



HAL
open science

Prévision journalière de la consommation d'électricité à l'aide d'un modèle à espace d'états fonctionnel

Komi Nagbe, Jairo Cugliari, Julien Jacques

► **To cite this version:**

Komi Nagbe, Jairo Cugliari, Julien Jacques. Prévision journalière de la consommation d'électricité à l'aide d'un modèle à espace d'états fonctionnel . Journées Des Statistiques, May 2018, Paris, France. hal-01785221

HAL Id: hal-01785221

<https://hal.science/hal-01785221v1>

Submitted on 4 May 2018

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.

PRÉVISION JOURNALIÈRE DE LA CONSOMMATION D'ÉLECTRICITÉ À L'AIDE D'UN MODÈLE À ESPACE D'ÉTATS FONCTIONNEL

Komi Nagbe ^{1,2} & Jairo Cugliari ² & Julien Jacques ²

¹*Enercoop ; komi.nagbe@enercoop.org* ²*Université de Lyon, Lyon 2 ERIC EA 3083 ; jairo.cugliari@univ-lyon2.fr, julien.jacques@univ-lyon2.fr*

Résumé. Ces dernières années, la multiplication des appareils électriques, la libération du marché de l'électricité, et le besoin de répondre en temps réel à la demande en électricité, ont fait de la prévision de la consommation électrique un enjeu important pour les fournisseurs d'électricité. Beaucoup de solutions ont été proposées pour la prévision journalière de l'électricité telles que les méthodes de reconnaissance des formes et les modèles de régressions. Dans ce document, nous proposons de prédire la consommation électrique en utilisant l'historique de consommation à travers un modèle autoregressifs à espace d'états fonctionnel. Pour estimer les paramètres de ce modèle, nous utilisons la maximisation de la vraisemblance, le lissage par splines, l'analyse en composantes principales fonctionnelles et le filtre de Kalman. Notre méthode est compétitive pour la prédiction de processus stationnaire. De plus, en cas de non-stationarité, notre méthode permet de prendre en compte des variables exogènes sources de la non-stationnarité.

Mots-clés. Prévision de la demande d'électricité ; Modèle d'espace d'états fonctionnel ; Filtre de Kalman ; Données fonctionnelles ; Lissage par splines ; Analyse en composantes principales fonctionnelles.

Abstract. In the last past years the liberalization of the electricity supply, the increasing variability of electric appliances and their use, and the need to respond to the electricity demand in the real time had made electricity demand forecasting a challenge. To this challenge many solutions are being proposed. For electricity demand forecasting, methods such as regression and pattern recognition methods were proposed. In this paper we propose to take only the past electricity consumption information embedded in a functional vector autoregressive state space model to forecast the future electricity demand. To estimate the parameters of this model, we use likelihood maximization, spline smoothing, functional principal components analysis and Kalman filtering. The principal advantage of this model is to forecast electricity demand without taking into account exogenous variables in case of stationarity. But in case of non stationarity, we integrate in this model exogenous variables which are known as the cause of non stationarity of the load demand.

Keywords. Electricity demand forecasting ; Functional state space model ; Kalman filtering ; Functional data ; Spline smoothing ; Functional principal components analysis

1 Présentation des données de consommation d'électricité de la société ENERCOOP

ENERCOOP est un fournisseur d'électricité d'origine renouvelable. Ses clients sont des particuliers, des entreprises et des commerces. Dans ce travail, nous nous intéressons à la consommation agrégée d'électricité pour ENERCOOP, définie comme la somme des consommations de l'ensemble de ses clients. Disposer d'une prévision journalière de l'électricité permet aux fournisseurs d'équilibrer en temps réel la production et la consommation d'électricité. La Figure 1 montre l'évolution de la consommation journalière d'électricité par les clients d'ENERCOOP pendant 6 ans.

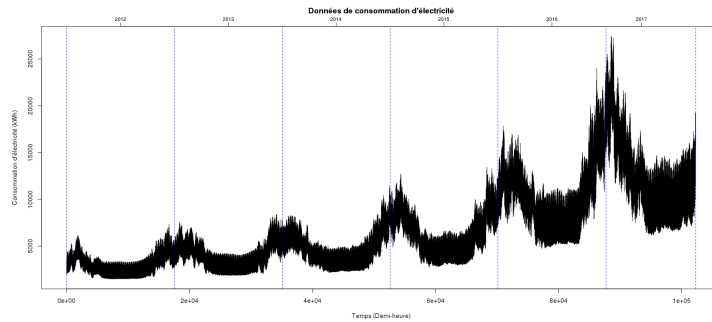


FIGURE 1 – Demande d'électricité pour le fournisseur ENERCOOP pendant six ans.

Les données de consommation sont enregistrées au pas de 30 minutes, ce qui donne 48 points de consommation enregistrés dans la journée. Sur la Figure 1, on devine une évolution journalière (plus visible en effectuant un zoom que nous ne présentons pas ici par manque de place), ainsi qu'une tendance annuelle. Par ailleurs, une caractéristique de la consommation d'électricité est la distinction entre weekends, jours fériés et jours ouvrables d'une part, et la distinction entre les mois doux et les mois plus froids d'autre part. Pendant les jours fériés et les weekends, on consomme moins d'électricité que pendant les jours ouvrable. Les mois d'hiver sont plus énergivores que les mois d'été, à cause du fait que beaucoup de clients utilisent l'électricité pour se chauffer.

2 Données fonctionnelles

2.1 Consommation de l'électricité comme séquence de fonctions

Bien que l'on n'enregistre la consommation qu'à certains instants de la journée (chaque demi-heure), la consommation électrique est un continuum. Par conséquent, nous considérerons la courbe de charge comme une fonction du temps.

Le point de départ de notre modèle est un processus stochastique à temps continu $Z = \{Z(t), t \in \mathbb{R}\}$. Pour étudier ce processus, un outil utile [1] est de considérer un processus stochastique de second ordre $X = \{X_i(t), i \in \mathbb{N}, t \in [0, \delta]\}$ qui est un processus à temps discret prenant ses valeurs à chaque instant sur un certain espace fonctionnel. Le processus X est extrait

de Z comme suit. Pour une trajectoire Z observée sur l'intervalle $[0, T]$, $T > 0$, nous considérons n sous-intervalles sous la forme $I_i = [(i-1)\delta, i\delta]$, $i = 1, \dots, n$ tels que $\delta = T/n$. Nous pouvons ainsi écrire que,

$$X_i(t) = Z((i-1)\delta + t), \quad t \in [0, \delta] \quad i = 1, \dots, n.$$

Par conséquent, anticiper le comportement de Z sur $[T, T + \delta]$ équivaut à prédire la fonction suivante $X_{n+1}(t)$ de X .

Dans la suite, δ correspondra à une journée (48 valeurs). Ainsi, nous chercherons à prédire une journée de consommation à partir de l'observation de n journées.

2.2 Projection des fonctions dans les bases de dimension finie et faible

En pratique, on ne dispose que d'un échantillonnage fini $\mathbf{x}_i = \{x_i(t_j), j = 1, \dots, N\}$ de chaque trajectoire $x_i(t)$, observée éventuellement avec du bruit. Une façon classique de reconstruire les courbes $\mathbf{x}_i(t)$ est de les projeter dans une base de fonctions de dimension finie. Parmi les nombreuses bases généralement utilisées, nous choisissons de travailler avec une base B-spline.

Si le nombre de splines utilisées est élevé, le temps de calcul du modèle présenté dans la section suivante peut être long. Afin de réduire la dimension du problème, et ainsi le temps de calcul, nous utilisons l'analyse en composantes principales fonctionnelles (ACPF) [2]. Nous cherchons à représenter les fonctions initiales dans une base de dimension plus faible que la base de splines. Ainsi, chaque courbe $x_i(t)$ peut être approximée dans une base de fonctions propres :

$$x_i(t) = \sum_{k=1}^p y_{ik} \xi_k(t) \quad (1)$$

où le nombre p de fonctions propres, supposé être relativement faible, sera choisi en fonction de l'erreur d'approximation des courbes. Notons \mathbf{y}_i le vecteur des p scores obtenus par la projection de $x_i(t)$ (la courbe de charge pour le jour i), dans la base de fonctions propres.

3 Modèle à espace d'états fonctionnel

Le modèle à espace d'états fonctionnel (MEEF) que l'on propose est une extension du modèle à espace d'états [3] permettant de manipuler des objets d'un espace fonctionnel. Pour cela, nous travaillons sur les scores \mathbf{y}_i des fonctions $x_i(t)$ dans la base de fonctions propres de l'ACPF. Le modèle proposé s'écrit alors

$$\begin{cases} \mathbf{y}_i = \mathbf{y}_{i-1} \alpha_i + \epsilon_i \\ \alpha_{i+1} = \alpha_i + \eta_i, \end{cases} \quad (2)$$

où ϵ_i et η_i sont des bruits gaussiens centrés indépendants avec respectivement des matrices de covariances H_i et Q_i inconnues, $\alpha_i \in \mathbb{R}^{m+1}$ est l'état du système à l'instant i .

Afin d'introduire de la parcimonie dans le modèle que l'on propose, tout en gardant une certaine flexibilité sur la modélisation, nous nous intéressons à trois formes de structures des matrices H_i et Q_i , pleine, diagonale et nulle, qui définissent ainsi six modèles possibles. Le Tableau 1 détaille ces six variantes du modèle (2). Ce tableau commence par le modèle le plus simple (H_i diagonale et Q_i nulle) et termine par le modèle le plus complexe (H_i pleine et Q_i pleine).

TABLE 1 – Variantes considérées pour le modèle (2) montrant les différentes structures des matrices H_i et Q_i , et le nombre de paramètres inconnus en fonction de p .

Variantes	H_i	Q_i	nb. de param.
1	Diagonale	Nulle	p
2	Diagonale	Diagonale	$p + p^2$
3	Diagonale	Pleine	$p + p^4$
4	Pleine	Nulle	p^2
5	Pleine	Pleine	$p^2 + p^2$
6	Pleine	Pleine	$p^2 + p^4$

4 Expériences sur les données d'ENERCOOP

Nous détaillons ici une des variantes mise en place pour rendre l'hypothèse de stationnarité raisonnable. Nous effectuons un certain nombre de correctifs au modèle original afin d'incorporer de l'information calendaire sous la forme de variables exogènes.

4.1 Correctifs pour non stationnarité

Pour prendre en compte la dépendance calendaire du profil de consommation nous avons intégré des variables indicatrices pour les jours, transformées en un tableau disjonctifs de 8 colonnes (7 variables jours de la semaine plus une variable rassemblant les jours fériés), ainsi qu'une variable numérique contenant le numéro du mois dans l'année. Toutes ces variables partagent une même nature, elles sont complètement connues au moment de la prévision. Dans ce travail nous n'avons pas intégré de variables météorologiques, dont la prévision nécessiterait l'utilisation de modèles dédiés. De manière concrète, le modèle de prévision que nous utilisons est le suivant :

$$\begin{cases} \mathbf{y}_i = \mathbf{Day}_i \beta_i^D + \mathbf{Month}_i \beta_i^M + \mathbf{y}_{i-1} \alpha_i + \epsilon_i \\ \alpha_{i+1} = \alpha_i + \eta_i. \end{cases} \quad (3)$$

où $\mathbf{Day}_i \in \{0, 1\}^{1 \times 8}$, $\beta_i^D \in \mathbb{R}^{8 \times p}$, $\mathbf{Month}_i \in \{1, \dots, 12\}$ et $\beta_i^M \in \mathbb{R}^{1 \times p}$. La variable \mathbf{Day}_i est le vecteur disjonctif des jours et prend ses valeurs dans $\{0, 1\}$ où 1 indique que le vecteur \mathbf{y}_i a été observé sur le jour correspondant, et 0 sinon. \mathbf{Month}_i est la variable correspondante aux mois de l'année, et indique dans quel mois le vecteur \mathbf{y}_i a été observé.

4.2 Résultats

Nous commençons par discuter l'apport de la projection dans la base de l'ACPF. Nous avons utilisé le package `fda` [4] pour représenter les données dans les bases splines ainsi que pour réaliser l'ACPF. Comme critères de mesure de qualité de prévision, nous utilisons les critères classiques MAPE (Mean Absolute Proportional Error) et RMSE (Root Mean Square Error).

La Table 2 montre les erreurs de reconstruction des données de consommation dans la base de composantes principales fonctionnelles. On remarque qu'une bonne reconstruction des données peut être obtenue avec un faible nombre de composantes principales liées à un nombre élevé

de splines. Nous avons choisi de travailler avec 45 splines et 10 composantes principales. Le choix est guidé par un examen minutieux de la qualité de la reconstruction de courbes. On peut remarquer donc que nous sommes passés d’une dimension de 48 à une dimension de 10, ce qui permet de réduire considérablement le temps de calcul.

TABLE 2 – Erreurs RMSE et MAPE pour la reconstitution des données de consommation en fonction du nombre de splines et du nombre de composantes principales.

nb PC	MAPE (en %)					RMSE (in kWh)				
	nb. de splines					nb. de splines				
	12	24	40	45	47	12	24	40	45	47
2	3.400	3.250	3.320	3.770	4.190	343	331	332	351	385
5	1.770	1.440	1.550	1.400	2.280	176	145	149	141	217
10	1.350	0.760	0.680	0.750	0.790	132	69	58	64	71

Nous évaluons maintenant la tâche de prévision. L’estimation de notre modèle à espace d’états fonctionnel (3) est réalisée à l’aide du package `KFAS` [5].

La Table 3 présente les résultats de prévision pour le modèle (3). Pour obtenir ces résultats, nous avons considéré cinq années de données d’apprentissage et une année de données test pour évaluer notre modèle. Pour chaque jour de l’année de test nous obtenons la prévision à horizon 1 jour, c’est-à-dire pour les prochaines 48 demi heures. Ensuite, on calcule l’erreur de prévision avec les enregistrements effectivement observées. Ces derniers sont ensuite incorporés aux données d’apprentissage et une nouvelle prévision est obtenue.

TABLE 3 – Erreur MAPE (en %) journalière de prévision du modèle à espace d’états fonctionnel.

	Min.	1er quar.	Médian	3ème quart.	Max.	Moyenne (Écart-type)
Lundi	1.48	2.35	2.81	3.96	8.90	3.30 (1.39)
Mardi	0.65	1.51	2.16	2.93	8.06	2.56 (1.46)
Mercredi	0.65	1.44	2.04	2.73	8.06	2.21 (1.12)
Jeudi	0.78	1.41	1.87	2.97	7.50	2.42 (1.41)
Vendredi	0.77	1.42	1.93	3.39	8.39	2.39 (1.43)
Samedi	0.77	2.14	3.04	4.44	13.11	3.55 (1.96)
Dimanche	0.96	3.02	4.49	6.02	13.11	4.67 (1.92)
Jours Fériés	1.83	3.10	4.28	7.12	8.14	4.90 (2.03)
Global	0.65	2.05	2.83	4.20	13.11	3.25 (1.43)

Afin d’obtenir un comparaison avec une variante simple, nous utilisons une version d’un modèle de persistance, connu pour être une méthode de référence, définie par :

$$\hat{X}_{n+1}(t) = \begin{cases} X_n(t) & \text{si le jour } n \text{ est Lundi, Mardi, Mercredi ou Jeudi} \\ X_{n-7}(t) & \text{sinon} \end{cases} . \quad (4)$$

L’erreur globale de la prévision par persistance est de 5.60 % MAPE avec un écart type de 1.78%. Cette erreur est sensiblement supérieure à celle obtenue par notre procédure.

5 Conclusion

Nous avons présenté un modèle de prévision de la consommation d'électricité à court terme. Le modèle utilise la forme fonctionnelle de la courbe de charge pour représenter dans un espace de dimension faible les transitions entre jours. Un certain nombre de correctifs sont nécessaires afin de prendre en compte la nature non stationnaire du signal, induite par l'information calendaire ou la saisonnalité de la demande.

La prévision des jours fériés et weekends reste difficile pour notre modèle. Néanmoins, dans tous les cadres de prévisions étudiées (par jour de semaine, par mois) le modèle proposé fournit des résultats meilleurs que la prévision par persistance.

A l'étape actuelle de notre modélisation, nous n'avons pas encore tenu compte de l'information météorologique, notamment de la température qui est connue pour avoir une influence importante sur la consommation d'électricité. Dans de futurs travaux, nous allons intégrer cette information météorologique dans le modèle et mesurer son apport dans la prévision de la consommation d'électricité.

Références

- [1] Denis Bosq. *Linear Processes in Function Spaces*. Springer New York, 2000.
- [2] James O. Ramsay, Giles Hooker, and Spencer Graves. *Functional data analysis with R and MATLAB*. Springer New York, 2009.
- [3] James Durbin and Siem Jan Koopman. *Time series analysis by state space methods*, volume 38. OUP Oxford, 2012.
- [4] James O Ramsay, Hadley Wickham, Spencer Graves, and Giles Hooker. fda : Functional data analysis, 2011. *R package <http://CRAN.R-project.org/package=fda>*. Cited on, page 24.
- [5] Jouni Helske. Kfas : Exponential family state space models in r. *Journal of Statistical Software, Articles*, 78(10) :1–39, 2017.