



HAL
open science

Analyse exploratoire de séries temporelles d'utilisation de systèmes électriques par chaînes de Markov d'ordre variable

Hugo Le Picard, Fabrice Rossi

► **To cite this version:**

Hugo Le Picard, Fabrice Rossi. Analyse exploratoire de séries temporelles d'utilisation de systèmes électriques par chaînes de Markov d'ordre variable. 54es Journées de Statistique de la SFdS, Société Française de Statistique, Jul 2023, Bruxelles (BE), Belgique. hal-04335897

HAL Id: hal-04335897

<https://hal.science/hal-04335897v1>

Submitted on 11 Dec 2023

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.



Distributed under a Creative Commons Attribution 4.0 International License

ANALYSE EXPLORATOIRE DE SÉRIES TEMPORELLES D'UTILISATION DE SYSTÈMES ÉLECTRIQUES PAR CHÂÎNES DE MARKOV D'ORDRE VARIABLE

Hugo Le Picard ¹ & Fabrice Rossi ²

¹ *LEDa-CGEMP, Université Paris-Dauphine, PSL University,*
hugo.le-picard@dauphine.psl.eu

² *CEREMADE, CNRS, UMR 7534, Université Paris-Dauphine, PSL University,*
Fabrice.Rossi@dauphine.psl.eu

Résumé. Cet article propose d'utiliser des chaînes de Markov d'ordre variable dans leur version classique (VLMC) et dans leur version avec covariables exogènes (COVLMC) pour analyser l'utilisation de générateurs auxiliaires par des commerces nigériens pour pallier les défaillances du réseau électrique centralisé. Les résultats obtenus avec l'approche classique sont comparés à ceux produits par la prise en compte de covariables afin de déterminer si les schémas d'utilisation des générateurs auxiliaires suivent leur propre dynamique ou s'ils sont influencés par des facteurs externes, notamment par le réseau. Comme il n'existe pas (à la connaissance des auteurs) d'implémentation librement accessible pour ajuster des modèles de COVLMC, un nouveau package R, `mixvlmc`, est proposé.

Mots-clés. Séries temporelles discrètes, chaînes de Markov d'ordre variable, réseau électrique

Abstract. This paper proposes the use of variable order Markov chains in their classical version (VLMC) and in their version with exogenous covariates (COVLMC), to analyze the use of auxiliary generators by Nigerian businesses. The goal is to investigate how these generators mitigate the impact on their activity of power outages on the centralized electricity network. The results obtained with the classical approach are compared with those produced by taking into account covariates to determine whether the usage patterns of the auxiliary generators follow their own dynamics or whether they are influenced by external factors, in particular by the grid. As there is (to the best of the authors' knowledge) no freely available implementation for COVLMC estimation, a new R package named `mixvlmc` is introduced.

Keywords. Discrete-valued time series, variable length markov chain, power grid

1 Introduction

Nous nous intéressons dans cet article à la modélisation de séries temporelles à valeurs discrètes dans un objectif exploratoire. Il s'agit de caractériser de façon simple le comportement de telles séries afin de distinguer notamment si elles suivent une dynamique propre (de nature markovienne) ou si elles sont influencées par des facteurs extérieurs. Cette approche est

motivée par l'étude de l'utilisation de générateurs électriques de secours au Nigéria pour pallier les défauts du réseau électrique centralisé. On pourrait naïvement penser que les générateurs sont allumés quand le réseau centralisé fait défaut et sont éteints quand le courant est rétabli. Une observation de l'historique d'utilisation de générateurs montre que la situation est beaucoup plus complexe en réalité.

Nous précisons dans un premier temps le cadre applicatif (Section 2), puis nous présentons les outils statistiques utilisés pour l'analyse : les chaînes de Markov d'ordre variable dans leur version classique et dans leur version avec covariables exogènes (Section 3). Les résultats obtenus avec l'approche classique (Section 4) sont ensuite comparés à ceux produits en prenant en compte des covariables adaptées au problème (Section 5).

2 Générateurs électriques de secours au Nigéria

2.1 Éléments de contexte sur le Nigéria

Situé dans la partie occidentale de l'Afrique, qui borde le golfe de Guinée, le Nigéria est le pays le plus peuplé du continent et la première économie africaine, avec une population estimée à plus de 206 millions d'habitants en 2020 (United Nations, 2019) et un PIB de 432 milliards de dollars en 2020 (World Bank, 2022).

Le pays est le premier producteur de brut du continent et son économie est structurellement dépendante de l'exploitation du pétrole. Pour assurer une croissance multidimensionnelle et diminuer sa dépendance au secteur extractif, le développement des infrastructures, en particulier de l'électricité, est considéré comme essentiel (Adesanya and Schelly, 2019; International Monetary Fund, 2022). Le développement du secteur électrique nigérian est toutefois confronté à de nombreux défis.

Le secteur électrique national cumule de nombreux problèmes : problèmes de maintenance des centrales électriques et des infrastructures réseau, pertes techniques et non techniques particulièrement élevées, et pannes de courant régulières et prolongées. Les pannes d'électricité surviennent plus de 45 % du temps dans le pays. Elles ont lieu le plus souvent en milieu de journée, lorsque les entreprises et les industries ont le plus besoin d'électricité (Babajide and Brito, 2021).

En conséquence, les secteurs entreprises du pays utilisent des générateurs auxiliaires pour produire et consommer de l'électricité lorsque le réseau est en panne. La capacité totale installée des générateurs auxiliaires dans le pays est estimée entre 10 et 15 GW, soit deux à trois fois plus que la capacité installée centralisée réellement disponible sur le réseau national (Arik, 2019; Arowolo and Perez, 2020).

Du fait de la difficulté à obtenir des données dans la région, il existe peu de littérature sur l'utilisation de générateurs de secours par des agents confrontés à un réseau peu fiable. En conséquence, les comportements d'utilisation des générateurs auxiliaires vis-à-vis des états du réseau sont peu connus. Ce projet de recherche vient combler ce manque en analysant une base de données originale sur l'utilisation de générateurs auxiliaires par des commerces

connectés au réseau nigérian.

2.2 Présentation des données

Les données utilisées sont des séries temporelles de consommations d'électricité de générateurs auxiliaires d'entreprises nigérianes. Elles proviennent de 210 compteurs intelligents installés par l'A2EI (Access To Energy Institute) sur des générateurs auxiliaires (196 sur 210) et sur le réseau (14 sur 210) dans différentes régions du Nigéria. Ces compteurs mesurent en continu, à intervalles d'une à six minutes : la tension (V), le courant (A), le facteur de puissance, la fréquence (Hz), la puissance (kW) et la consommation cumulée (kWh). Chaque compteur a son propre identifiant et indique si le compteur a été installé sur un générateur ou sur le réseau, la capacité du générateur, le type d'entreprise à laquelle le générateur est associé, la région dans laquelle le compteur est installé et ses coordonnées GPS.

Les compteurs installés sur le réseau ne couvrent malheureusement pas toutes les zones où sont présents des générateurs. En raison de la grande hétérogénéité de la fiabilité de l'électricité fournie par le réseau central qui existe entre les différentes localités du Nigéria, nous sélectionnons donc les générateurs qui se trouvent à une distance maximale de 1 km d'un compteur installé sur le réseau. Cela doit nous permettre de nous assurer que nous comparons bien l'utilisation des générateurs en fonction des états du réseau afférents.

Nous avons donc dû restreindre notre analyse à 68 générateurs dans 5 sous-groupes régionaux où nous disposons à la fois de données sur l'état de fonctionnement du réseau central et de données sur l'utilisation des générateurs auxiliaires. Nous segmentons chaque journée en tranches de 10 minutes afin d'homogénéiser nos séries temporelles et de les rendre comparables¹.

Pour récapituler, nous procédons aux simplifications suivantes :

1. nous regroupons les compteurs des générateurs et des réseaux en 5 groupes géographiques où la distance maximale est de 1 000 mètres entre deux compteurs et où nous avons des données réseau ;
2. nous segmentons chaque jour en tranches de 10 minutes et prenons la moyenne des variables ;
3. nous simplifions notre série temporelle en donnant à un générateur ou à un compteur réseau la valeur 1 lorsqu'il est « allumé » et 0 dans le cas contraire.

3 Chaînes de Markov d'ordre variable

Chaque générateur est donc décrit par une série temporelle binaire $(X_i^c)_{i \geq 1}$ dont le comportement est piloté *a priori* par les préférences de l'entreprise concernée et par la qualité de l'électricité fournie par le réseau. Nous proposons d'utiliser des chaînes de Markov d'ordre

1. Dans la base de données brute, les générateurs peuvent envoyer des informations à des intervalles de temps différents en fonction du réglage qui a été effectué par l'opérateur. Certains générateurs envoient par exemple des informations toutes les 2 minutes, tandis que d'autres le font toutes les 6 minutes.

variable pour modéliser ces séries. Nous rappelons dans cette section les deux principaux modèles de ce type.

Dans la suite de l'article, pour toute séquence $(t_i)_i$ indexée sur \mathbb{N} ou \mathbb{Z} , on note t_i^j la séquence $(t_i, t_{i+1}, \dots, t_j)$, avec $i < j$. En particulier $t_{-\infty}^i$ désigne la séquence infinie à gauche.

3.1 Les chaînes de Markov d'ordre variable

Les VLMC (pour *Variable Length Markov Chain*) ont été proposés par [Rissanen \(1983\)](#) et développés notamment par [Bühlmann and Wyner \(1999\)](#). On peut les voir comme des chaînes de Markov d'ordre supérieur parcimonieuses.

Soit S un ensemble fini et S^∞ l'ensemble des séquences de longueurs arbitraires sur S . Une série temporelle $(X_i)_{i \in \mathbb{Z}}$ indexée par les entiers relatifs et à valeurs dans S est un VLMC s'il existe une fonction l de S^∞ dans $\{0, \dots, l_{\max}\}$ telle que pour tout t et toute séquence $x_{-\infty}^t$

$$\mathbb{P}(X_t = x_t \mid X_{-\infty}^{t-1} = x_{-\infty}^{t-1}) = \mathbb{P}\left(X_t = x_t \mid X_{t-l(x_{-\infty}^t)}^{t-1} = x_{t-l(x_{-\infty}^t)}^{t-1}\right). \quad (1)$$

La série est donc une chaîne de Markov d'ordre l_{\max} mais dont la mémoire est dépendante du contexte. On la caractérise par la fonction de contexte c de S^∞ dans lui-même qui a un passé quelconque $x_{-\infty}^t$ associe $x_{t-l(x_{-\infty}^t)}^{t-1}$. La propriété ci-dessus indique que les probabilités de transition de la chaîne dépendent seulement de $c(x_{-\infty}^t)$, une mémoire de longueur variable et inférieure ou égale à l_{\max} . Au lieu de spécifier S_{\max}^l lois décrivant les transitions de la chaîne, on se contente d'autant de lois qu'il y a de valeurs différentes produites par la fonction de contexte (ces valeurs sont les contextes du modèle). Il est ainsi possible d'observer des dépendances temporelles longues conjointement à des dépendances courtes.

L'algorithme dit de « contexte » proposé par [Rissanen \(1983\)](#) permet d'estimer la fonction de contexte c (et donc l) ainsi que les probabilités de transition à partir d'une série temporelle de longueur finie. L'algorithme utilise deux paramètres : le nombre minimal de répétitions d'une séquence finie dans la série pour qu'elle puisse être considérée comme un contexte valide (c'est-à-dire une valeur de $c(x_{-\infty}^t)$) et un paramètre de coupe qui rejette les contextes qui n'apportent pas assez d'information (au sens d'un test de rapport de vraisemblance). En pratique, le nombre minimal de répétitions est fixée à une valeur faible (2 par exemple). Le paramètre de coupe est généralement choisi en optimisant une vraisemblance pénalisée comme l'AIC ou le BIC [Mächler and Bühlmann \(2004\)](#). On peut aussi chercher à choisir directement le modèle en maximisant une vraisemblance pénalisée, comme proposé dans [Csiszar and Talata \(2006\)](#), ce qui conduit à un estimateur consistant avec le BIC [Garivier \(2006\)](#); [Csiszar and Talata \(2006\)](#). Les deux approches sont unifiées dans [Garivier and Leonardi \(2011\)](#).

3.2 Prise en compte de covariables

Le manque de fiabilité du réseau électrique se traduit par un comportement « aléatoire » de la fourniture de courant. Comme on suppose que les entreprises qui utilisent de façon limitée leur générateur le font en réaction aux défaillances du réseau, un modèle stationnaire

de leur comportement seul ne semble pas approprié. Pour pallier cette limitation des VLMC, [Zanin Zambom et al. \(2022\)](#) ont proposé une extension de ceux-ci qui prend en compte des covariables exogènes.

En plus de la série temporelle étudiée, $(X_i)_{i \in \mathbb{Z}}$, on suppose donnée une série $(Y_i)_{i \in \mathbb{Z}}$ à valeurs dans $\prod_{k=1}^K \mathcal{Y}_k$ où chaque \mathcal{Y}_k est soit \mathbb{R} soit un ensemble fini. On suppose qu'il existe deux fonctions l et h , de S^∞ dans $\{0, \dots, l_{\max}\}$, telles que $h(s^\infty) \leq l(s^\infty)$ pour tout $s \in S^\infty$ et que pour tout t et toutes séquences $x_{-\infty}^t$ et $y_{-\infty}^t$

$$\begin{aligned} \mathbb{P}(X_t = s_0 \mid X_{-\infty}^{t-1} = x_{-\infty}^{t-1}, Y_{-\infty}^{t-1} = y_{-\infty}^{t-1}) = \\ \mathbb{P}\left(X_t = x_t \mid X_{t-l(x_{-\infty}^t)}^{t-1} = x_{t-l(x_{-\infty}^t)}^{t-1}, Y_{t-h(x_{-\infty}^t)}^{t-1} = y_{t-h(x_{-\infty}^t)}^{t-1}\right). \end{aligned} \quad (2)$$

$(X_i)_{i \in \mathbb{Z}}$ suit alors un modèle COVLMC (pour *Variable Length Markov Chain with exogeneous COvariates*) par rapport à $(Y_i)_{i \in \mathbb{Z}}$.

La comparaison avec l'équation (1) fait apparaître plusieurs différences importantes. Tout d'abord seule la série temporelle discrète $(X_i)_{i \in \mathbb{Z}}$ est modélisée, $(Y_i)_{i \in \mathbb{Z}}$ est seulement observée. De plus les deux fonctions de longueur de mémoire, l et h utilisent le même passé, celui de la série modélisée, pas celui de la série des covariables. Enfin, la mémoire sur les covariables peut être plus courte que celle sur la série modélisée.

Comme dans [Zanin Zambom et al. \(2022\)](#), nous représentons

$$\mathbb{P}\left(X_t = x_t \mid X_{t-l(x_{-\infty}^t)}^{t-1} = x_{t-l(x_{-\infty}^t)}^{t-1}, Y_{t-h(x_{-\infty}^t)}^{t-1} = y_{t-h(x_{-\infty}^t)}^{t-1}\right)$$

par une régression logistique² pour chaque contexte $c(x_{-\infty}^t)$. En d'autres termes, les probabilités de transition dans un contexte donné sont calculées par une régression logistique appliquée aux covariables $Y_{t-h(x_{-\infty}^t)}^{t-1}$.

[Zanin Zambom et al. \(2022\)](#) proposent et analysent les propriétés asymptotiques d'une variante de l'algorithme de contexte adapté aux COVLMC. Une particularité de l'algorithme est qu'il règle la mémoire de covariables h pour chaque contexte accepté. Comme pour les VLMC, l'algorithme utilise un paramètre de nombre d'observations minimales et un paramètre de coupe. Pour garantir une estimation acceptable d'une régression logistique pour chaque contexte, le nombre d'observations minimales doit être beaucoup plus élevé que dans le cas du VLMC. Il doit aussi croître avec le nombre de paramètres des régressions. En pratique, les auteurs proposent de ne considérer que les contextes qui apparaissent au moins s fois par paramètre à estimer dans la régression. Le paramètre de coupe et s sont réglés en maximisant le BIC du modèle.

Dans notre contexte applicatif, les covariables représentent l'état du réseau et le moment de la journée. Nous attendons qu'une entreprise sensible à la qualité du réseau présente une mémoire non nulle par rapport aux covariables.

2. multinomiale dans le cas général.

3.3 Implémentation

Les VLMC sont notamment implémentés dans deux *packages* R, VMLC³ de Mächler and Bühlmann (2004) et PST⁴ de Gabadinho and Ritschard (2016). Il n'existe pas à notre connaissance d'implémentation publique des COVLMC. Nous proposons donc un nouveau *package* *mixvlmc*⁵. Outre l'accès aux COVLMC, il propose un mécanisme de post-élagage : après l'estimation d'un (CO)VLMC pour une valeur donnée du paramètre de coupe, nous estimons un ensemble de valeurs qui garantissent une réduction de la complexité du modèle (toutes les réductions possibles dans le cas des VLMC). Cela permet de sélectionner efficacement ce paramètre au moyen d'un critère informationnel, sans avoir besoin de réajuster un nouveau modèle.

4 Étude du comportement propre de chaque générateur

Dans un premier temps, nous ajustons donc à chaque générateur étudié un VLMC peu contraint (le paramètre de coupe correspond à un test au niveau 10 %), sur l'espace d'état $\{0, 1\}$ (0 pour le générateur éteint et 1 pour l'état allumé). Le modèle initial est ensuite élagué pour retenir le modèle optimal au sens du BIC. La figure 1 représente le nombre de contextes différents et la séquence constante la plus longue dans un contexte en fonction de l'ordre maximal de chaque générateur.

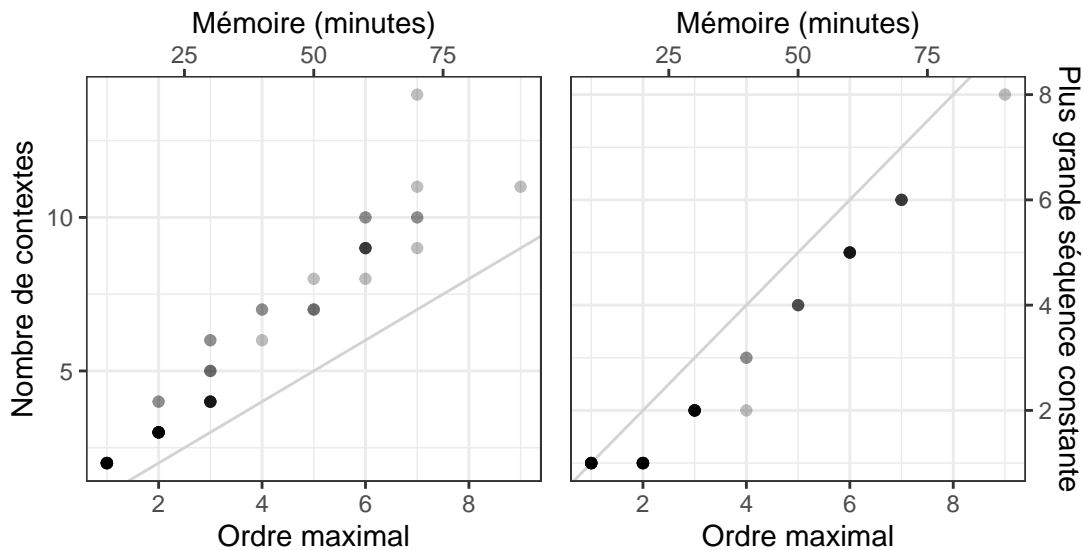


FIGURE 1 – Nombre de contextes (à gauche) et séquence constante la plus longue (à droite) en fonction de l'ordre maximal pour les VLMC ajustés sur les générateurs

3. <https://cran.r-project.org/package=VMLC>
4. <https://cran.r-project.org/package=PST>
5. <https://github.com/fabrice-rossi/mixvlmc>

Comme attendu, les modèles avec un ordre élevé sont parcimonieux : l'ordre maximal de 9 correspond à $2^9 = 512$ contextes potentiels alors que seuls 11 sont retenus dans ce cas. D'autre part, on constate que dans presque tous les cas, le contexte le plus long correspond à un état constant sur tout le contexte (éteint ou allumé) suivi ou précédé par l'état contraire. Par exemple le compteur avec l'ordre le plus grand (9), est caractérisé par des contextes de la forme $(1, 0, 0, \dots, 0)$ avec une probabilité croissante d'extinction en fonction de la longueur du contexte et un contexte de longueur 8 ne contenant que des états éteints avec une grande probabilité d'extinction (0.98). En termes concrets, cela signifie que la probabilité de s'allumer est calculée en cherchant jusqu'à 9 états dans le passé le premier instant auquel le générateur était allumé. Plus cet état est ancien, plus la probabilité de se rallumer est faible mais elle reste toujours plus grande que si le générateur est resté éteint depuis 1h20 (8 états).

Les modèles complexes peuvent donc rendre compte de comportements relativement logiques comme une tendance au rallumage du générateur si celui-ci a été utilisé récemment, une tendance à garder le générateur allumé un temps minimum, etc. Cependant, les générateurs avec un modèle complexe sont peu nombreux puisque 49 d'entre eux ont un ordre maximal de 3 au plus. Pour 18 d'entre eux, l'ordre est même de 1. On peut noter aussi, comme le montre la figure 2 que la complexité des modèles obtenus ne peut pas se déduire simplement de statistiques simples calculées sur la série temporelle d'origine.

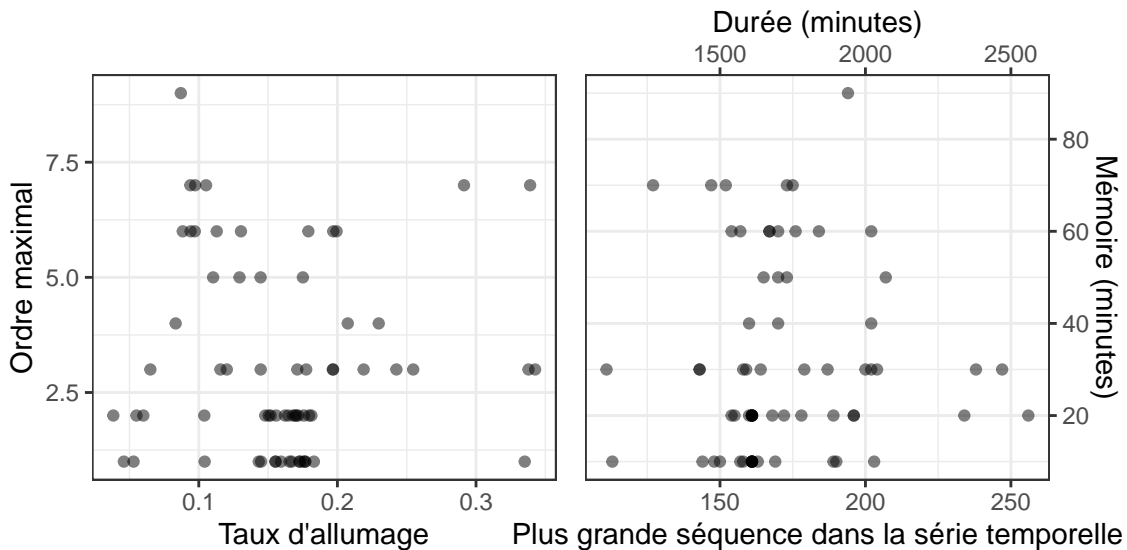


FIGURE 2 – Ordre du VLMC sélection par BIC en fonction du taux d'allumage du générateur (à gauche) et de la longueur de la plus grande séquence constante dans la série temporelle d'origine (à droite).

L'analyse détaillée des modèles obtenus peut permettre de caractériser partiellement le comportement des entreprises mais le nombre élevé de simples modèles markoviens d'ordre 1 ou 2 ne donne pas l'espoir de pouvoir capturer finement des classes de comportement. Il semble indispensable de tenir compte de l'état du réseau centralisé pour comprendre les attitudes des entreprises.

5 Étude de l'influence des covariables

Les modèles COVLMC utilisés dans notre analyse incorporent deux covariables : une covariable représentant l'état du réseau à chaque instant de la journée, représenté soit sous la forme d'un indicateur binaire (éteint ou allumé), soit sous la forme d'un indicateur représentant la qualité de l'électricité fournie par le réseau central (éteint, normal, sous-voltage et sur-voltage), et un indicateur binaire indiquant la période de la journée, « jour » entre 6h00 et 19h30 et « nuit » sinon ⁶.

De façon analogue au VLMC, nous ajustons à chaque générateur étudié un COVLMC peu contraint avec un paramètre de coupe qui correspond à un test au niveau 10 %. Ce premier modèle est ensuite élagué pour retenir un modèle plus simple et optimal au sens du critère BIC. Quinze modèles de COVLMC sont ajustés sur chaque générateur avec les différentes combinaisons de covariables et du paramètre s . On conserve ensuite la meilleure combinaison au sens du BIC (cf la table 1 pour la répartition des combinaisons retenues en terme de covariables). On constate que pour tous les générateurs, le modèle COVLMC obtenu a un meilleur BIC que celui du meilleur VLMC.

Covariables	Réseau détaillé	Réseau simple	Réseau simple et jour/nuit	Jour/nuit
Nombre	1	25	15	17

TABLE 1 – Nombre de générateurs par combinaisons de covariables

D'autre part, la figure 3 illustre une limitation importante des VLMC et confirme l'intérêt de prendre en compte les covariables. On voit en effet que les générateurs dont le comportement est influencé par l'état du réseau seulement correspondent comme attendu à des ordres faibles pour le VLMC associé. Par contre, la réciproque n'est pas vraie : le fait que le VLMC d'un générateur possède un ordre faible n'est pas prédictif de sa dépendance à l'instant de la journée ou au réseau.

Outre le classement des générateurs en grandes catégories de comportement en fonction des covariables retenues, l'utilisation des COVLMC apporte des gains descriptifs variés selon les générateurs. Nous pouvons citer l'exemple d'un générateur avec un VLMC optimal d'ordre 6, comportant 8 contextes qui reviennent essentiellement à chercher la dernière occurrence d'un allumage (de façon similaire à l'exemple d'ordre 9 décrit précédemment). Le COVLMC correspondant utilise les covariables jour/nuit et réseau (binaire), pour un modèle d'ordre 2 seulement et avec 4 contextes très simples (les 4 possibilités dans cette situation). Les contextes qui utilisent les covariables ont un comportement facile à interpréter : le fait d'être la nuit diminue fortement la probabilité d'allumage du générateur. De la même façon, le fait que le réseau central fournisse du courant réduit la probabilité d'allumage. On observe aussi une tendance à maintenir le générateur allumé quand il l'est déjà. De façon plus générale, l'analyse des coefficients des modèles logistiques permet de distinguer des classes de comportement

6. Cela tient compte des caractéristiques particulières des pays situés sous l'équateur, où le soleil se couche à un angle plus prononcé, ce qui entraîne des levers et couchés du jour plus rapide et où les heures de journée sont relativement stables tout au long de l'année.

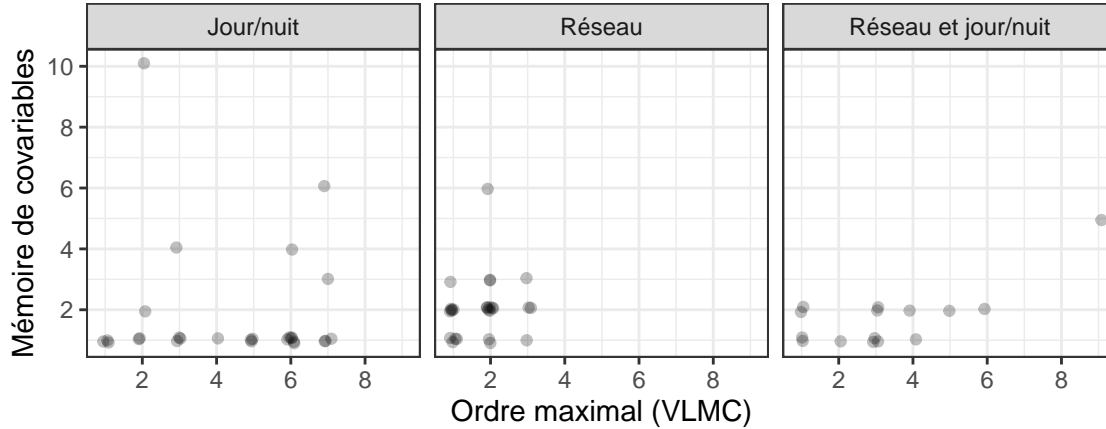


FIGURE 3 – Memoire maximale en covariables du COVLMC pour chaque générateur en fonction de l’ordre maximal du VLMLC retenu pour ce générateur. Les générateurs sont répartis selon la combinaison de covariable retenue.

plus fines, comme par exemple les entreprises qui ont besoin de courant la nuit par opposition à celles qui tendrons à éteindre systématiquement leur générateur la nuit.

6 Conclusion et perspectives

Nous avons présenté dans ce travail les premières étapes d’une analyse exploratoire du comportement d’entreprises nigérianes confrontées à une fourniture d’électricité instable et de qualité médiocre. Les séries temporelles décrivant l’allumage et l’extinction des générateurs utilisés par ces entreprises ne peuvent pas être modélisées de façon satisfaisante avec des chaînes de Markov d’ordre variable. Il est indispensable de tenir compte de l’influence de l’état du réseau électrique central et de la période de la journée pour obtenir des modèles plus réalistes. On constate alors que les comportements des entreprises sont très variables entre une sensibilité forte à l’état du réseau central et une utilisation basée plutôt sur l’heure de la journée exclusivement.

Pour aller plus loin dans l’analyse, il est semble nécessaire de prolonger le développement des modèles de chaînes de Markov d’ordre variable avec covariables. En effet, l’estimation du modèle proposée dans [Zanin Zambom et al. \(2022\)](#) n’intègre pas la sélection des covariables importantes mais seulement de l’ordre du modèle relativement à l’historique des covariables. De ce fait, les modèles logistiques associés à chaque contexte peuvent conserver plus de covariables que nécessaire et conduire en conséquence à un rejet du contexte correspondant.

Bibliographie

A. A. Adesanya and C. Schelly. Solar pv-diesel hybrid systems for the nigerian private sector : An impact assessment. *Energy Policy*, 132 :196–207, 2019.

- E. Arik. Le marché des groupes électrogènes dans les Suds. Research report, Agence Française de Développement ; CERI Sc Po Paris, 2019.
- W. Arowolo and Y. Perez. Market reform in the nigeria power sector : A review of the issues and potential solutions. *Energy Policy*, 144 :111580, 2020.
- A. Babajide and M. C. Brito. Solar pv systems to eliminate or reduce the use of diesel generators at no additional cost : A case study of lagos, nigeria. *Renewable Energy*, 172 : 209–218, 2021.
- P. Bühlmann and A. J. Wyner. Variable length markov chains. *The Annals of Statistics*, 27 (2) :480–513, April 1999. doi : 10.1214/aos/1018031204.
- I. Csiszar and Z. Talata. Context tree estimation for not necessarily finite memory processes, via bic and mdl. *IEEE Transactions on Information Theory*, 52(3) :1007–1016, March 2006. ISSN 1557-9654. doi : 10.1109/TIT.2005.864431.
- A. Gabadinho and G. Ritschard. Analyzing state sequences with probabilistic suffix trees : The PST R package. *Journal of Statistical Software*, 72(3) :1–39, 2016. doi : 10.18637/jss.v072.i03.
- A. Garivier. Consistency of the unlimited bic context tree estimator. *IEEE Transactions on Information Theory*, 52(10) :4630–4635, Oct 2006. ISSN 1557-9654. doi : 10.1109/TIT.2006.881742.
- A. Garivier and F. Leonardi. Context tree selection : A unifying view. *Stochastic Processes and their Applications*, 121(11) :2488–2506, 2011. ISSN 0304-4149. doi : <https://doi.org/10.1016/j.spa.2011.06.012>.
- International Monetary Fund. Nigeria : 2021 Article IV Consultation-Press Release ; Staff Report ; Staff Statement, and Statement by the Executive Director for Nigeria, Country Report No. 2022/033. report, International Monetary Fund, 2022.
- M. Mächler and P. Bühlmann. Variable length markov chains : Methodology, computing, and software. *Journal of Computational and Graphical Statistics*, 13(2) :435–455, 2004. doi : 10.1198/1061860043524.
- J. Rissanen. A universal data compression system. *IEEE Transactions on Information Theory*, 29(5) :656–664, Sep. 1983. ISSN 1557-9654. doi : 10.1109/TIT.1983.1056741.
- United Nations. World Population Prospects 2019. Database, United Nations, 2019. URL <https://population.un.org/wpp/>. Accessed : 2022-05-12.
- World Bank. World Development Indicators. Database, World Bank, 2022. URL <https://databank.worldbank.org/source/world-development-indicators>. Accessed : 2022-05-11.
- A. Zanin Zambom, S. Kim, and N. Lopes Garcia. Variable length markov chain with exogenous covariates. *Journal of Time Series Analysis*, 43(2) :312–328, 2022. doi : 10.1111/jtsa.12615.