

Modèle de régression et prévision de séries temporelles pour la demande en électricité à Djibouti

Mohamed Abdillahi Isman^{1,2} & Anne-Françoise Yao² & Paul-Marie Grollemund² & Julien Ah Pine³

¹ *Université de Djibouti, Djibouti, magouleh@gmail.com*

² *Laboratoire de Mathématiques Blaise Pascal, Université Clermont Auvergne, France, magouleh@gmail.com, anne.yao@uca.fr, paul_marie.grollemund@uca.fr*

³ *Université Lyon 2, Laboratoire ERIC, julien.ah-pine@univ-lyon2.fr*

Résumé

Le fonctionnement du système électrique repose sur la nécessité d'assurer l'équilibre entre la production et la demande en électricité à chaque instant, pour une grande variété de situations (pannes, maintenance des groupes de production et des infrastructures de réseau etc). L'offre et la demande en électricité à Djibouti n'est toujours pas satisfaite à cause de différents facteurs comme l'augmentation du nombre de ménages, les conditions climatiques et les habitudes de consommation. A Djibouti, aucune étude n'a été réalisé jusqu'à présent. Aussi, notre objectif est de proposer des modèles d'aide à la maîtrise de l'équilibre entre production et demande pour Electricité De Djibouti (EDD). Dans cet exposé, nous présenterons les résultats de la première étape de l'étude où nous avons testé différentes méthodes permettant comprendre l'évolution de la demande à Djibouti sur les dix dernières années. Nous nous intéressons ici aux performances de quelques modèles classiques de séries temporelles en amont d'étapes intégrant des co-variables qui seront par la suite défini en accord avec la compagnie d'électricité de Djibouti.

Mots-clés. Modélisation de série temporelle, prévision, Modèles SARIMA, TBATS, NNAR, STL

Abstract

The operation of the electrical system is based on the need to ensure the balance between production and demand for electricity at all times, for a wide variety of situations (breakdowns, maintenance of production units and network infrastructures, etc.). Electricity supply and demand in Djibouti is still not satisfied due to various factors such as the increase in the number of households, climatic conditions and consumption habits. In Djibouti, no study has been carried out so far. Also, our objective is to propose models to help control the balance between production and demand for Electricité De Djibouti (EDD). In this presentation, we will present the results of the first stage of the study where we tested different methods to understand the evolution of demand in Djibouti over the last ten years. We are interested here in the performances of some classic models of time series upstream of stages integrating co-variables which will be defined thereafter in agreement with the electricity company of Djibouti.

Keywords. Time Series Modeling, Forecasting, SARIMA Models, TBATS, NNAR, STL

1 Introduction

L'électricité a atteint une place très importante dans le monde en particulier à Djibouti. Il est devenu un contributeur majeur à l'amélioration du niveau de vie des habitants, du développement économique et

industriel. Sa demande dépend fortement de l'endroit où elle est utilisée mais la régularité de son offre pose des défis particuliers puisqu'il s'agit d'une énergie difficile à stocker à grande échelle. Elle doit, en permanence, faire l'objet d'un équilibre entre la production et la demande pour répondre au besoin d'approvisionnement d'électricité. De ce fait, le gestionnaire de réseau doivent anticiper la demande en réalisant des prévisions sur la demande d'électricité. A l'échelle Djiboutienne, la production d'électricité a atteint 627.111 Mwh en 2020 contre 605.150 Mwh en 2019, soit une progression de 3.63% [4]. La part de l'énergie importée de l'Ethiopie en 2017 s'est établie à 508.198 Mwh, soit un accroissement de 13% par rapport à 2016 où l'énergie importée avait atteint 449.510 Mwh. En proportion, la part de l'énergie importée représente 89.6% en 2017, contre 83.6% en 2016 du volume de production. La consommation électrique était, pour sa part, fixée à 517.035 Mwh à fin 2020, en hausse de 2.9% par rapport à l'énergie consommée en 2019 qui se situait à 502.244 Mwh.

Il en résulte que Djibouti, petit pays situé dans la Corne de l'Afrique avec une superficie de 23 200 km² et avec une population de 962 451 habitants, connaît une insuffisance de production d'électricité qui provoque régulièrement de pannes d'électricité et une haute dépendance vis-à-vis des importations énergétique. Les deux tiers de la population vivent en zone urbaine, essentiellement dans la capitale; le reste de la population vit principalement dans les villes secondaires. Son climat se classe parmi les plus chauds au monde avec une température moyenne annuelle de 30.1° C. En été, de mai à septembre, les températures extérieures atteignent entre 35° à 45° et par conséquent entraînent une forte consommation d'électricité. Pendant cette période, on observe souvent un pic de consommation d'électricité qui est dû à l'utilisation des équipements domestiques liés à la réfrigération (climatisation et ventilation). Ces derniers représentent le principal usage de l'électricité et correspondent à 78.6% de la consommation d'électricité des ménages [6]. Durant ces périodes, le gestionnaire de réseau est fortement sollicité pour faire face aux défis de la hausse de la demande.

Le secteur de l'électricité à Djibouti est un secteur en constante évolution et la demande en électricité est croissante d'une année à l'autre. Aujourd'hui malgré les efforts (l'élargissement du parc de production d'électricité) prises par les responsables de la société de production et de distribution de l'électricité à Djibouti, la demande d'électricité n'est toujours pas satisfaite complètement. Ceci est dû à l'insuffisance des productions en énergie électrique et par conséquent, on voit une panne d'électricité régulière programmée pendant les périodes de pic de consommation d'électricité. Le nombre de client de la société nationale d'électricité (EDD) ne cesse de croître ainsi que la consommation d'électricité au fil du temps. De ce fait, En vue d'assurer une adéquation entre les capacités de production d'électricité et la demande de sa clientèle, il est important pour l'EDD de développer des stratégies d'anticipation de la demande d'électricité. Dans ce contexte, Il est donc fondamental de construire une modélisation de la demande d'électricité à la fois pour décrire finement le présent (demande actuelle) mais aussi pour être en mesure de prédire leur évolution sur un plan technique (une meilleure gestion des données énergétiques et d'analyses, ainsi que dans l'obtention de prévisions plus précises).

2 Méthodes de modélisation et prévision

Nous décrivons dans cette partie quelques méthodes de prévision que nous avons utilisé dans la prévision de la demande d'électricité de Djibouti.

2.1 Modèle SARIMA

Le Modèle SARIMA est une extension du modèle ARIMA qui prend en charge les séries chronologiques saisonnières. Le modèle ARIMA est une classe de modèles de série chronologique linéaire non stationnaire proposés par Box et Jenkins en 1976. Le principe du modèle ARIMA est que les valeurs futures sont des combinaisons linéaires de valeurs passées.

Une série chronologique saisonnière peut être détectée par le graphique temporel de la série ainsi que par le graphique ACF. La présence de la saisonnalité dans la série donne au graphique temporel de la série une composition ondulatoire. On peut aussi déduire la saisonnalité à partir du comportement de l'ACF qui semble être décroissance sinusoidal.

Un modèle SARIMA peut s'écrire mathématiquement comme suivant :

$$\phi(L)\Phi(L^S)\Delta^d\Delta_S^D Y_t = \theta(L)\Theta(L^S)\epsilon_t$$

où $\phi(L)$ et $\theta(L)$ sont des polynômes caractéristiques d'ordre p et q respectivement pour la partie autorégressif (AR) et moyenne mobile (MA) non-saisonnaire. $\Phi(L^S)$ et $\Theta(L^S)$ sont des polynômes caractéristiques d'ordre P et Q pour la partie autorégressive et moyenne mobile saisonnière. $\Delta^d = (1 - L)^d$ et $\Delta_S^D = (1 - L^s)^D$ représentent respectivement les opérateurs de différenciations réguliers et saisonniers s avec d le nombre de fois que la série est différenciée pour éliminer l'effet de la tendance et D le nombre de fois que la série est différenciée pour éliminer l'effet de la saisonnalité sur la série. Y_t est la série étudiée et ϵ_t est l'erreur aléatoire.

La procédure itérative de la modélisation ARIMA proposé par Box et Jenkins sont (i) l'identification du modèle (ii) Estimation des paramètres du modèle choisi et (iii) validation du modèle/vérifications de diagnostic. Ces trois étapes ont été effectuées dans cette étude.

2.2 Méthode STL [9]

La méthode STL (seasonal-Trend decomposing using Loess) est un algorithme de decomposition d'une série chronologique basé sur une séquence de procédure de lissage en utilisant des modèles de régression ajustés localement. La méthode a été présenté par Cleveland et al. au Journal Officiel de Statistique en 1990. L'algorithme STL utilise la méthode LOESS (LOcal regrESSion) pour lisser la série chronologique en composants. Si la série chronologique Y_v a la partie tendance, la partie saisonnière et la partie reste notée respectivement T_v, S_v, R_v , pour $v = 1$ à N , alors la forme mathématique de la méthode est donnée par

$$Y_v = T_v + S_v + R_v$$

L'algorithme de décomposition STL consiste en deux procédures récursives.

Une boucle interne qui met à jour la composante tendance et la composante saisonnière de manière itérative à l'aide du lissage LOESS.

- 1^{ère} étape : L'enlèvement de la tendance. $Y_v - T_v^{(k)}$. Pour le premier passage $T_v^{(0)} = 0$.
- 2^{ème} étape : Cette étape utilise la méthode de LOESS pour lisser les valeurs de chaque cycle de sous-série de la série sans tendance avec $q = n_s$ et $d = 1$. Le résultat est défini comme C_v^{k+1} .
- 3^{ème} étape : Un filtre passe-bas composé de filtres de moyenne mobile et de LOESS, est utilisé pour capturer tous les mouvements à basse fréquence depuis C_v^{k+1} . Le résultat est marqué comme L_v^{k+1} .
- 4^{ème} étape : Décomposition de cycle de sous-série lissés en detrending : $S_v^{k+1} = C_v^{k+1} - L_v^{k+1}$ pour $v = 1$ à N .
- 5^{ème} étape : Désaisonnalisation à partir de la série initiale : $Y_v - S_v^{k+1}$.
- 6^{ème} étape : Cette étape utilise La méthode de LOESS pour lisser la série désaisonnalisée de l'étape 5 pour obtenir la composante de tendance avec $q = n_t$ et $d = 1$. (q est la fenêtre de lissage et d est le degré du polynôme).

Une boucle externe qui calcule les coefficients de robustesse pour minimiser l'effet des valeurs aberrantes (elle remplacera les LOESS à la deuxième et sixième étapes de la boucle intérieure avec le robuste LOESS).

2.3 Modèle TBATS [3]

TBATS (T: Saisonnalité trigonométrique, B: Transformation de Box-Cox, A: Erreurs ARMA, T: Tendance, S: Composants saisonniers) est un modèle de série chronologique qui est utile pour traiter des données avec plusieurs modèles saisonniers c'est-à-dire les données qui affichent un comportement saisonnier complexe, comme plusieurs périodes saisonnières, une saisonnalité à haute fréquence et une saisonnalité non entière.

Modèle :

$$y_t^{(\lambda)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^T S_{t-m_i} + d_t$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t$$

$$b_t = b_{t-1} + \beta d_t$$

$$d_t = \sum_{i=1}^p \varphi_i d_{t-i} + \sum_{i=1}^q \theta_i e_{t-i} + e_t$$

Où

- $y_t^{(\lambda)}$: Série temporelle au moment t (transformé Box-Cox)
- $S_t^{(i)}$: i-ème composante saisonnière
- l_t : niveau local
- b_t : tendance avec amortissement
- d_t : processus $ARMA(p, q)$ pour les résidus
- e_t : gaussien avec du bruit

Partie saisonnière :

$$S_t^{(i)} = \sum_{j=1}^{k_i} S_{j,t}^{(i)}$$

$$S_{j,t}^{(i)} = S_{j,t-1}^{(i)} \cos(\omega_i) + S_{j,t-1}^{*(i)} \sin(\omega_i) + \gamma_1^{(i)} d_t$$

$$S_{j,t}^{*(i)} = -S_{j,t-1}^{(i)} \sin(\omega_i) + S_{j,t-1}^{*(i)} \cos(\omega_i) + \gamma_2^{(i)} d_t$$

$$\omega_i = \frac{2\pi}{m_i}$$

Paramètres du modèle :

- T : quantité de saisonnalités
- m_i : durée de la i-ème période saisonnière
- k_i : quantité d'harmoniques pour la i-ème période saisonnière
- λ : transformation Box-Cox
- α, β : lissage
- φ_i, ϕ_i : $ARMA(p, q)$ coefficients
- ϕ : Tendance amortie
- $\gamma_1^{(i)}, \gamma_2^{(i)}$: lissage saisonnier (deux pour chaque période)

2.4 Modèle NNAR [13]

Le modèle NNAR (Neural Network Auto-Regression) est un réseau de neurones Feed-Forward avec une seule couche cachée et des valeurs décalées de la série chronologique utilisées comme entrées. Le modèle est noté comme $NNAR(p, P, k)_T$ où p est le nombre d'entrées des valeurs décalées de la série, P les décalages saisonniers, k le nombre de nœuds dans la couche cachée et T la longueur de la période saisonnière. La formule du modèle s'écrit comme

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}, y_{t-T}, y_{t-2T}, \dots, y_{t-PT}) + \epsilon_t$$

où f est la fonction d'activation sigmoïde avec une seule couche cachée de k neurone et ϵ_t est la série résiduelle.

Les valeurs p , P et k sont trouvées automatiquement par la fonction si elles ne sont pas spécifiés. Pour les séries temporelles saisonnières, la valeur par défaut est $P = 1$, et p est choisi à partir du modèle linéaire optimal ajusté aux données désaisonnalisées. Si k n'est pas spécifié, alors il est défini sur $k = (p + P + 1)/2$ (arrondi à l'entier le plus proche).

2.5 Métriques d'évaluations des performances de cinq modèles

Les modèles ajustés sur l'ensemble d'entraînement ont été utilisés pour prévoir à l'avance 36 mois. Trois indicateurs ont été calculés pour mesurer l'exactitude de la prévision pour les cinq modèles : l'erreur quadratique moyenne (RMSE), erreur absolue moyenne (MAE) et erreur absolue moyenne en pourcentage (MAPE). Ces indicateurs sont exprimés comme suit :

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad , \quad MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad \text{et} \quad MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{y_t}$$

où y_t est la série observée au temps t et \hat{y}_t la série ajustée.

3 Résultats

3.1 Données de l'étude

Les données utilisées dans cette étude sont la demande d'électricité collectés à partir de différentes postes de transformation installées dans la capitale de Djibouti. Ces données sont celles enregistrées à chaque mois de l'année sur la période de 2012 à 2020 par l'Électricité de Djibouti (EDD), qui est la plus grande compagnie de production et de distribution d'électricité à Djibouti. Nous nous sommes intéressés particulièrement les clients qui sont connectés au réseau électrique à basse tension (BT) avec une puissance souscrite entre 1kva et 9kva. Les données de l'ensemble d'apprentissage utilisées pour la construction de modèles sont de la période de janvier 2012 à décembre 2017 et celles de l'ensemble de test sont de la période de janvier 2018 à décembre 2020.

3.2 Comparaison de performance des modèles

On a comparé cinq modèles dont les modèles SARIMA, NNAR, TBATS, STL et la combinaison SARIMA-NNAR-TBATS-STL qui est le modèle donnant la moyenne des prévisions par tous les modèles ci-dessous. Les métriques d'erreur de prévision montrent que le modèle SARIMA est plus performant (en sens de RMSE, MAE ou MAPE) pour prédire la demande en électricité au moins pour la période 2018-2020. Ces mesures sont calculées à partir de l'ensemble de test, donc uniquement pour les observations jusqu'à la ligne noire en pointillés dans le graphique ci-dessous. Les observations après cette ligne pointillée noire sont les prévisions futures de la demande d'électricité.

Table 1: Comparaisons de performance des cinq modèles

	RMSE	MAE	MAPE
STL	11.066197	9.807536	13.728329
SARIMA	2.608943	2.177854	3.267496
NNAR	4.475172	3.626537	5.374722
TBATS	4.531631	3.905998	5.745926
Combinaison	4.565369	3.721192	5.281873

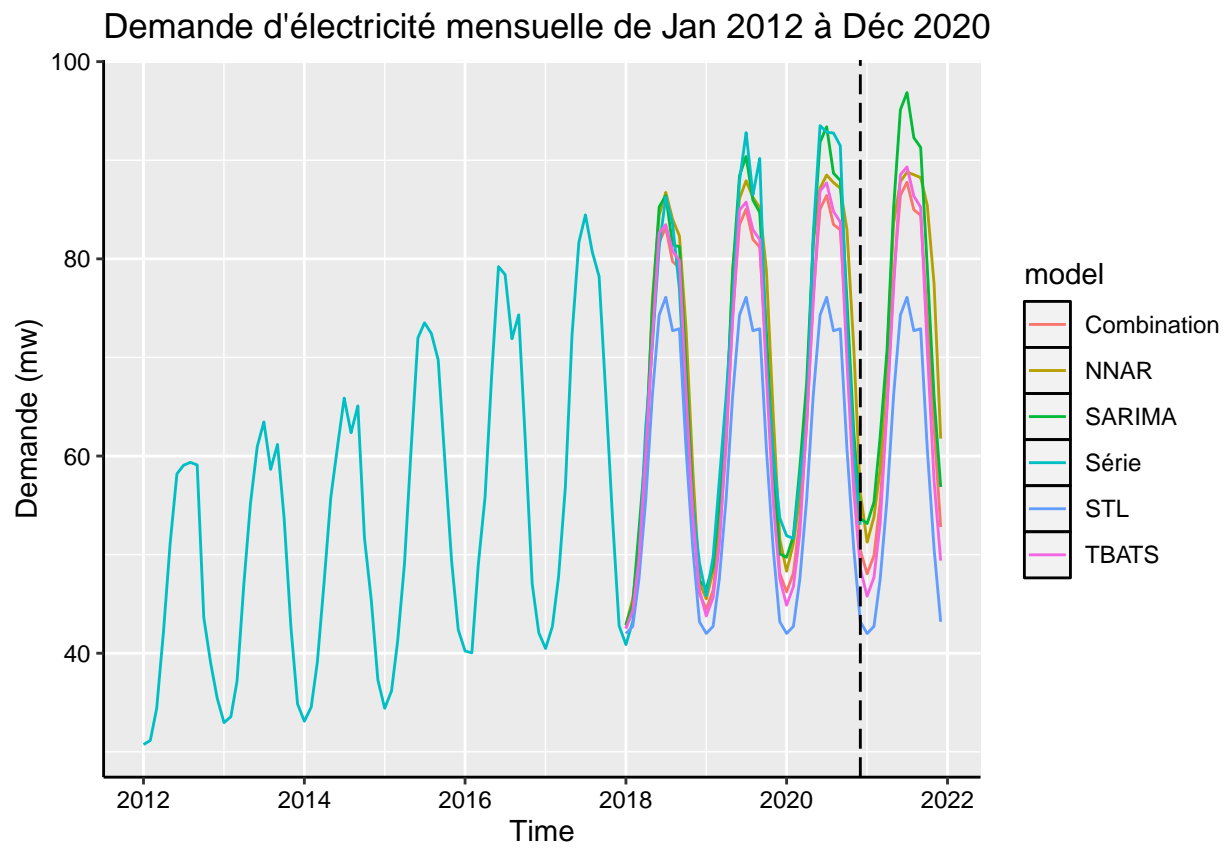


Figure 1: Prédiction avec cinq méthodes

Remerciements : Ce travail a été soutenu par la compagnie d'électricité de Djibouti en collaboration avec l'université Clermont Auvergne en France.

References

- [1] Petek Aydemir. 2018. Performance of ensemble forecasting tools for analysis turkish consumer price index. Master's thesis. Middle East Technical University.
- [2] Petek Aydemir. 2018. Performance of ensemble forecasting tools for analysis turkish consumer price index. Master's thesis. Middle East Technical University.
- [3] Jacek BROŻYNA, Grzegorz Mentel, Beata Szetela, and Wadim Strielkowski. 2018. Multi-seasonality in the TBATS model using demand for electric energy as a case study. *Economic Computation & Economic Cybernetics Studies & Research* 52, 1 (2018).
- [4] Électricité de Djibouti. 2020. *Rapport annuel de la compagnie d'électricité de djibouti*.
- [5] Rakshitha Godahewa, Christoph Bergmeir, Geoffrey I Webb, and Pablo Montero-Manso. 2020. A strong baseline for weekly time series forecasting. *arXiv preprint arXiv:2010.08158* (2020).
- [6] Institut National de la Statistique de Djibouti (INSD). 2004. *Enquête djiboutienne auprès des ménages - energie (EDAM-energie)*. INSD.
- [7] Maksat Kalimoldayev, Aleksey Drozdenko, Igor Kopylyk, T Marinich, Assel Abdildayeva, and Tamara Zhukabayeva. 2020. Analysis of modern approaches for the prediction of electric energy consumption. *Open Engineering* 10, 1 (2020), 350–361.
- [8] Farhan Mohammad Khan and Rajiv Gupta. 2020. ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in india. *Journal of Safety Science and Resilience* 1, 1 (2020), 12–18.
- [9] Abhinav Mishra, Ram Sriharsha, and Sichen Zhong. 2021. OnlineSTL: Scaling time series decomposition by 100x. *arXiv preprint arXiv:2107.09110* (2021).
- [10] Chukwudike C Nwokike, Bright C Offorha, Maxwell Obubu, Chukwuma B Ugoala, and Henry I Ukomah. 2020. Comparing SANN and SARIMA for forecasting frequency of monthly rainfall in umuahia. *Scientific African* 10, (2020), e00621.
- [11] Daniel Ollech. 2021. Seasonal adjustment of daily time series. *Journal of Time Series Econometrics* 13, 2 (2021), 235–264.
- [12] Gaetano Perone. 2021. Comparison of ARIMA, ETS, NNAR, TBATS and hybrid models to forecast the second wave of COVID-19 hospitalizations in italy. *The European Journal of Health Economics* (2021), 1–24.
- [13] Gongchao Yu, Huifen Feng, Shuang Feng, Jing Zhao, and Jing Xu. 2021. Forecasting hand-foot-and-mouth disease cases using wavelet-based SARIMA–NNAR hybrid model. *Plos one* 16, 2 (2021), e0246673.
- [14] Kun Zhou, Wenyong Wang, Lisheng Huang, and Baoyang Liu. 2021. Comparative study on the time series forecasting of web traffic based on statistical model and generative adversarial model. *Knowledge-Based Systems* 213, (2021), 106467. DOI:<https://doi.org/https://doi.org/10.1016/j.knosys.2020.106467>