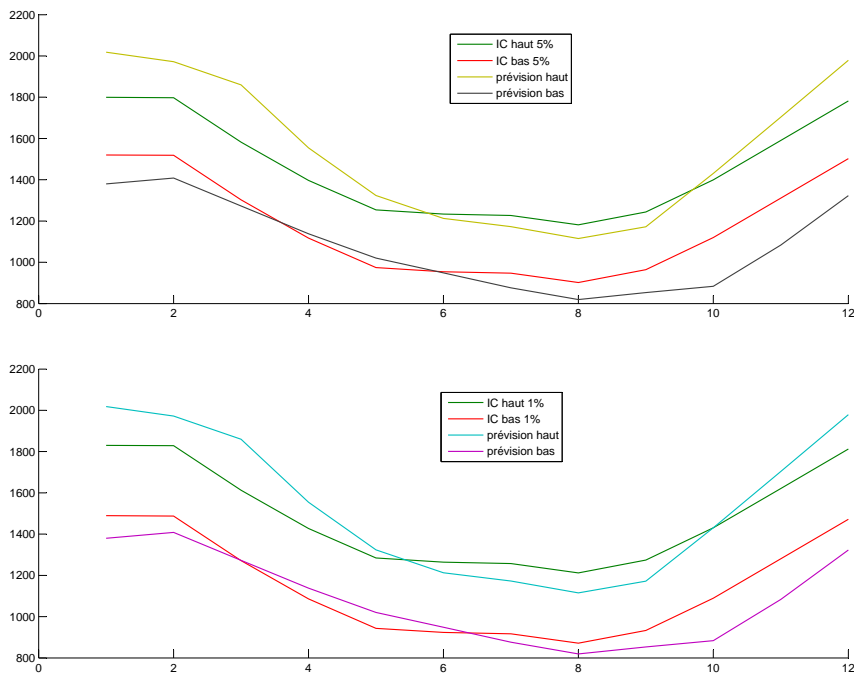


FIGURE 9: Évolution des valeurs prévisionnelles de l'intervalle et l'IC à 5% et 1% de la prévision sur données classiques

rait impossible de la prédire en effectuant une analyse économétrique prenant en compte les données classiques. Sur le second graphique, nous accordons 99% de chance que la vraie valeur tombe dans l'intervalle de confiance. Cette révision du niveau de confiance a pour conséquence d'élargir l'intervalle de confiance. Malgré cet élargissement, l'IC continue d'évoluer de manière à ne pas garantir une forte probabilité que la vraie valeur tombe dans la région couverte par le minimum et le maximum. Ainsi, la prise de décision sous l'hypothèse que les variables sont représentées par des valeurs uniques sera fortement impactée. Par contre, la prise en compte des données intervalles nous permettra d'effectuer des prévisions qui suivent l'évolution des bornes de l'intervalle. Nous illustrons ce résultat sur le graphique (10).

Nous observons que les valeurs prévisionnelles suivent l'évolution des vraies valeurs avec une marge d'erreur assez faible. Ainsi, la probabilité de se tromper sur la prévision de la vraie valeur est plus faible en considérant les données intervalles qu'avec des données à valeurs uniques.

La prévision de la demande d'énergie électrique est un résultat clé dans l'industrie électrique car elle constitue un élément fondamental dans la prise de décision dans le système de planification énergétique. C'est pourquoi nous avons jugé nécessaire de repenser de nouvelles techniques afin de réduire l'incertitude dans la prédiction. Il apparaît par conséquent que l'utilisation des données intervalles dans le cadre de la prévision de la demande d'électricité (le cas de notre illustration) constitue une progression encourageante car ; les données sous un

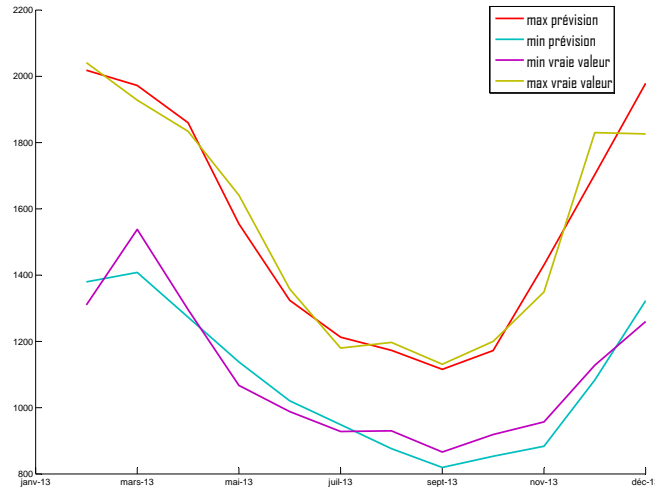


FIGURE 10: Évolution des valeurs prévisionnelles des bornes de l'intervalle et les vraies valeurs

format intervalle prennent en compte plus d'information que celles représentées par des valeurs uniques. Ainsi, **l'incertitude dans la prévision diminue avec la réduction de l'incertitude dans les données initiales.**

## 7 Conclusion

L'objet de cette étude consistait à présenter les méthodes d'estimation des paramètres d'un modèle de régression lorsque les variables sont observées sous un format intervalle puis, d'illustrer cette procédure d'estimation sur les données de consommation d'électricité en France afin d'effectuer une comparaison entre les valeurs prévisionnelles issues des données intervalles et l'IC de la prévision issus des données ponctuelles. A l'issue de ce travail, les principaux enseignements sont présentés de la façon suivante.

Dans un premier temps, nous avons montré que l'utilisation des valeurs uniques pour caractériser une variable économique entraîne une perte d'information. Comme l'économie cherche à comprendre le comportement des agents économiques, la prise en compte des données intervalles permettra d'accumuler plus d'information et ainsi, facilitera la compréhension des comportements des agents économiques.

La procédure d'estimation basée sur les limites inférieures et supérieures de l'intervalle fait abstraction de l'utilisation précoce de la régression classique pour adopter une méthode qui combine l'estimateur du maximum de vraisemblance et des MCO, ou le maximum de vraisemblance et l'estimateur des distances minimums. Pour garantir l'absence de croisement des bornes de l'intervalle, la restriction  $y_{lt} \leq y_{ut}$  est imposée au système bi-varié des bornes de l'intervalle.

Dans un deuxième temps, ces méthodes d'estimation ont été appliquées aux

données intervalles de consommation d'électricité en France dans une optique de prévision. La mise en œuvre de cette prévision a nécessité la recherche des composantes de la chronique. L'examen de l'évolution de nos données via des tests nous a permis de déceler un effet saisonnier mais pas d'extra-saisonnier. Ainsi, les procédures d'estimation sont mise en œuvre sur les données désaisonnalisées par la méthode Census X12 et la prévision prendra en compte les coefficients saisonnier obtenus après désaisonnalisation. Pour mesurer l'intérêt de l'analyse sur les données intervalles, nous avons comparé la prévision sur les données intervalles, à l'IC de la prévision sur les données ponctuelles.

La moyenne arithmétique simple sur les données intervalles nous a permis de construire notre série de données ponctuelles. La représentation graphique de cette série laisse apparaître une cyclicité dans l'évolution de la consommation d'électricité d'où la nécessité de modéliser par une série de Fourier. Les coefficients de Fourier ont été estimés par sur Matlab. Les intervalles de confiance de la prévision ont été calculés pour un niveau de confiance de 95 et 99%. Dans la dernière section, l'observation du profil des IC prévisionnels et celui des prévisions sur les données intervalles a révélé l'inclusion de l'IC de la prévision sur les données classiques, dans les bornes prévues par les données intervalles. Cette inclusion persiste avec l'augmentation du niveau de confiance accordé à l'IC de la prévision. Nous avons ainsi conclu sur l'impossibilité des données classiques à couvrir l'intervalle de la vraie valeur de la variable étudiée.

## Références

- [1] Arroyo, J., Gonzalez-Rivera, G., and Maté, C. (2010), Forecasting With Interval and Histogram Data. Some Financial Applications, in *Handbook of Empirical Economics and Finance*, eds. A. Ullah and D. Giles, Boca Raton, FL : Chapman and Hall, pp. 247-280.[473,483]
- [2] Billard, L., and Diday, E. (2000), Regression Analysis for Interval-Valued Data, in *Data analysis, Classification and related Methods, Proceedings of the Seventh Conference of the international Federation of Classification Societies(IFCS00)*. Belgium : Springer, pp. 369-374. [473]
- [3] Lima Neto, E., and de Carvalho, F. (2008), Centre and Range Method for Fitting a Linear Regression Model to Symbolic Interval Data, *Computational Statistics and Data Analysis*, 52, 1500-1515.[473,482].
- [4] Lima Neto, E., and de Carvalho, F. (2010), Constrained Linear Regression Models for Symbolic Interval-valued Variables, *Computational Statistics and Data Analysis*, 54, 333-347.[473,482].
- [5] Nath, G. (1972), Moments of a linearly Truncated Bivariate Normal Distribution, *Australian Journal of Statistics*, 14, 97-102.[479].
- [6] Garcia-Ascanio C., Maté C.(2010), Electric Power Demand Forecasting using Interval Time Series : A Comparison between VAR and iMLP, *Energy policy*, 38, 715-725.



- [7] Abdel-Aal, R.E., Al-Garni, A.Z.(1997), Forecasting Monthly Electric Consumption in Eastern Saudi Arabia using Univariate Time-series Analysis, *Energy*, 22, pp. 1059-1069.
- [8] Amemiya, T. (1973), Regression Analysis When the Dependent Variable is Truncated Normal, *Econometrica*, 41, 997-1016.
- [9] Maia, A., de Carvalho, F., Ludermir, T. (2008), Forecasting Models for Interval-Valued Time Series, *Neurocomputing*, 71, 3344-3352.
- [10] Saab,S., Badr, E., Nasr G. (2001), Univariate Modeling and Forecasting of Energy Consumption : The Case of Electricity in Lebanon, *Energy*, 26,1-14.
- [11] Brown,T. M. (1952), Habit Persistence and Lags in Consumer Behaviour, *Econometrica*, 20, 355-371.
- [12] Arroyo, J., Maté C. (2006), Introducing Interval Times Series : Accuracy Methods, *Computational Statistic*, pp.1139-1146.
- [13] Chavent M., Saaracco J. (2008), On Central Tendency and Dispersion Measures for Intervals and Hypercubes, *Communications in Statistics-Theory and methods*, 37, 1471-1482.
- [14] Arroyo J., Espinola R., Maté C. (2011), Different Approaches to Forecast Interval Times Series : A Comparison in Finance, *Computational Economics*, 37, 169-191.
- [15] Hyndman R.J., Koehler B.A. (2006), Another Look at Measures of Forecast Accuracy, *International Journal of Forecasting*, 22, 679-688.
- [16] Bourbonnais R., Terraza M. (2010), Analyse des Séries Temporelles : Applications à l'économie et à la gestion, 3e édition, Dunod, Paris.
- [17] Bourbonnais R. (2009), Econométrie, 7e édition, Dunod, Paris.
- [18] Gioia F., Lauro C.N. (2005), Basic Statistical Methods for Interval Data, *Statistica Applicata*, Vol 17, n°1.
- [19] Chueng Yin-Wong (2007), An Empirical Model of Daily Highs and Lows, *International Journal of Finance and Economics*, 12, pp. 1-20.
- [20] Wei Xu (2010),Symbolic Data Analysis : Interval-Valued Data Regression, Thesis of Doctorate of the University of Georgia.
- [21] Alefeld G., Mayer G.(2000), Interval Analysis : Theory and Applications, *Journal of Computation and Applied Mathematics*, 121, 421-464.
- [22] Tobin J.(1958), Estimation of Relationships for Limited Dependant Variables, *Econometrica*, 26, 24-36.
- [23] Gioia F. (2012), Arbitrage and Pricing in Financial Markets with Interval Data, Vol. 11, N° 5, PP. 678-685.
- [24] Lawson, C.L., Hanson, R.J. (1974), Solving Least Squares Problem. Prentice-Hall, New York.
- [25] Cazes, P. Chouakria, A., Diday, E., Schektman, S. (1997), Extension de l'analyse en composantes principales des données de type intervalle, *Revue de Statistique Appliquée*, Tome 45, N°3, P. 5-24.

- [26] Blanco-Fernandez A., Colubi A., Gonzalez-Rodriguez G. (2012), Confidence Sets in a linear Regression Model for Interval Data, *Journal of Statistical Planning and Inference*, 142, PP. 1320-1329.