

ÉTUDE DE CAS POUR LA MODÉLISATION DE LA CONSOMMATION DOMESTIQUE D'EAU CHAUDE

Aurore LOMET ¹ & Frédéric Suard ¹ & David Chèze ²

¹ *CEA, LIST, F-91191 Gif-sur-Yvette, France*

² *CEA, LITEN, Le Bourget du Lac, France*
{*aurore.lomet, frederic.suard, david.cheze*}@cea.fr

Résumé. Ce papier présente une étude sur la modélisation de la consommation d'eau chaude sanitaire (ECS) de résidences équipées de ballons d'eau. L'objectif est de prévoir les besoins journaliers des habitants afin d'anticiper ces besoins dans un système de pilotage visant à réduire les dépenses énergétiques. Dans le cadre de cette étude, les usages réels de plusieurs résidences en France ont été analysés sur une période de deux ans. Les données mesurent en continu le volume journalier consommé, la température à l'entrée et à la sortie du ballon par jour. Cette étude préliminaire a mis en évidence la présence de variations aléatoires, une faible influence saisonnière, une périodicité d'une semaine pour l'ensemble des résidences qui sont cependant distinctes par différents profils journaliers suivant la résidence observée et le jour de la semaine. N'ayant aucune information a priori, nous proposons des modèles de séries temporelles avec une partie auto-régressive dont la forme est basée sur l'analyse de données. Les estimations de ces modèles nécessitent peu de ressources et peuvent être aisément embarquées. Ces modèles sont comparés à l'état de l'art sur l'ensemble des données réelles disponibles et sur des données de consommation de résidences en Suède. Les résultats montrent une nette amélioration de la qualité de la prévision pour les processus auto-régressifs et la généralité de notre approche.

Mots-clés. consommation d'eau chaude sanitaire, série temporelle, prévision.

Abstract. In this paper, we study the real domestic hot water (DHW) consumptions from single family house equipped with hot water tank. We model it forecast the daily needs of inhabitants in order to integrate it in a control strategy to reduce the energy cost by heating only the necessary DHW volume. At first, we realized a data analysis from real uses of several dwellings in France over two years between 2009 and 2011. The timestamps, the inlet and outlet temperatures of the boiler and the consumed volume are given. This study highlights a weekly periodicity, random fluctuations and the different profiles of consumption following the residence, the season, and the day of the week. Otherwise, having no prior information on the residents, we propose to model the daily DHW consumption by an adaptive time series model which does not require strong a priori and computational time. The output of our models are compared to estimations given by another model already existing in the literature. The good-promising results obtained on real data show that our model using the time series allows to take into

account the periodicity of the DHW consumption and random fluctuations. The tests on real consumption from Sweden dwellings show genericity of our approach.

Keywords. Domestic hot water consumption, time series, forecast.

1 Introduction

La consommation d'eau chaude sanitaire (ECS) représente une part importante du coût énergétique d'une résidence (entre 10% et 25% selon la localisation, EIA 2013). En France, elle représente en 2012 près de 11,2% du coût total des dépenses énergétiques (CEREN 2014). Ce coût énergétique pourrait dans le cas de résidence équipée de ballon d'eau chaude facilement diminuer en optimisant le volume d'ECS à chauffer. En effet, actuellement, le stock d'ECS est généralement contenu dans un réservoir de taille variable (de 100 à 200L) dont la quantité totale est chauffée à une température approximative de 60°C souvent pendant les heures creuses. Cependant, cette quantité est rarement utilisée en totalité et les déperditions de chaleur dans la journée impliquent des apports énergétiques pour la maintenir au delà d'une température seuil. Une solution pour limiter ces dépenses énergétiques est de prédire la consommation d'ECS afin de ne chauffer que la quantité d'eau nécessaire.

L'objectif est par conséquent de modéliser la consommation d'ECS afin de réaliser des prévisions du volume journalier d'ECS nécessaire aux résidents. Ainsi, ces prévisions peuvent être intégrées dans un système de pilotage pour la réduction du coût énergétique d'une résidence individuelle. Pour ce faire, on dispose des relevés journalier d'ECS de plusieurs résidences à usage réel fournis par le LITEN|LETh du CEA Chambéry, recueillis entre 2009 et 2011 en France et en Suède. Les données mesurent le volume journalier consommé, la température à l'entrée et à la sortie du ballon par jour. Toutes autres informations relatives aux résidences telles que la localisation ou le nombre d'habitants ne sont pas connues.

Plusieurs types de modélisation de consommation d'ECS ont été proposés tels que les modèles de produits de probabilités (Jordan and Vajen, 2001), de moyennes mobiles (Prud-homme and Gillet, 2001), de régression (Suganthi and Samuel, 2012) ou bien encore des réseaux de neurones (Aydinalp et al., 2004, Swan et al., 2011). Cependant, la plupart de ces modèles ont besoin en entrée d'informations a priori sur les habitants ou le type d'installation qui ne sont pas toujours disponibles et dont on souhaite s'affranchir pour garantir la genericité de l'approche. D'autres peuvent nécessiter d'importantes ressources de calculs et sont donc difficilement embarquables.

Dans cette étude, n'ayant pas d'a priori sur les spécificités des résidences et dans l'objectif d'embarquer le modèle de prévision, on réalise une modélisation des données par des processus de séries temporelles ayant une partie auto-régressive (ARIMA, ARMA, ARIMAX et ARMAX) présentés dans la section suivante. Ces différents modèles proposés s'appuient sur les résultats de l'analyse des données des résidences en France qui a

mis en évidence la présence de variations aléatoires, une faible influence saisonnière, une périodicité d’une semaine et différents profils suivant la résidence observée et le jour de la semaine. Puis, les conditions expérimentales sont présentées et les résultats de prévisions des résidences suédoises et françaises sont comparés aux prévisions du modèle issu de l’état de l’art proposé par Prud homme et Gillet (2001) (nommé ici modèle PG).

2 Modèles de prévision pour la consommation d’ECS

Le volume consommé d’ECS est donné sous la forme de relevés réalisés tous les jours. Ces observations (x_t) réalisées suivant une composante temporelle t à temps discret peuvent être exprimées mathématiquement par des processus autorégressifs tels que ARIMA (*Auto Regressive Integrated Moving Average*) et ARIMAX (*Auto Regressive Integrated Moving Average with eXternal inputs*) afin d’en analyser le comportement passé et d’en prévoir le futur.

Le processus ARIMA qui permet de prendre en compte la non-stationnarité des données est une forme généralisée du processus ARMA (*Auto Regressive Moving Average*). Il est défini par trois composantes (le processus autorégressif d’ordre p , le degré d’intégration d’ordre d et le processus de moyenne mobile d’ordre q) : $\forall t \in \mathbb{Z}, Y_t = \nabla^d X_t$, où Y_t est un processus ARMA (p, q) : $Y_t = \sum_{i=1}^p \varphi_i Y_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \epsilon_t$, où les coefficients φ_i et θ_i sont constants et où les ϵ_i sont les termes d’erreurs. La première somme du modèle ARMA représente la partie autorégressive qui exprime linéairement les estimations Y_t selon son historique. La seconde somme représente la partie moyenne mobile qui est une moyenne pondérée d’un certain nombre de valeur antérieures. Ainsi, un processus ARIMA dont le degré d’intégration est égal à zéro est un processus ARMA.

Considérant une fenêtre d’apprentissage de douze semaines d’observations afin d’obtenir un historique représentatif des profils de consommation, les données sont stationnaires (ce qui est vérifié par le test *augmented Dickey-Fuller*). Le processus ARMA est par conséquent retenu pour la modélisation de la consommation d’ECS. La forme générale de ce modèle est fixée selon les résultats de l’analyse de données des sept résidences qui a mis en évidence une périodicité d’une semaine. De plus, pour prendre en compte les variations de consommation dues à l’absence des résidents, lors des vacances par exemple, on considère aussi la consommation des jours précédents. Plusieurs formes de modèles respectant ces critères ont été testées et comparées en faisant varier le nombre de jours précédents et le nombre de semaines. Finalement, dans les phases de tests sur les données réelles françaises, nous avons sélectionné parmi les différentes formes de modèles respectant les a priori décrits précédemment, celle ayant été validée par le test *Ljung-Box* et pour laquelle l’erreur d’estimation était la plus faible. Ainsi, pour toutes les résidences, le modèle fixé est défini par : $X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \varphi_7 X_{t-7} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \theta_7 \epsilon_{t-7} + \epsilon + Cst$, où Cst est une valeur constante (la moyenne du volume journalier d’ECS consommé).

Seuls les valeurs des coefficients φ_i , θ_i et de la constante Cst varient selon la rési-

dence. En pratique, ces coefficients sont estimés dans une phase d'apprentissage basée sur un historique de données. Puisque qu'aucun autre paramètre n'influe sur le modèle, cet ensemble d'apprentissage doit être correctement choisi : il est préférable qu'il présente peu de données manquantes ou aberrantes et qu'il soit représentatif de la consommation d'ECS (absence des résidents, fluctuations, comportements périodiques).

Disposant de mesures de température qui ont une influence sur la consommation d'ECS, on propose également d'appliquer le processus ARMAX qui permet d'intégrer au processus ARMA des variables exogènes. Ainsi, on intègre la variable la température de l'eau à l'entrée du ballon de la veille que l'on considère comme représentative des variations de la température extérieure. Comme précédemment, seuls les coefficients varient et la forme générale du modèle est fixée pour l'ensemble des résidences suivant : $X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \varphi_7 X_{t-7} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \theta_7 \epsilon_{t-7} + \epsilon + Cst + \psi T_{t-1}$, où T est la température extérieure et ψ le coefficient associé à estimer.

3 Résultats et perspectives

Les processus ARMA et ARMAX présentés précédemment sont appliqués aux données de consommation d'ECS réelle de sept résidences en France et cinq en Suède afin de tester la généralité de l'approche. Pour chaque résidence, une fenêtre d'apprentissage de douze semaines est utilisée pour l'estimation des valeurs des coefficients. Il s'agit des trois mois précédant la phase de test. L'estimation des coefficients est réalisée par la méthode du maximum de vraisemblance.

Après cette phase d'apprentissage, les modèles sont appliqués pour l'estimation du volume total journalier sur une période d'un an avec une mise à jour quotidienne des données. Les résultats sont comparés aux estimations du modèle PG qui est une moyenne mobile sur les jours de la semaine correspondants, observés sur les deux mois précédents (Prud-homme et Gillet (2001)).

Les figures 1 et 2 donnent les estimations et la distribution des erreurs du modèle ARMA et ARMAX pour deux résidences en France et en Suède. À gauche de chaque figure, sont représentées les données en noir et les estimations en bleu ; à droite, sont représentées la distribution des erreurs et leur valeur médiane en rouge. Les valeurs négatives (respectivement positives) des histogrammes correspondent à une sous-estimation de la consommation (respectivement une sur-estimation) qui entraîne une perte de confort pour les résidents (respectivement une sur-consommation énergétique pour chauffer l'eau).

Les résultats montrent que les modèles ARMA et ARMAX corrigent les problèmes de latence du modèle PG notamment dans le cas de l'absence pendant quelques jours des résidents et des fortes variations de la consommation d'ECS. Les histogrammes mettent aussi en évidence que la distribution des erreurs d'estimation est plus centrée en zéro que celle du modèle PG et que le nombre de grandes erreurs est réduit. En effet, la partie autorégressive des modèles ARMA et ARMAX permet de retrouver en quelques jours la

tendance de la consommation des jours précédents comme illustré avec les vacances d’été sur la figure 1 (début août) et avec les fluctuations importantes (d’octobre et novembre, par exemple). Dans ce dernier cas et bien que le modèle PG soit performant pour une consommation régulière, il met plus d’un mois à retrouver la tendance contre seulement quelques jours pour ARMA et ARMAX. Contrairement à ce que l’on pouvait attendre après la phase d’étude qui montrait une influence de la température, l’intégration de cette variable exogène n’apporte pas une amélioration significative des prévisions. Cela s’explique par le fait que l’effet de la température est sous-jacent dans la consommation des jours précédents (X_{t-1}, X_{t-2}). Les résultats sur les données de résidences suédoises montrent également une amélioration de la prévision et la généralité de notre approche.

En conclusion, le modèle ARMA se montre être le plus performant quant à l’estimation du volume journalier d’ECS consommé. Ces prévisions pourraient être combinées à un modèle de consommation *intra-day* afin d’obtenir une prévision toutes les heures. L’objectif serait alors d’ajuster en continu et sur un horizon de quelques heures les besoins prévisionnels en ECS.

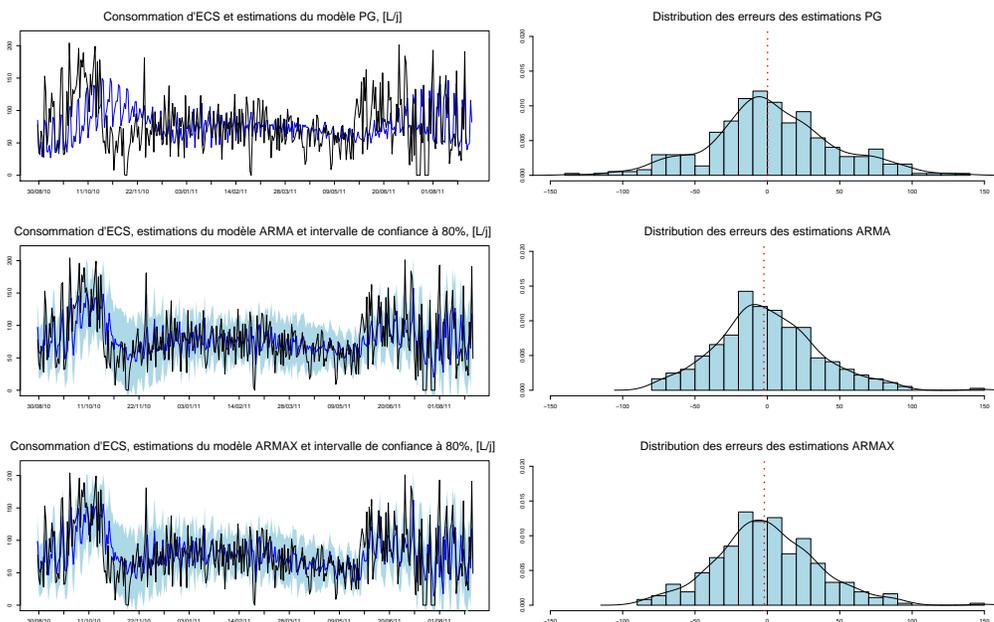


FIGURE 1 – Estimations des modèles PG, ARMA et ARMAX d’une résidence en France.

Remerciements : Les données ont été collectées depuis 2009 dans le cadre du projet Combisol de *Intelligent Energy Europe*, <http://www.combisol.eu>. Elles ont été mises à disposition par l’Institut National de l’Energie Solaire – Pôle Formation pour les installations en France et par le centre *Solar Energy Research Center SERC - Dalarna University* pour la Suède.

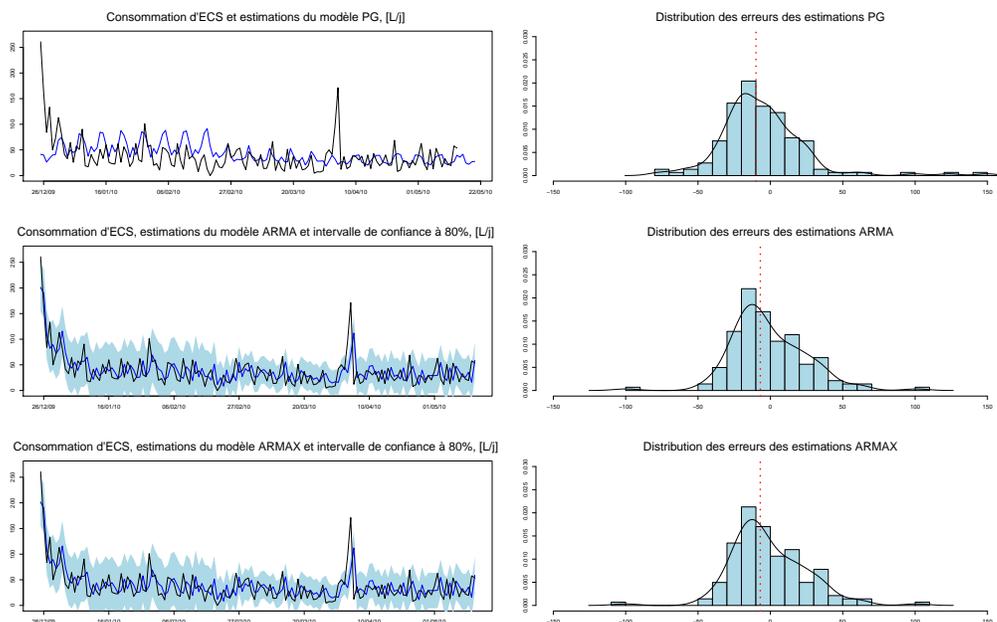


FIGURE 2 – Estimations des modèles PG, ARMA et ARMAX d'une résidence en Suède.

Bibliographie

- [1] M. Aydinalp, V. Ismet Ugursal, and A. S. Fung (2004), Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks. *Applied Energy*, 79(2), 159-178.
- [2] Centre d'études et de recherches économiques sur l'énergie (2014), Données statistiques du CEREN, <http://www.ceren.fr/stats>.
- [3] EIA (2013), U.S. Energy Information Administration. <http://www.eia.gov/consumption/residential>.
- [4] U. Jordan and K. Vajen (2001), Influence of the dhw load profile on the fractional energy savings : A case study of a solar combi-system with trnsys simulations. *Solar Energy*, 69, 197-208.
- [5] T. Prud-homme and D. Gillet. Advanced control strategy of a solar domestic hot water system with a segmented auxiliary heater (2001), *Energy and buildings*, 33(5), 463-475.
- [6] L. Suganthi and A. A. Samuel. Energy models for demand forecasting - a review (2012), *Renewable and Sustainable Energy Reviews*, 16(2), 1223-1240.
- [7] L. G. Swan, V. I. Ugursal, and I. Beausoleil-Morrison (2011), Occupant related household energy consumption in canada : Estimation using a bottom-up neural-network technique. *Energy and Buildings*, 43(2), 326-337.