

Caractérisation statistique des erreurs de prédiction de production d'énergie renouvelable

Enzo Le Bouëdec

28 août 2018

Table des matières

1	Introduction	4
1.1	Présentation du Laboratoire	4
1.2	Contexte	5
2	Présentation des données	7
2.1	Production Éolienne	7
2.2	Production Photovoltaïques	9
3	Modèles et métriques	12
3.1	Modèles de Markov cachés	12
3.1.1	Chaîne de Markov	12
3.1.2	Modèle de Markov cachés	13
3.1.3	Modèles Markov Switching Auto-Regressive	20
3.2	Modèles GARCH	21
3.3	K-moyennes (Kmeans)	21
3.4	Modèle de mélange gaussien (GMM)	22
3.5	Analyse par Composante Principale	23
3.6	Métriques	23
3.6.1	BIC	23
3.6.2	Erreur moyenne absolue (MAE)	23
3.6.3	Erreur quadratique moyenne (RMSE)	23
4	Résultats pour la production éolienne	25
5	Application aux erreurs de prévisions photovoltaïques	26
5.1	Méthode des plus proches voisins	27

5.2 Résultats avec un GMM	30
5.3 Résultats avec un HMM	34
A Première version de la publication concernant les erreurs de prévisions éoliennes	39

Liste des symboles

π	Probabilités initiales d'appartenance à l'un ou l'autre des états cachés (HMM et MSAR)
Γ	Matrice de transition entre les états cachés (HMM et MSAR)
σ	Variance des densités de probabilités Gaussiennes de dimension $d \times d$ (HMM et MSAR)
\hat{d}	Indice de la journée indiquant l'heure de la correction
a_0	Moyenne des densités de probabilités (HMM) ou coefficient constant du processus auto-regréssif (MSAR)
d	Dimension de l'espace considéré
m	Nombre d'états cachés (HMM et MSAR)
ACP	Analyse par Composante Principale
BIAS	Erreur moyenne
BIC	Bayesian Information Criterion
GMM	Modèle de mélange gaussien
HMM	Modèle de Markov Caché (3.1.2)
Kmeans	Méthode de partitionnement K-moyennes
MAE	Erreur moyenne absolue
MS-AR	Modèle Markov Switching Auto Regressive (3.1.3)
RMSE	Erreur quadratique moyenne

CHAPITRE 1

Introduction

1.1 Présentation du Laboratoire

J'ai réalisé mon stage au sein du laboratoire Systèmes et Applications des Technologies de l'Information et de l'Énergie (SATIE) qui est une unité mixte de recherche du CNRS comptant environ 200 personnes. Ses tutelles sont le CNRS, l'ENS Paris Saclay, l'ENS Rennes, le CNAM Paris, l'université de Cergy Pontoise, l'université de Paris Sud et l'IFFSTAR. Ce stage a été encadré par Roman Le Goff Latimier, Hamid Ben Hamed et Bernard Multon qui font partie de l'ENS de Rennes, lieu dans lequel s'est déroulé le stage. Le laboratoire SATIE comporte de nombreuses équipes de recherches dont les thématiques dérivent toutes de problématiques liées à l'*Electrical Engineering*. Les travaux que j'ai eu à réaliser s'intégraient dans les thématiques de l'équipe Système d'Énergies pour les Transports et l'Environnement. L'objectif principal de cette équipe est de mettre au point de nouveaux systèmes (ou des les améliorer) de création, de transport ou de conversion d'énergie afin d'augmenter leur efficacité tout en minimisant l'impact environnemental.

1.2 Contexte

La part des énergies renouvelables à production variable et **incertaine** -notamment l'éolien et le photovoltaïque- est en augmentation constante au cours des dernières années comme le montre la figure 1.1. En effet leur part dans la production d'énergie renouvelable passe d'une quantité négligeable au début des années 2000 à environ 20% en 2016.

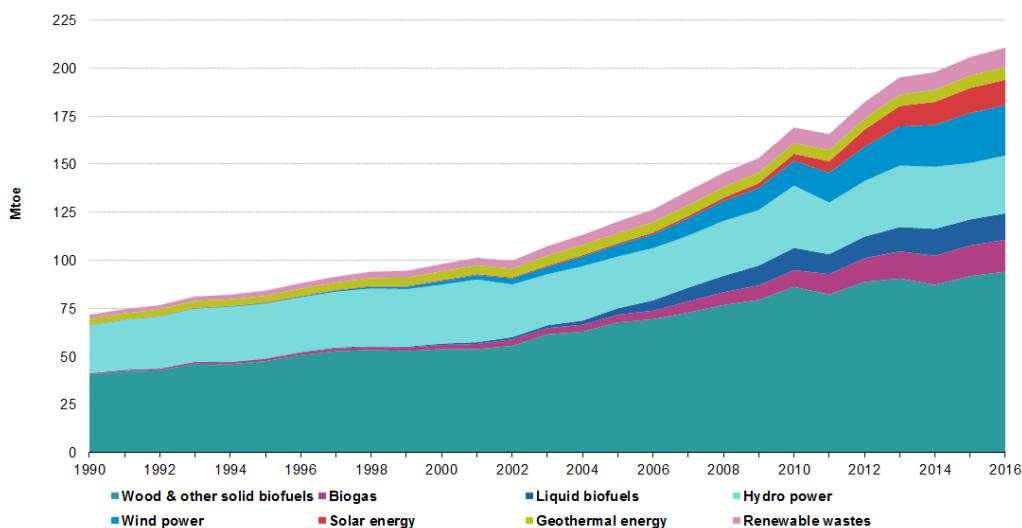


FIGURE 1.1 – Production d'énergie primaire à partir de sources renouvelables en Europe de 1990 à 2016 d'après les chiffres d'Eurostat

Cette intégration croissante des énergies renouvelables à production incertaine demande d'être faite avec minutie car elle impose de nouvelles formes d'incertitudes.

Afin de comprendre pourquoi, il est nécessaire de faire un point sur le réseau électrique Français, et par extension, Européen. Ces réseaux se sont construits autour de l'impératif d'être en mesure de fournir, à n'importe quel instant, la puissance demandée par les utilisateurs. Or comme l'énergie qui circule sur les réseaux électriques est un flux instantané, une contrainte physique impose que demande et production soient égales à tout instant.

Les sources de productions d'énergies d'alors étaient principalement délocalisées, constantes et avec une faible inertie (centrales thermiques) ou forte inertie (centrale nucléaires). Contraintes auxquelles il faut ajouter la difficulté de stockage de l'électricité ainsi que des infrastructures qui limitent la puissance

de certains échanges. Pour pallier à ces difficultés et contraintes, plusieurs systèmes correspondant à différentes échelles de temps ont été mis en place.

L'essentiel de l'organisation pour égaliser production et consommation aux temps moyens et courts réside dans la mise en place de différents marchés. Un premier ,le marché *day-ahead*, sur lequel des accords non engageants sont pris pour le lendemain basés sur les particularismes de la journée (vague de froid, jour férié,... etc). Un autre marché a lieu au cours de la journée où des offres sont effectuées dès qu'elles sont compatibles. Cependant des écarts plus ou moins grands par rapport aux engagements ont lieux. Le gestionnaire de réseau fait alors appel à un second marché plus onéreux, dit marché de *balancing*, pour combler ces écarts. En fonction des législations, les écarts par rapports aux engagements sur les différents marchés peuvent être soumis à des pénalisations.

Dans ce contexte, de nombreuses études avaient été réalisées sur la prévision de la consommation des ménages qui était à l'époque la principale source d'incertitude. Avec l'arrivée des énergies renouvelables à production incertaine, un important effort de recherche a été orienté autour de la prévision de production. Dans le cas de l'éolien et du photovoltaïque, ces prévisions sont souvent basées sur des prévisions météorologiques.

Étudier les erreurs de prévisions prend son sens au travers de plusieurs éléments. Tout d'abord, le fait que les prévisions de production soient basées sur des prévisions météorologiques empêche un rafraîchissement continu des prévisions. En effet, les calculs météorologiques étant particulièrement lourds, ils ne sont réalisés que quelques fois par jour. Déterminer une dynamique dans les erreurs de prévisions permettrait donc d'affiner les prévisions de production entre deux mises à jour successives des calculs météorologiques. Enfin, au vu de la conception des marchés, ce sont bien les erreurs de prévisions qui doivent être compensés sur le second marché et qui peuvent perturber le réseau côté gestionnaire ou donner lieu à des amendes côté producteur. C'est donc bien, in fine, cette valeur qu'il est impératif de minimiser. Caractériser les erreurs de prévision est donc une approche complémentaire et une seconde source d'information valable et pertinente.

CHAPITRE 2

Présentation des données

2.1 Production Éolienne

L'étude est réalisée dans un premier temps sur des données issue de Bonneville Power Administration¹ entre 2009 et 2017. Ces données comportent, par pas de temps de 5 minutes, les productions électriques éoliennes réelles ainsi que les prévisions à 1 heure annoncées par les producteurs.

Afin de réduire la taille du jeu de données, il est réduit à un échantillonnage horaire. Pour ce faire la moyenne sur l'heure suivante est réalisée.

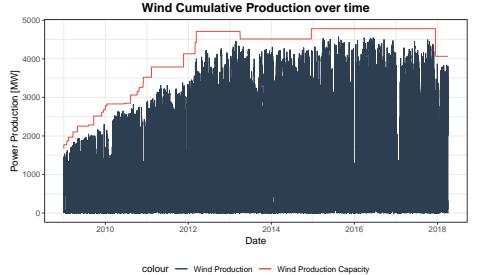
D'autre part, comme le nombre d'installations de production a évolué du simple au double (voir ici²) sur cette période, les données ont été normalisées par la capacité de production maximale (voir 2.1a). Les erreurs de prévisions sont donc reconstituées à partir de ces données en soustrayant les productions mesurées aux productions prévues (voir 2.1b).

La figure 2.2a présente l'histogramme des erreurs normalisées en échelle logarithmique. On constate que les erreurs sont bien centrées autour de zéro et qu'une grande partie réside dans l'intervalle $[-0.2, 0.2]$ ce qui témoigne d'une relativement bonne capacité de prévision. La figure 2.2b représente

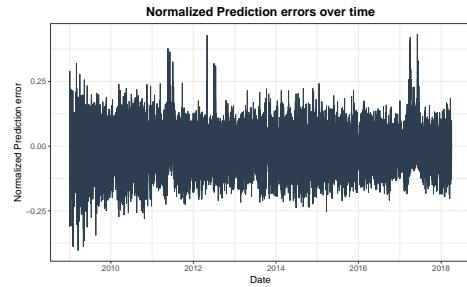
Préciser combien

1. <https://transmission.bpa.gov/Business/Operations/Wind/>

2. https://transmission.bpa.gov/business/operations/Wind/WIND_InstalledCapacity_Plot.pdf



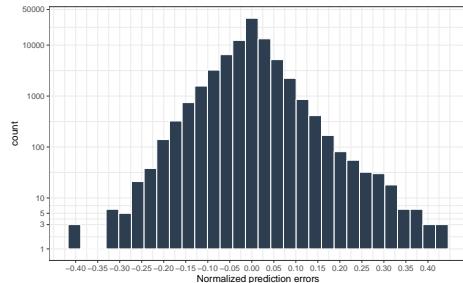
(a) Production réalisée et capacité de production installée



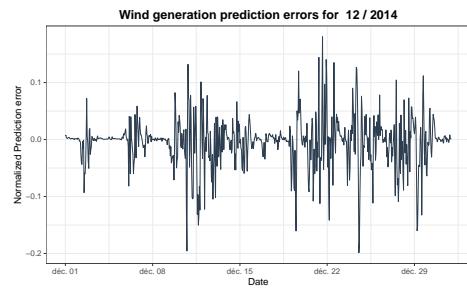
(b) Erreurs de prévisions normalisées

FIGURE 2.1 – Représentation du jeu de donnée utilisé pour l'étude sur la période allant de Janvier 2009 jusqu'à Mai 2018

quand à elle un zoom sur la série temporelle des erreurs de prévisions. Il est intéressant de noter que la série présente des périodes sur lesquelles la variance est relativement faible (comme autour du 3 ou après le 8 décembre) et d'autre où la variance est nettement plus élevée (comme sur la deuxième partie du mois).



(a) Histogramme des erreurs de prévisions



(b) Erreurs de prévisions sur un mois

FIGURE 2.2 – Représentation du jeu de donnée utilisé pour l'étude

2.2 Production Photovoltaïques

Des données agrégées de production d'énergie photovoltaïque ainsi que les prévisions $J - 1$ correspondantes sont mises à disposition par TenneT sur leur site web³. Les données fournies sur la Bavière ont été utilisées pour cette étude. Ces données sont libres d'accès et ouvertes à la publication scientifique. Elles vont du 1^{er} décembre 2010 jusqu'à aujourd'hui avec un pas d'échantillonnage de 15 minutes (voir figure 2.3). Aucune covariable n'est mise à disposition. N'ayant pas eu accès à la capacité de production installée en Bavière sur cette période, ces données sont normalisées par des maxima de prévisions/productions sur certaines périodes. Cette capacité de production installée estimée est aussi représentée sur la figure 2.3. Les données manquantes (190 pour les prévisions et 69 pour les réalisations) sont remplacées par la dernière valeur observée temporellement parlant.

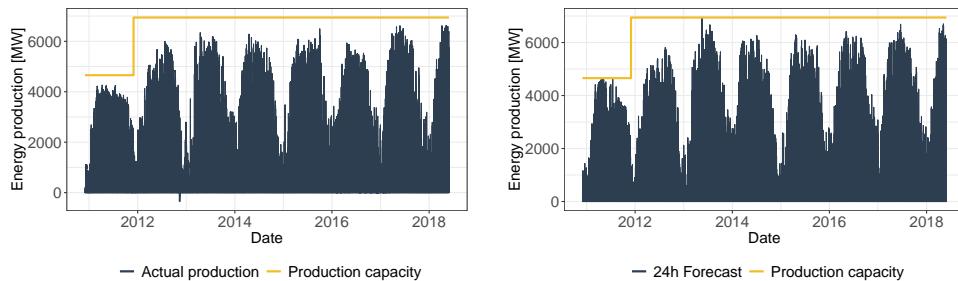


FIGURE 2.3 – Production réelle (à gauche) et prévue (à droite) d'énergie photovoltaïque en Bavière entre 2010 et aujourd'hui

La série temporelle des erreurs de prévisions alors obtenue est présentée sur la figure 2.4. Comme l'illustre la partie zoomée, la structure de ces données est particulière avec une alternance de plages à valeurs négligeables et non négligeables.

Cette structure correspond en fait à une saisonnalité quotidienne comme le montre le diagramme en boîte de la figure 2.5. Les erreurs de prévisions entre 20h et 4h du matin sont systématiquement nulles. Elles correspondent aux plages horaires pendant lesquelles le soleil ne brille jamais en Bavière.

3. <https://www.tennet.eu/electricity-market/transparency-pages/transparency-germany/>

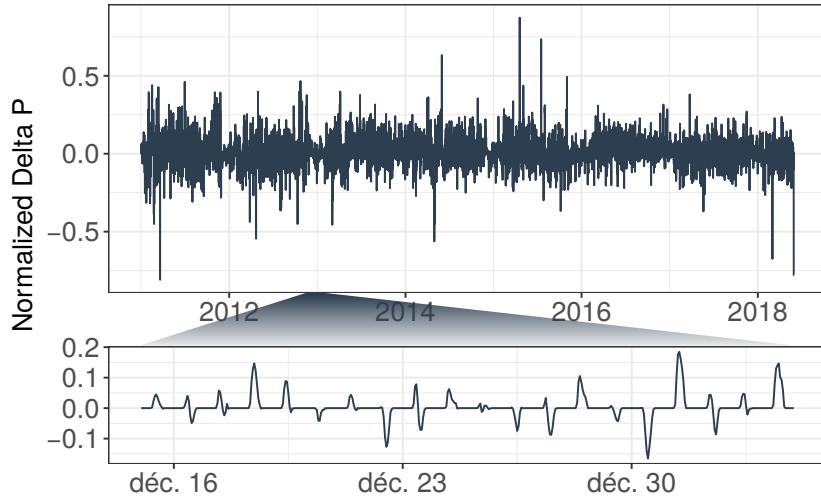


FIGURE 2.4 – Erreur de prévision de production d'énergie solaire sur la Bavière

En effet, à Munich (capitale de la Bavière), le soleil s'est levé à 5h12 et s'est couché à 21h17 le 21 Juin 2018. Sur ces périodes, aucune production d'énergie n'est prévue ni réalisée, et donc les erreurs sont systématiquement nulles. Il est donc logique que les erreurs moyennes jusqu'à 5h du matin soient nulles. Cependant, au vu de cette amplitude d'ensoleillement maximale, il est possible qu'un autre phénomène réduise systématiquement les erreurs de prédiction à 20h et 21h pour lesquelles cette analyse n'explique pas d'aussi faible valeurs. L'histogramme aussi présenté sur cette figure illustre aussi cette surreprésentation des valeurs très faible. D'autre part, il montre que la grande majorité des erreurs de prévisions moyennées sont assez faibles. Par exemple, 90% des données sont incluses dans l'intervalle $[-0.688, 0.0891]$.

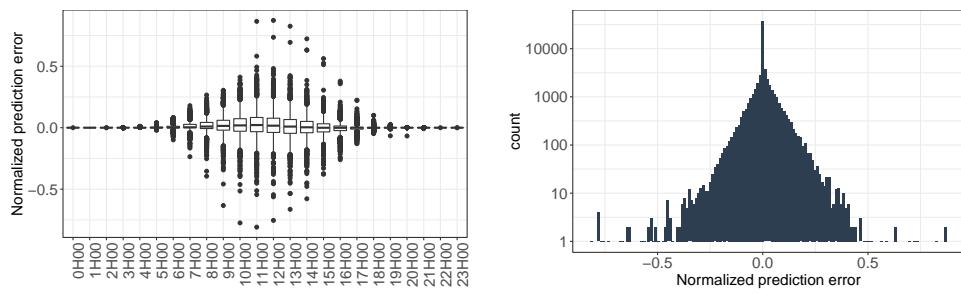


FIGURE 2.5 – Diagramme en boîte et histogramme pour les erreurs de prévisions moyennées sur chaque heure

CHAPITRE 3

Modèles et métriques

3.1 Modèles de Markov cachés

3.1.1 Chaîne de Markov

Afin d'introduire correctement les modèles de Markov cachés il est d'abord nécessaire d'introduire le concept de chaîne de Markov. Les éléments constitutifs d'une chaîne de Markov sont :

Chaîne de Markov

$\mathbf{E} = \{E_1, E_2, \dots, E_m\}$ Une liste de m états possible

Γ Matrice de probabilité de transition (de taille $m \times m$) d'un état à un autre. L'élément (i, j) , noté γ_{ij} , est la probabilité de transition de l'état i à l'état j . Comme ce sont des probabilités alors

$$\sum_{j=1}^m \gamma_{ij} = 1 \quad \forall i \in 1, 2, \dots, m$$

$\boldsymbol{\pi} = \pi_1, \pi_2, \dots, \pi_m$ où π_i représente la probabilité initiale d'être dans l'état i .

Une séquence de variables discrètes $\{C_t : t \in \mathbb{N}\}$ à valeurs dans \mathbf{E} forme

une chaîne de Markov si pour tout $t \in \mathbb{N}$ elle respecte l'hypothèse d'un processus Markovien (3.1) :

$$P(C_{t+1}|C_t, C_{t-1}, \dots, C_1) = P(C_{t+1}|C_t) \quad (3.1)$$

En d'autres termes, le processus n'a donc pas de mémoire et toute l'information utile pour prévoir l'état suivant est contenue dans l'état présent, comme le montre la figure 3.1.

D'autre part on introduit la terminologie suivante : si $P(C_{t+1}|C_t)$ est indépendant de t alors on parle de chaîne de Markov *homogène*.

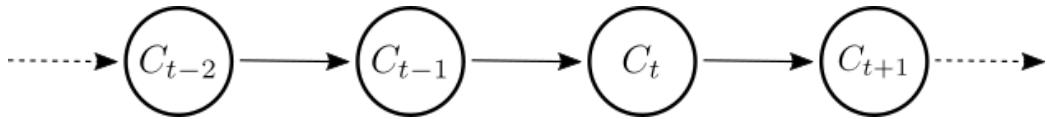


FIGURE 3.1 – Représentation graphique d'une chaîne de Markov. Figure inspirée de [1].

3.1.2 Modèle de Markov cachés

Une chaîne de Markov peut donc se révéler utile lors de la modélisation d'un phénomène observable. Cependant il arrive que le phénomène étudié ne soit que partiellement observable ou que l'on ne dispose pas de données pour le représenter. Un exemple didactique de l'utilité d'un HMM, imaginé par Jason Eisner, consistait en la détermination du caractère chaud ou froid de la journée en fonction du nombre de glaces mangées dans la journée. On a alors un processus caché à deux états (chaud ou froid) et une série d'observation (le nombre de glace mangées par jour). Un modèle de Markov caché est défini par les éléments suivants :

Modèle de Markov caché (HMM)

$\mathbf{E} = \{E_1, E_2, \dots, E_m\}$ Une liste de m états possible

Γ Matrice de probabilité de transition (de taille $m \times m$) d'un état à un autre. L'élément (i, j) , noté γ_{ij} , est la probabilité de transition de l'état i à l'état j . Comme ce sont des probabilités alors :

$$\sum_{j=1}^m \gamma_{ij} = 1 \quad \forall i \in 1, 2, \dots, m$$

$\mathbf{X} = X_1, X_2, \dots, X_T$ Une séquence représentant le phénomène observé

$p_i(x) = P(X_t = x | E_i) \quad \forall i = 1, 2, \dots, m$ La probabilité d'émission de l'observation x sachant l'état i . Dans le cas d'observations continues, ceci peut être remplacé par une densité de probabilité.

$\pi = \pi_1, \pi_2, \dots, \pi_m$ où π_i représente la probabilité initiale d'être dans l'état i .

L'idée fondamentale d'un modèle de Markov caché est de dire que l'observation ne dépend que de l'état dans lequel le système dynamique (la chaîne de Markov) se trouve. En considérant une séquence de T variables discrètes $\mathbf{C} = C_1, C_2, \dots, C_T$ à valeurs dans \mathbf{E} alors le système (3.2) est vérifié. Les hypothèses d'un modèle de Markov cachés sont représentées graphiquement dans la figure 3.2.

$$\begin{cases} P(C_{t+1}|C_t, C_{t-1}, \dots, C_1) = P(C_{t+1}|C_t) \\ P(X_t|\mathbf{X}, C_1, C_2, \dots, C_T) = P(X_t|C_t) \end{cases} \quad (3.2)$$

Lors de ce stage, ce modèle a été utilisé pour caractériser des séries temporelles à variables continues et non discrètes. Les densités de probabilités utilisées ont à chaque fois été Gaussiennes et donc entièrement caractérisées par leurs moyennes a_0 et leurs variances σ . La complexité d'un tel modèle, dans le cas d'observations continues de dimension d , est donc la suivante :

$$\underbrace{\text{Probabilités initiales}}_{m-1} + \underbrace{\text{Moyennes}}_{m \times d} + \underbrace{\frac{Kd(d-1)}{2}}_{\text{Matrices de variance covariance}} + \underbrace{m \times m}_{\text{Matrice de transition}} \quad (3.3)$$

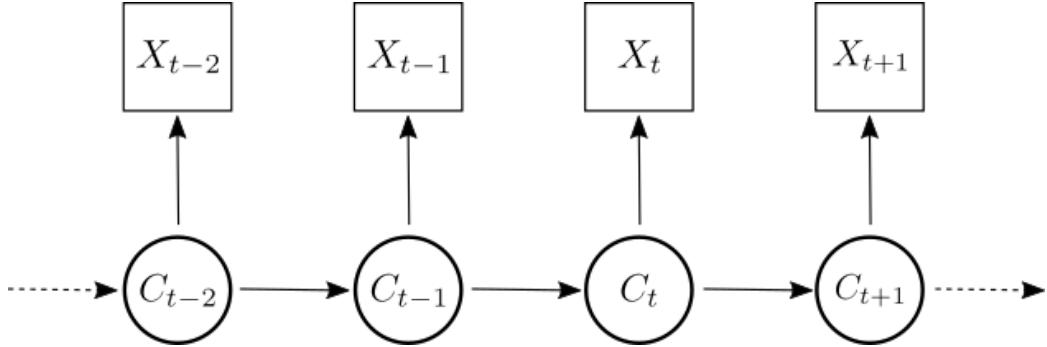


FIGURE 3.2 – Représentation graphique d'un modèle de Markov caché. Les flèches représentent les dépendances. Les carrés représentent les observations et les cercles le processus caché. Figure inspirée de [1].

Il a été montré précédemment dans [2] que les problèmes liés aux modèles de Markov cachés sont de trois natures différentes :

- Évaluation de la probabilité de l'observation d'une séquence.
- Recherche de la séquence le plus probable.
- Apprentissage

Probabilité d'observation d'une séquence

On se place dans le cadre où l'on dispose d'un HMM complet et on souhaite déterminer quelle est la probabilité d'obtenir une séquence d'observation précise. Plus formellement, en notant \mathbf{X} la séquence observée et λ le HMM considéré, on cherche $P(\mathbf{X}|\lambda)$.

En supposant que l'on connaisse la séquence d'états \mathbf{C} alors il vient le résultat suivant :

$$P(\mathbf{X}|C_1, C_2, \dots, C_T) = \prod_{i=1}^T P(X_i|C_i) \quad (3.4)$$

Et comme cette séquence d'états \mathbf{C} possède elle même une probabilité d'apparition, il vient :

$$P(\mathbf{X}, \mathbf{C}) = P(\mathbf{X}|\mathbf{C}) \times P(\mathbf{C}) = \prod_{i=1}^T P(X_i|C_i) \times \prod_{i=1}^T P(C_i|C_{i-1}) \quad (3.5)$$

Cependant, dans le cas des HMM, on ne connaît pas la succession d'états cachés qui a eu lieu pour générer cette séquence d'observation. Une méthode pour y remédier est de calculer la probabilité d'obtenir cette séquence par chaque combinaison de chemin. Traduit mathématiquement, cela donne :

$$P(\mathbf{X}, \lambda) = \sum_{\mathbf{C}} P(\mathbf{X}|\mathbf{C}) \times P(\mathbf{C}) \quad (3.6)$$

Le nombre total de séquences possibles dans le HMM λ est de m^T . Dans un cadre réaliste, où T est grand et potentiellement m aussi, ce calcul devient irréalisable. Pour y remédier, on utilise un algorithme progressif (communément appelé le forward algorithm) pour le calcul des probabilités. On introduit $\mathbf{P}(X)$ la matrice diagonale dont le ii -ième élément est $p_i(x)$ la probabilité d'obtenir l'observation X depuis l'état i . On introduit de plus le vecteur $\boldsymbol{\alpha}_t$ dont la j -ième composante $\alpha_t(j)$ est la probabilité d'être dans l'état j après t observations dans le cadre du HMM λ . Mathématiquement elle s'obtient de la façon suivante :

$$\boldsymbol{\alpha}_t = \boldsymbol{\pi} \mathbf{P}(X_1) \times \prod_{i=2}^t \boldsymbol{\Gamma} \mathbf{P}(X_i) \quad (3.7)$$

On obtient donc l'algorithme récursif suivant dont le coût calculatoire n'est plus que de $T \times m^2$:

$$\boldsymbol{\alpha}_1 = \boldsymbol{\pi} \mathbf{P}(X_1) \quad (3.8)$$

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} \boldsymbol{\Gamma} \mathbf{P}(X_t) \quad (3.9)$$

Recherche de la séquence la plus probable

On se place dans le cadre où l'on dispose d'un HMM λ et d'une série de T observations \mathbf{X} . On cherche alors à déterminer quelle est la séquence d'états cachés C_1, C_2, \dots, C_T qui a le plus de chance de générer cette séquence d'observation.

Encore une fois ce problème n'est pas viable en testant une à une les combinaisons afin de déterminer quelle est la plus probable. Pour y remédier, il est possible d'utiliser l'algorithme de Viterbi, dont la structure est assez similaire au forward algorithm. En effet il repose sur le même treillis mais la

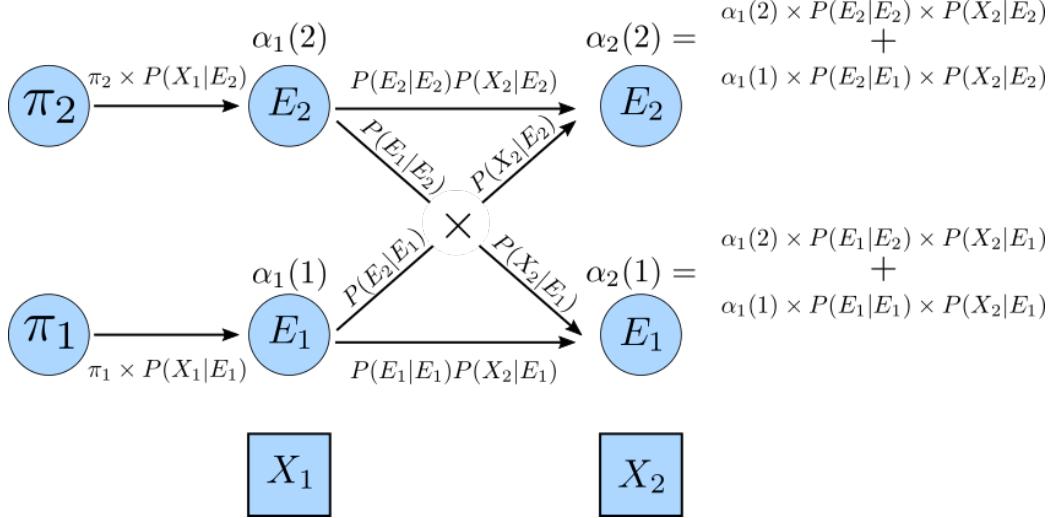


FIGURE 3.3 – Schéma du forward algorithm inspiré de [3]

valeur en chaque maille représente la probabilité d'observer X_t depuis l'état C_i en ayant emprunté le chemin le plus probable menant à C_i . A l'itération t le vecteur de probabilités de Viterbi \mathbf{v}_t se calcule donc de la manière suivante :

$$\mathbf{v}_1 = \boldsymbol{\pi} \mathbf{P}(X_1) \quad (3.10)$$

$$\mathbf{v}_t(j) = \max_{i=1}^m [\mathbf{v}_{t-1}(i) \times \gamma_{ij} \times p_j(X_t)] \quad \forall j \in 1, 2, \dots, m \quad (3.11)$$

La séquence recherchée qui maximise la probabilité de réaliser l'observation est donc celle qui mène au $\max(\mathbf{v}_T)$. Lors du calcul, il est donc nécessaire de garder en mémoire le parcours effectué.

Apprentissage

On se place dans le cadre où l'on dispose d'une liste de m états \mathbf{E} et d'une séquence de T observations \mathbf{X} . On cherche à déterminer la matrice de transition $\boldsymbol{\Gamma}$ et les probabilités d'émissions $p_i(x) \quad \forall i \in 1, 2, \dots, m \quad \forall x \in \mathbf{X}$.

Pour ce faire il est possible de faire appel à un algorithme itératif nommé "forward backward". Celui-ci nécessite l'introduction d'une autre entité β_t qui est défini par l'équation (3.12).

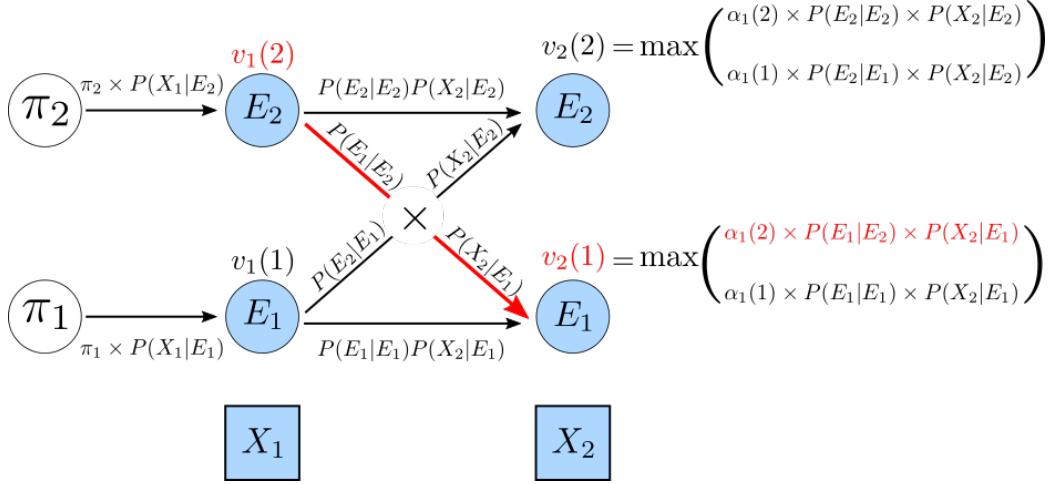


FIGURE 3.4 – Représentation du parcours du treillis par l’algorithme de Viterbi. Les cercles à fond blanc représentent les probabilités initiales ; les cercles à fond bleu les états et les carrés les observations. Les v sont les probabilités de Viterbi. Un maximum arbitraire à chaque t est représenté en rouge pour illustrer le cheminement de la séquence la plus probable. Schéma inspiré de [3].

$$\beta_t(i) = P(X_{t+1}, X_{t+2}, \dots, X_T | C_t = E_i) \quad (3.12)$$

C’est un algorithme itératif qui calcule une estimation initiale des probabilités et qui utilise cette même estimation pour générer de meilleures probabilités. La première étape est de définir une formule pour calculer une estimation de γ_{ij} que l’on notera $\hat{\gamma}_{ij}$:

$$\hat{\gamma}_{ij} = \frac{\text{Nombre attendu de transitions entre } E_i \text{ et } E_j}{\text{Nombre total attendu de transitions depuis } E_i} \quad (3.13)$$

En introduisant $\xi_t(i, j) = P(C_{t+1} = j, C_t = i | \mathbf{X}, \lambda)$ On remarque que le numérateur pourrait alors s’écrire comme étant :

$$\sum_{t=1}^{T-1} \xi_t(i, j) \quad (3.14)$$

D'autre part, on remarque qu'il est possible de calculer une quantité similaire à $\xi_t(i, j)$:

$$\hat{\xi}_t(i, j) = P(C_{t+1} = j, C_t = i, \mathbf{X} | \lambda) = \alpha_t(i) \times \gamma_{ij} p_j(X_{t+1}) \times \beta_{t+1}(j) \quad (3.15)$$

D'après l'identité $P(AB|C) = P(A|BC)P(B|C)$ on a :

$$\xi_t(i, j) = \frac{\hat{\xi}_t(i, j)}{P(\mathbf{X} | \lambda)} = \frac{\alpha_t(i) \times \gamma_{ij} p_j(X_{t+1}) \times \beta_{t+1}(j)}{\alpha_T(X_T)} \quad (3.16)$$

Au final, on obtient :

$$\hat{\gamma}_{ij} = \frac{\sum_{t=1}^{T-1} \frac{\alpha_t(i) \times \gamma_{ij} p_j(X_{t+1}) \times \beta_{t+1}(j)}{\alpha_T(X_T)}}{\sum_{k=1}^m \sum_{t=1}^{T-1} \frac{\alpha_t(i) \times \gamma_{ik} p_k(X_{t+1}) \times \beta_{t+1}(k)}{\alpha_T(X_T)}} \quad (3.17)$$

Il reste encore à déterminer les probabilités d'émissions. Pour cela on dérive une formule pour l'estimation de $p_i(x)$:

$$\hat{p}_i(X) = \frac{\text{Nombre attendu d'observations de } X \text{ depuis } i}{\text{Nombre attendu d'occurrences de } i} \quad (3.18)$$

On introduit :

$$\kappa_t(j) = P(C_t = E_j | \mathbf{X}, \lambda) = \frac{P(C_t = E_j, \mathbf{X} | \lambda)}{P(\mathbf{X} | \lambda)} \quad (3.19)$$

On est dans un cas particulier de la situation précédente avec $i = j$ et on obtient donc le résultat suivant :

$$\kappa_t(j) = \frac{\alpha_t(j) \beta_t(j)}{\alpha_T(X_T)} \quad (3.20)$$

Au final :

$$\hat{p}_i(X) = \frac{\sum_{t=1}^T \delta \kappa_t(i)}{\sum_{t=1}^T \kappa_t(i)} \quad \text{avec } \delta = \begin{cases} 1 & \text{si } X_t = X \\ 0 & \text{sinon} \end{cases} \quad (3.21)$$

Il est maintenant possible de raffiner l'estimation précédente de Γ et des $p_i(x)$ à chaque itération. Cependant cet algorithme converge vers un minimum local. Il peut donc être important de déterminer une "bonne" estimation initiale.

3.1.3 Modèles Markov Switching Auto-Regressifs

D'après le schéma 3.2, l'utilisation d'un HMM pour représenter l'évolution d'une variable météorologique suppose que toute la dynamique est représentée par l'état caché. Or, notamment dans le contexte météorologique, il est clair que la probabilité d'une observation dépend de l'observation précédente. Pour affaiblir cette hypothèse il est possible d'utiliser des modèles de type MSAR où la dynamique de la série temporelle est décrite par une collection de modèles auto-régressifs. Le choix du modèle auto-régressif est alors contrôlé par une chaîne de Markov cachée. Le schéma 3.5 représente graphiquement les dépendances d'un MSAR.

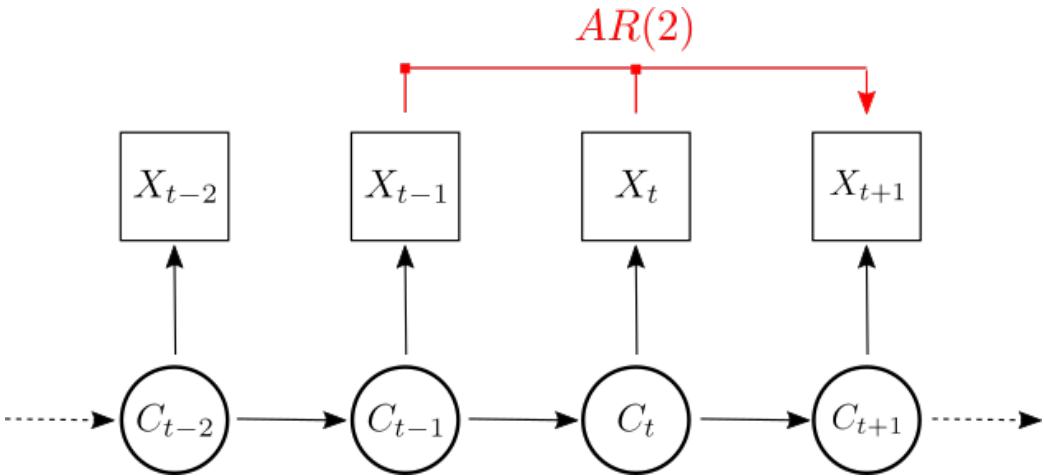


FIGURE 3.5 – Représentation des dépendances d'un modèle de type MSAR avec des modèles auto-régressifs d'ordre 2. Ici l'observation X_{t+1} est générée à partir du modèle pointé par l'état caché C_{t+1} qui prend en entrée les observations X_t et X_{t-1} .

D'autre part, le fonctionnement des HMM a été présenté dans la section précédente pour des phénomènes à valeurs discrètes. Or les séries temporelles

étudiées dans ce rapport sont continues. C'est pourquoi il convient de remplacer les probabilités d'émettre une observation spécifique $p_i(x)$ par les éléments constitutifs d'une densité de probabilité.

3.2 Modèles GARCH

Les modèles *Auto-Regressive Conditionnal Heteroskedasticity* (ARCH) ont été introduit pour prendre en compte le caractère variable de la variance d'une série temporelle. ils se basent sur l'étude des résidus d'un processus modélisant la moyenne (ARIMA par exemple), notés ϵ_t . Ces résidus ne doivent plus être auto-corrélatés. Le modèle sépare alors ces résidus en deux parties distinctes avec σ_t un écart type qui dépend du temps et z_t une partie stochastique (un bruit blanc).

$$\epsilon_t = \sigma_t z_t \quad (3.22)$$

Le modèle ARCH modélise l'écart type de la façon suivante :

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \dots + \alpha_p \epsilon_{t-p}^2 \quad (3.23)$$

Avec comme conditions que $\alpha_0 > 0$ et $\alpha_i \geq 0 \quad \forall i \in [1, \dots, n]$. Une structure plus générale, le modèle *Generalized ARCH* (GARCH), propose que l'écart type soit modélisé de la façon suivante :

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \dots + \alpha_p \epsilon_{t-p}^2 + \beta_{t-1} \sigma_{t-1}^2 + \dots + \beta_{t-q} \sigma_{t-q}^2$$

3.3 K-moyennes (Kmeans)

Le modèle des K moyennes est un est un modèle de partitionnement de données (ou souvent dit de "*clustering*" d'après la dénomination anglaise). Le problème consiste donc à diviser un ensemble de points en un certains nombre de groupe m de façon à ce qu'ils minimisent une certaine fonction.

Dans le cas présent, j'ai fait appel à la routine R `Kmeans`, qui cherche à minimiser la somme des erreurs quadratiques au sein de chaque groupe grâce à l'algorithme heuristique décrit dans [4]. Il fonctionne de la façon suivante : soit une partition de l'ensemble des points C . Alors, un point aléatoire x est choisi et le centre du cluster auquel il appartenait est mis à jour comme si ce point ne lui appartenait plus. Le point choisi aléatoirement est ensuite

réassigné au cluster qui permet de minimiser le plus possible la fonction de coût total, c'est à dire la somme des erreurs quadratiques au sein des clusters. Par construction, cet algorithme permet de ne jamais accroître la fonction de coût. Ainsi, il donne assurance qu'une partition dont les points sont fixés sera atteinte après un nombre fini d'itérations.

3.4 Modèle de mélange gaussien (GMM)

Le modèle de mélange de mélange gaussien est un modèle statistique qui tente d'expliquer une distribution statistique comme étant une somme de sous distributions gaussiennes. La figure ?? présente un cas d'application idéaliste d'un tel modèle.

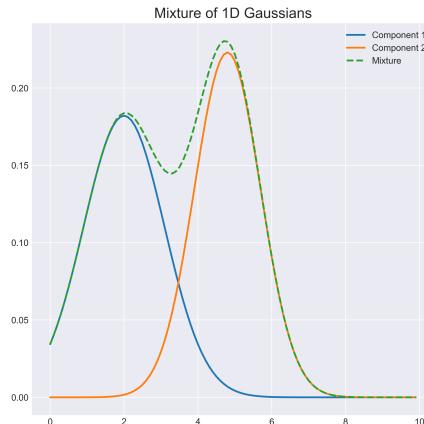


FIGURE 3.6 – Exemple d'application d'un GMM dans un cas idéaliste

Ce modèle cherche donc à identifier, pour chaque gaussienne, sa variance, sa moyenne et les proportions. La gestion des problèmes en haute dimension avec ce modèle peut s'avérer difficile car il est sujet à la malédiction de la dimension. En effet, si d est la dimension de l'espace et m le nombre de clusters, alors le nombre de paramètres à identifier est :

$$\underbrace{m - 1}_{\text{Proportions}} + \underbrace{m \times d}_{\text{Centres}} + \underbrace{\frac{md(d - 1)}{2}}_{\text{Matrices de variance covariance}} \quad (3.24)$$

On a donc le nombre de paramètres qui évolue de façon linéaire avec le nombre de clusters m et quadratique avec la dimension d . Comme le montre [5], il est préférable d'utiliser des modèles réduits ou par de la sélection de variable plutôt que de réduire l'espace dans lequel vivent les données. C'est pourquoi, au cours de ce stage, les paramètres des modèles de mélanges gaussiens ont été estimés via l'utilisation du package `Mclust`. En effet, il dispose d'une liste de modèles de mélange gaussien qui facilite la prise en compte de la haute dimension en réduisant le nombre de paramètres, comme explique dans [?]. Il utilise, tout comme les modèles MS-AR, un algorithme EM pour identifier les jeux de paramètres maximisant la vraisemblance.

3.5 Analyse par Composante Principale

3.6 Métriques

3.6.1 BIC

Le *Bayesian Information Criterion* permet de faire un compromis entre la vraisemblance des modèles vis-a-vis d'une série temporelle et leurs nombre de paramètres. Il est défini dans (3.25) avec L la vraisemblance, k le nombre de paramètres du modèle et N le nombre d'observations dans la série temporelle. Au vu de sa définition c'est un score orienté négativement. Le meilleur modèle au sens du critère BIC est donc celui qui obtiendra les valeurs les plus faibles.

$$BIC = -2 \log L + k \log N \quad (3.25)$$

3.6.2 Erreur moyenne absolue (MAE)

L'erreur moyenne absolue est définie comme étant :

$$MAE = \frac{\sum_{i=1}^T |\tilde{P}_i - P_i|}{T} \quad (3.26)$$

3.6.3 Erreur quadratique moyenne (RMSE)

L'erreur quadratique moyenne est définie comme étant :

$$RMSE = \sqrt{\frac{\sum_{i=1}^T (\tilde{P}_i - P_i)^2}{T}} \quad (3.27)$$

CHAPITRE 4

Résultats pour la production éolienne

L’application des modèles MS-AR aux erreurs de prévisions de production d’énergie éolienne a donné lieu à la rédaction d’une première version d’un article scientifique. Le lecteur est donc invité à lire ce document en annexe A. Les travaux que j’ai effectués sont présentés dans la section intitulé *Modelling of the wind power forecast error*. Les poursuites d’études, notamment celles concernant les transitions non homogènes, ont été commencées mais elles ne sont pas assez abouties pour être intégrées telles quelles dans ce rapport.

CHAPITRE 5

Application aux erreurs de prévisions photovoltaïques

Les erreurs de prédiction de production photovoltaïques diffèrent par nature de celles de production éolienne. Alors que le vent peut être d'intensité non négligeable à tout instant, l'irradiance sera quant à elle complètement nulle la nuit. Comme l'alternance jour/nuit est un phénomène purement déterministe les prévisionnistes ne commettent pas d'erreurs de prévisions la nuit. Ceci modifie la structure des erreurs de prévisions en introduisant régulièrement des plages de prévisions nulles. Or les outils utilisés dans la section 4 se heurtent à des problèmes de convergence numérique s'ils sont utilisés tels quels sur la série d'erreur de prédiction de production photovoltaïque.

Afin de se ramener à une structure de série temporelle moins délicate en termes de résolution numérique, une approche simpliste aurait consisté à supprimer les périodes nocturnes de la série temporelle. Cependant, cette méthode fait part de deux désavantages qui à fait qu'elle a été écartée au cours de ce stage. Tout d'abord, même si elle paraît très simple, il faut malgré tout tenir compte que la durée de la nuit change tous les jours. De surcroît, cela imposerait un pré-traitement des données qui dépend aussi de leur origine géographique. Ainsi la méthodologie et le modèle obtenu paraît être trop difficilement transposable pour être un jour utilisé en opérationnel. C'est

Expliquer
pourquoi
clairement

pourquoi une approche légèrement différente a été choisie.

En observant la série de donnée, il est possible de remarquer que des trajectoires quotidiennes typiques apparaissent. Cette approche à déjà été utilisée dans [8], ce qui permet de se soustraire aux problèmes de convergences numériques en introduisant un autre difficulté : la haute dimension.

5.1 Méthode des plus proches voisins

Dans un premier temps, je me suis attaché à voir s'il était possible de corriger la prédiction fournie un jour à l'avance en fonction de l'allure des erreurs déjà observées dans la journée courante.

On note \hat{d} l'indice de dimension indiquant quand la correction doit être réalisée. Alors nécessairement $\hat{d} < d$. Pour le contexte applicatif, afin de se placer dans un cadre qui pourrait être réaliste au sens opérationnel, la correction doit être effectuer à 10h.

L'identification des journées les plus proches a été réalisé en sélectionnant celles qui minimisaient (5.1) où z représente la trajectoire de la journée courante et \tilde{z} une des trajectoires passées.

$$\sum_{i=1}^{\hat{d}} |z_i - \tilde{z}_i| \quad (5.1)$$

La première étape dans l'utilisation de la méthode des plus proches voisins consiste à identifier quelle est le nombre optimal de plus proches voisins à prendre en compte. Dans cet objectif, il est nécessaire d'identifier une métrique pour évaluer le gain en terme de prédiction. J'ai décidé d'utiliser comme critère (5.2) où \bar{z}_i^{NN} représente la moyenne des plus proches voisins considérés.

$$\frac{\sum_{i=\hat{d}}^d |z_i - \bar{z}_i^{NN}|}{\sum_{i=\hat{d}}^d |z_i|} \quad (5.2)$$

La Figure 5.1 montre des exemples où cette méthode de correction d'erreur fonctionne plus ou moins bien. Il est quand même intéressant de constater que même dans le cas où elle amplifierait l'erreur commise, plusieurs trajectoires parmi les voisins identifiés avaient capturées le bon comportement.

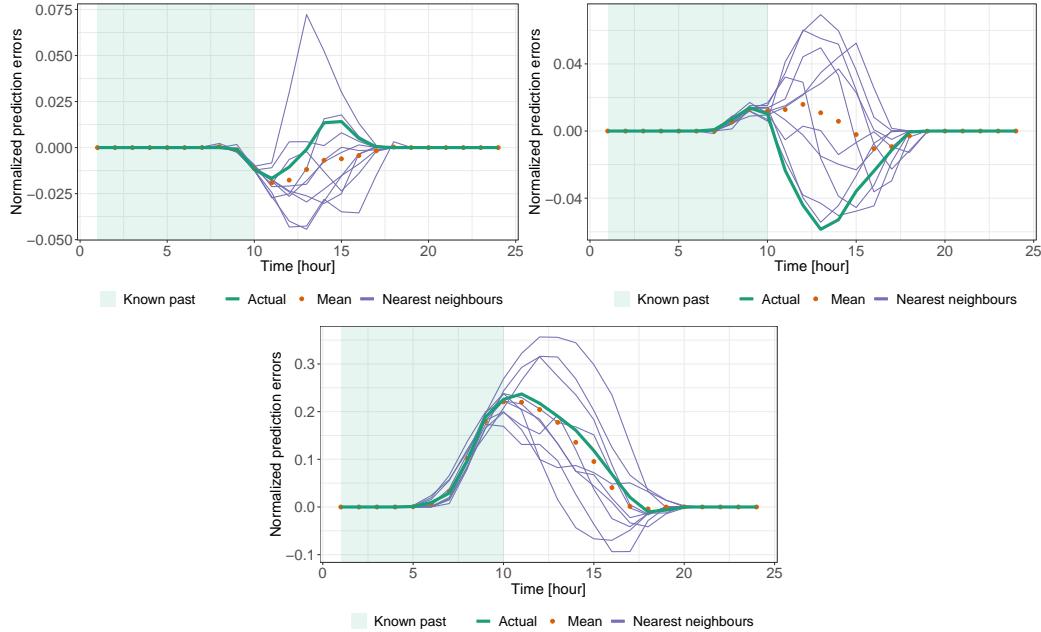


FIGURE 5.1 – Exemples de trajectoires d’erreurs de prévisions ainsi que les 10 trajectoires les plus proches identifiés via (5.1) et leur moyenne

La comparaison des performances en fonction du nombre de plus proches voisins est présentée dans la Figure 5.2, avec le label *all*. On remarque que, dans la limite des cas testés, l’augmentation du nombre de plus proches voisins ne dégrade pas les résultats obtenus. Cependant, en se limitant aux 5 premiers voisins on obtient une réduction de l’erreur sur le reste de la journée de l’ordre de 30%.

Cependant, comme cette méthode requiert d’identifier la trajectoire courante parmi toutes les trajectoires connues, elle peut s’avérer très coûteuse en temps de calcul au fil du temps. Par ailleurs, son comportement va par définition évoluer au cours du temps, ce qui peut amener à des déconvenues. Une option pour limiter cet effet négatif réside dans la détermination de trajectoires types. En effet, en regardant la figure 5.1, on peut observer des trajectoires d’erreurs typique en forme de paraboles positive ou négative centrée autour de 14h. D’où l’idée d’identifier les plus proches voisins non plus parmi l’ensemble des trajectoires disponibles mais depuis des trajectoires représentatives.

L’identification de ces trajectoires a été réalisée grâce à la méthode des Kmeans (voir 3.3). La figure 5.3 présente les trajectoires types obtenues pour

différents nombres de classes. Au vu du contexte, le nombre optimal de classe a été choisi en comparant les résultats obtenus pour de la correction d'erreur. Ces résultats sont présentés dans la figure 5.2. Tout d'abord, l'identification des plus proches voisins parmi ces groupes détériore systématiquement les résultats par rapport à l'identification sur tout le passé connu. Cependant en identifiant la moyenne d'environ 5 trajectoires types parmi 20, une correction d'erreur de l'ordre de 25% peut être atteinte.

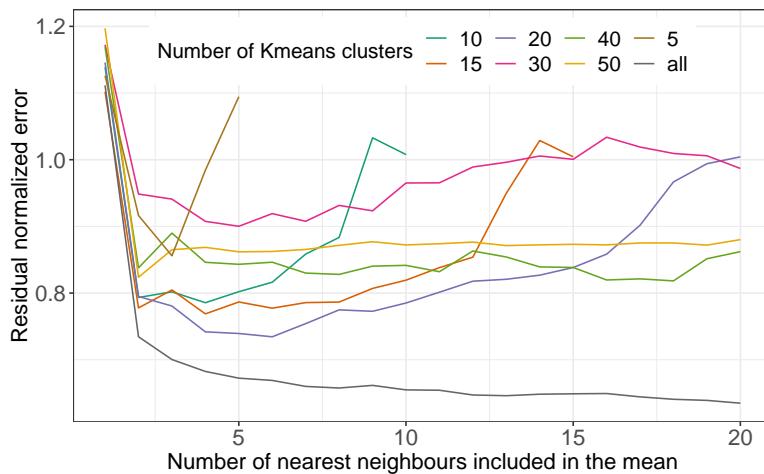


FIGURE 5.2 – Comparaison des corrections de prévisions en fonction du nombre de plus proches voisins. La série *all* présente les résultats quand les plus proches voisins sont identifiés parmi toutes les trajectoires du jeu de données d'entraînement. Les autres, quand ils sont identifiés parmi un certain de clusters obtenus par la méthode des Kmeans (voir section 3.3)

Dans cette section, une approche assez simple de la méthode des plus proches voisins a été mise en œuvre et elle a permis d'en démontrer le potentiel. En effet elle a permis d'atteindre des corrections des prévisions au cours de la journée qui permettent de réduire les erreurs de prévisions entre 25 et 35%. De meilleurs résultats auraient probablement pu être obtenus en affinant la méthode. En effet, il aurait pu être intéressant de pondérer certaines heures dans l'identification des plus proches voisins afin de donner plus de poids aux dernières heures connues. Ou encore de regarder si la prise en compte du jour précédent pouvait amener de l'information supplémentaire ? Finalement, combiner l'utilisation de l'erreur absolue avec l'erreur quadratique aurait aussi pu amener plus d'informations. Cependant, le potentiel de cette méthode

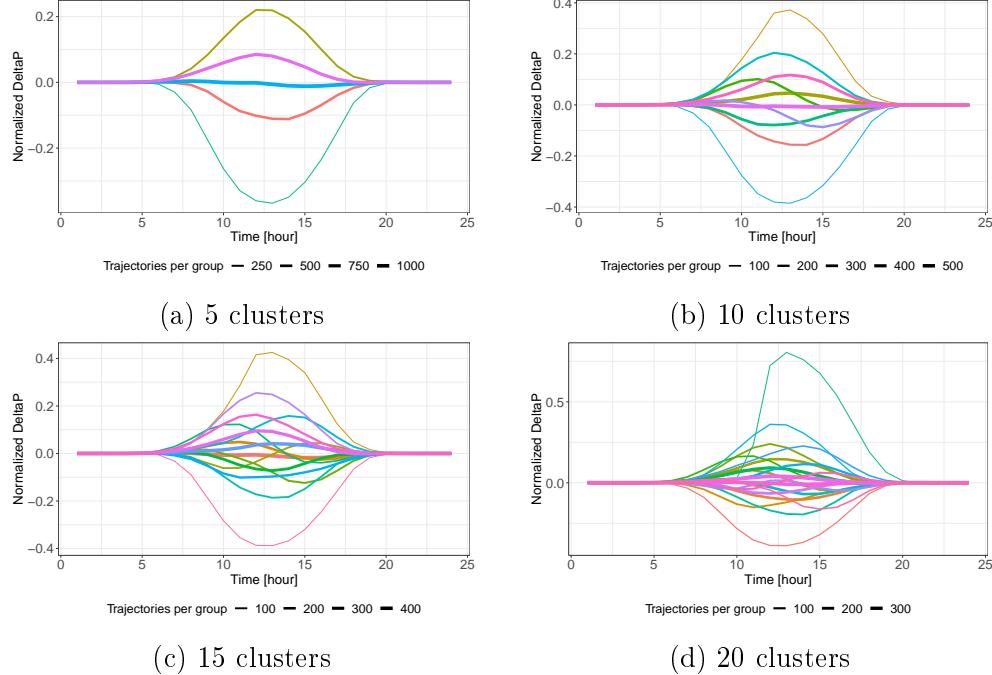


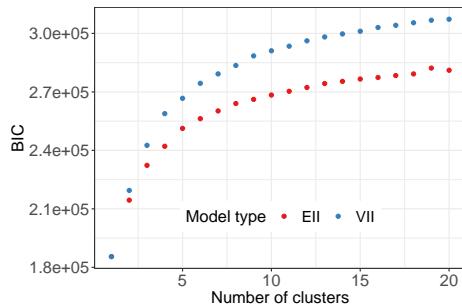
FIGURE 5.3 – Centroïdes des clusters obtenus par la méthode des Kmeans sur la série d’entraînement. la population des clusters est matérialisée par l’épaisseur du trait.

m'a conduit à explorer une autre piste. En effet, plutôt que de n'identifier que les trajectoires types via la méthode des Kmeans, de l'information supplémentaire sur les variances associées pouvaient se révéler utile en utilisant la méthode de mélange Gaussien.

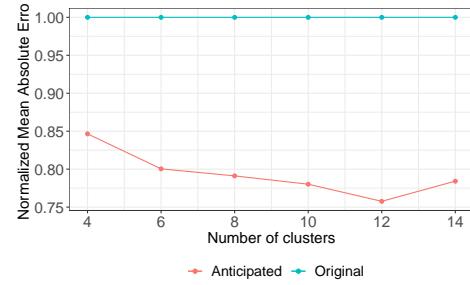
5.2 Résultats avec un GMM

Le modèle de mélange Gaussien, présenté dans 3.4, a été ajusté sur la série temporelle des erreurs de prédiction photovoltaïques. Comme pour les Kmeans, il est nécessaire d'identifier le nombre optimal de clusters K . Pour ce faire, le critère BIC est à nouveau utilisé comme premier indicateur. On peut constater sur la figure 5.4a que le modèle de type VII (matrice de variance covariance diagonale et à valeur unique par cluster) semble surpasser le modèle de type EII (matrice de variance covariance diagonale et à valeur

unique commune pour tous les cluster). Cette figure ne présente pas les résultats obtenus par les autres modèles car ils n'ont pas convergé, probablement à cause d'un manque de données. En effet, même si le jeu d'entraînement se compose de 6 années complètes, cela ne représente qu'un total de 2190 journées. Ce nombre est assez vite relativement faible face aux nombres de paramètres à estimer des GMM plus évolués (voir 3.4). Par exemple, si chaque valeur sur la diagonale est distincte alors on passe de $(K - 1) + K \times d + K$ à $(K - 1) + K \times d + K \times d$ paramètres et il est donc nécessaire d'avoir nettement plus de données pour les estimer. Dans ce contexte précis, il aurait pu être intéressant de pouvoir décrire la diagonale de la matrice de variance covariance par deux valeurs. Cela aurait pu permettre à l'algorithme de dissocier les variances pour les heures diurnes et nocturnes tout en restreignant le nombre de paramètres. Mais cette option n'est pas offerte par le package **Mclust** et n'a pas fait l'objet d'une implémentation de ma part par manque de temps.



(a) Valeurs du critère BIC pour différents type de mélange gaussiens et nombres de groupes



(b) Corrections d'erreurs à 10 heures obtenus avec plusieurs GMM dont le nombre de clusters K diffèrent

Le critère BIC ne favorise pas un nombre spécifique de cluster mais donne plutôt une tendance générale favorisant systématiquement l'augmentation du nombre de clusters. Une explication plausible réside dans le fait que le phénomène sous-jacent, l'évolution du temps au sens météorologique du terme, est par définition un phénomène continu et non discret. Plus on discrétiserait son espace de représentation, plus on se rapprocherait d'une représentation continue et ainsi meilleurs sont les résultats. Afin de garder l'interprétabilité du modèle obtenu, et d'en limiter sa complexité, il est impératif de choisir un nombre de cluster relativement restreint. La figure 5.4a montre une croissance plus faible en terme de critère BIC à partir du sixième cluster.

Afin de combiner des informations provenant de deux indicateurs différents, il est possible de voir comment le GMM répond à une application spécifique en fonction de K . J'ai donc choisi de les comparer sur la base de leur potentiel à réaliser de la correction au cours de la journée. Des exemples de ces corrections sont présentées dans la figure 5.5. Ces résultats sont présentés sur la figure 5.4b. Un net gain, d'environ 5%, est constaté en passant de 4 clusters à 6. Ensuite les gains sont relativement plus faible et pour réobtenir un gain similaire il faudrait prendre 12 clusters. En comparant ces deux indicateurs, il semble que le modèle à 6 clusters soit un bon compromis entre complexité et précision.

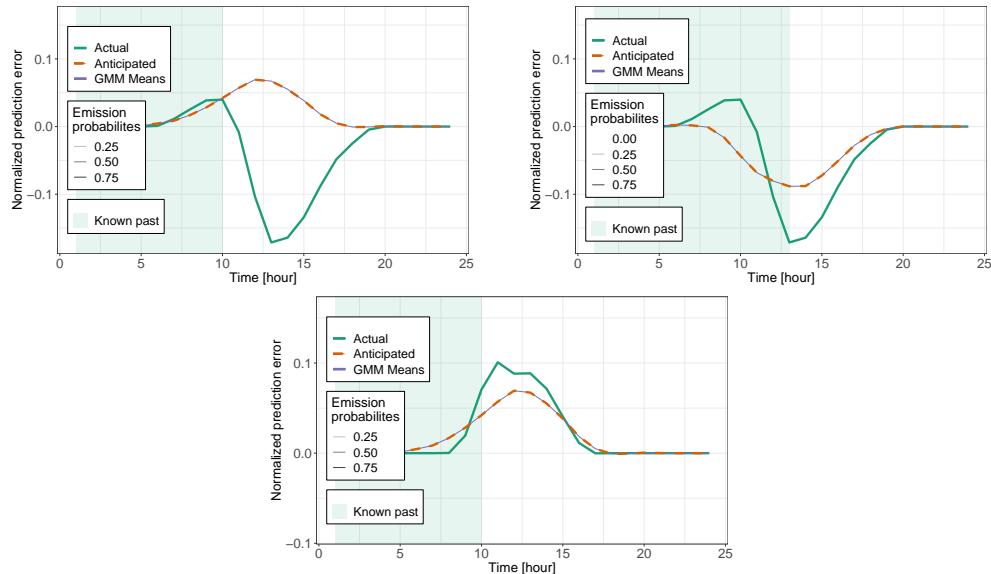


FIGURE 5.5 – Corrections au cours de la journée avec le GMM de la figure 5.6.

Les paramètres du modèle obtenus sont présentés graphiquement dans la figure 5.6. Il est possible de constater que les trajectoires moyennes des états sont assez semblables à celles obtenus par la méthode des Kmeans (voir figure 5.3). Enfin, l'information supplémentaire réside dans la connaissance des variances pour chaque clusters. Elle sont relativement faible et globalement assez similaires les unes des autres pour les groupes allant de 1 à 5. Par contre, le groupe 6, a une variance nettement plus forte. Ce cluster, qui est significativement moins peuplé que les autres, regroupe les individus au

comportement inhabituels et/ou chaotiques.

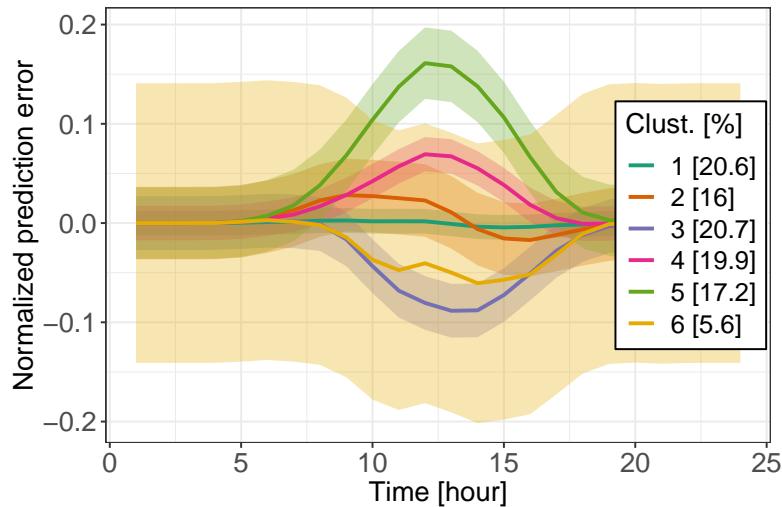


FIGURE 5.6 – Représentation graphique des paramètres du modèle GMM à 6 clusters obtenu sur la série des erreurs prévisions photovoltaïques. Les lignes continues représentent les moyennes, les fonds colorés leur intervalles de confiance de 68% respectifs. Les proportions respectives sont intégrées dans la légende.

Finalement, l'utilisation de modèle GMM n'a pas permis d'améliorer les résultats obtenus en terme de corrections de prévisions au cours de la journée par rapport aux plus proches voisins. En effet, le modèle GMM retenu parvient à réduire de 20% les erreurs pour une correction à 10 heures alors que les plus proches voisins permettaient une réduction de l'ordre de 35%. La variante en utilisant une identification de clusters type déterminés par la méthode des Kmeans conduisait à une réduction de l'ordre de 25% dans les mêmes conditions. Cependant, le modèle GMM apporte quand même de l'information supplémentaire non négligeable avec les précisions sur les variances de chaque clusters. Ceci permet par exemple d'identifier les probabilités d'appartenance au cours de la journée via les densités de probabilités, ce qui est plus fiable que d'utiliser l'erreur absolue ou l'erreur quadratique. Par ailleurs, une fois les probabilités d'appartenance, cela permet de générer des scénarios de comportements sur le reste de la journée et ainsi de réaliser une prédiction stochastique sur le reste de la journée par opposition aux Kmeans qui proposaient une prévision déterministe. Mais, mis à part donner un

comportement moyen, ce modèle ne permet toujours pas d'anticiper l'avenir au delà de la journée en cours. C'est pourquoi il pourrait être intéressant d'identifier un HMM dont la structure est similaire si ce n'est que la matrice de transition permette de prévoir des scénarios sur plusieurs journées.

5.3 Résultats avec un HMM

En gardant à l'esprit que ce modèle est identifié dans le but de générer des prévisions sur plusieurs journée, il est impératif de donner plus de liberté au HMM qu'au GMM quant à la structure de sa matrice de variance covariance. En effet, si on ne relaxe pas la contrainte d'avoir une valeur unique sur la diagonale de variance covariance alors le modèle ne sera jamais en mesure de capturer les erreurs nulles ou négligeables la nuit. Les scénarios générés n'auraient pas alors cette caractéristique et les performances en termes d'applications seraient grandement réduites. Cependant, ce faisant, le nombre de paramètres passe de chaque modèle devient alors :

$$\underbrace{K - 1}_{\text{Probabilités initiales}} + \underbrace{K \times d}_{\text{Centres}} + \underbrace{K \times d}_{\text{Matrices de variance covariance}} + \underbrace{K \times K}_{\text{Matrice de transition}} \quad (5.3)$$

Par exemple, en prenant le cas d'une observation toutes les heures ($d = 24$) et de 6 états cachés ($K = 6$) par identification avec le GMM, il faut donc identifier 329 paramètres. C'est donc plus du double des 155 paramètres du GMM sélectionné dans la section 5.2. Pour pallier à ce problème, j'ai procédé à une réduction de dimension de l'espace par la méthode de l'ACP (voir 3.5). Les résultats obtenus en fonction du nombre de composantes sélectionnées sont présentés dans la figure 5.7. J'ai choisi de garder un nombre de composante principales expliquant 99% de la variance totale, soit 6 composantes principales.

Ces 6 composantes sont présentées sur la figure 5.8. Les 6 composantes principales partagent le fait d'être nulles ou négligeables avant 5 heures et après 20 heures. Ces plages horaires sont celles pour lesquelles les erreurs de prédiction sont nulles toute l'année comme le montre la figure ???. La première composante, si multipliée par le facteur adéquat, peut permettre de représenter la plupart des trajectoires types obtenues par les différents modèles des sections précédentes. On comprend donc pourquoi elle permet, à elle seule, d'expliquer 80% de la variance. Physiquement, elle traduit une

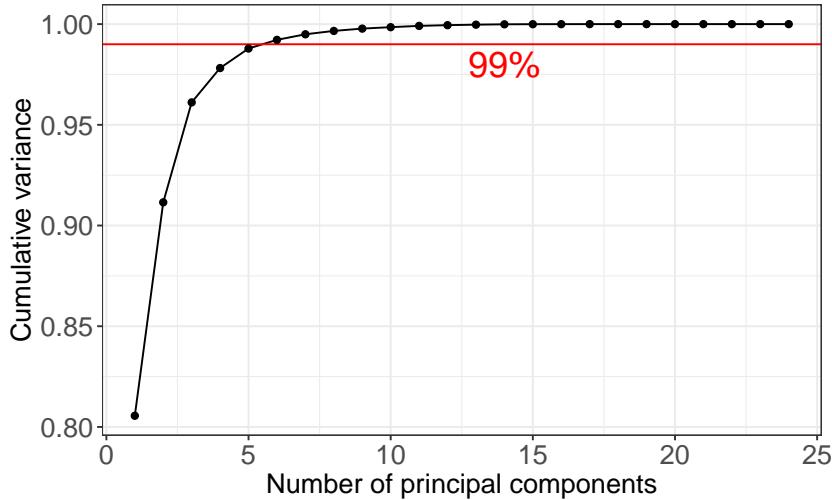


FIGURE 5.7 – Variance cumulée expliquée en fonction du nombre de composante principales

pure sur ou sous estimation de la puissance productible au jour considéré. La deuxième composante principale cherche à représenter les journées avec surestimation le matin et sous estimation l'après-midi et inversement. Certaines trajectoires types obtenues précédemment traduisaient la relativement forte présence de journées d'erreurs de prévision avec ce profil. Les composantes principales d'ordres supérieurs ajoutent à chaque fois une alternance entre surestimation et sous-estimation. Procéder à une ACP dans le cas présent revient donc à faire décomposition en série de Fourier sur la plage d'heures ensoleillées tout en laissant toujours nulles les heures qui ne le sont jamais.

Dans cet espace réduit, le nombre de paramètres d'un HMM à états cachés ne serait donc plus que de 113 contre les 329 dans l'espace de départ. L'initialisation des paramètres pour le HMM a été faite à partir de ceux obtenus pour le GMM. Les paramètres du modèle obtenu sont présentés graphiquement dans la figure 5.9a.

L'interprétation des régimes sous cette forme n'est pas aisée mais deux éléments sont intéressant à noter. Tout d'abord, les régimes sont caractérisés par des valeurs non négligeables sur la première composante uniquement. D'autre part, la variance diminue en augmentant l'ordre de la composante principale, et ce dans tous les régimes. Ce phénomène est inhérent à l'utilisation de l'ACP qui détermine les composantes principales en s'appuyant sur une

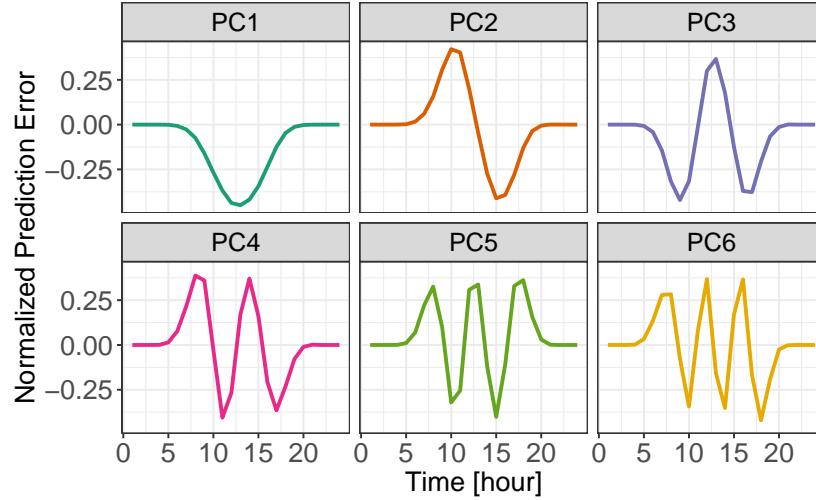


FIGURE 5.8 – Représentation graphique des 6 composantes principales obtenues par la l'ACP dans l'espace d'origine de dimension 24

maximisation de la variance qu'elles expliquent. Les régimes, transformés dans l'espace d'origine, sont présentés dans la figure 5.9b. On retrouve les profils obtenus dans par les modèles utilisés précédemment. Notamment deux modèles de surestimation plus ou moins forte, deux modèles de sous-estimation plus ou moins forte et deux modèles à moyenne nulles mais avec des variances très distinctes. Comme dans le cas du GMM, le régime 6 sert à caractériser les journées aux profils atypiques.

Utiliser ce modèle pour faire de la correction au cours de la journée n'apporterait que peu de gain par rapport au GMM de la section 5.2 au vu de leurs nombreuses similitudes. Cependant, il est possible de l'utiliser pour faire des prévisions d'ensembles sur une ou plusieurs journées. Comme pré-requis, il est nécessaire de pouvoir déterminer les probabilités d'appartenance d'une journée à chacun des états cachés. Elles sont déterminées par l'utilisation de l'algorithme forward backward décrit dans 3.1.2. Des exemples de probabilités d'appartenance sont présentées dans la figure 5.10 dans l'espace réduit (figure 5.10a) et dans l'espace d'origine (figure 5.10b).

On constate que, dans cet exemple, les probabilités d'appartenance mettent en évidence une compétition entre plusieurs états. En effet, comme aucune des trajectoires ne correspond exactement à une trajectoire type, c'est l'information sur les variances de chaque états qui est prépondérante. Par exemple, le profil

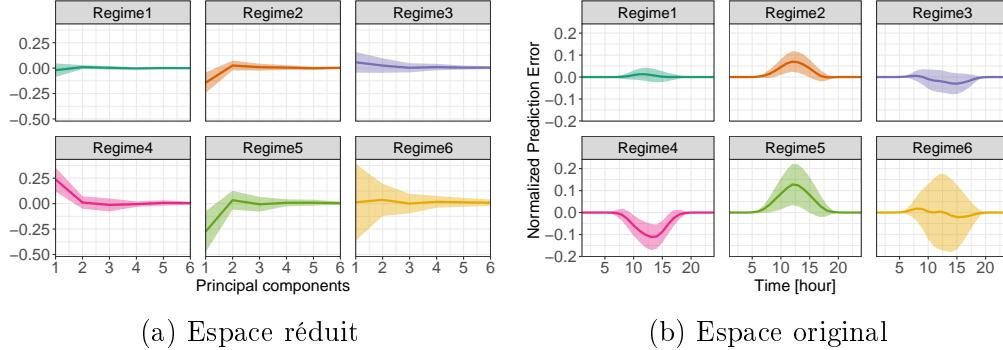


FIGURE 5.9 – Représentation des moyennes et variances d'un HMM à 6 états cachés

moyen de la première journée semble, dans l'espace de dimension réduit, être le plus proche du Régime 1. Cependant, comme il a une très faible variance notamment sur les composantes principales d'ordre supérieur à 1, il n'a pas une plus grande probabilité d'avoir émis cette observation que les régimes 2 et 3. La troisième journée se rapproche très fortement de la trajectoire moyenne du régime 4 et il est donc nettement plus probable que ce soit lui qui ait la probabilité d'émission la plus forte. La sixième journée représente un autre cas intéressant. En effet, la première composante principale est assez proche de celle de l'état 5. Mais, au vu de la forte valeur sur la troisième composante principale, une compétition entre le régime 5 (forte erreurs positive) et le régime 6 (caractérisé par sa forte variance) s'installe.

A partir de ces probabilités d'appartenance, il est possible de faire de la prévision d'ensemble, c'est à dire de la génération de scénarios probables. En effet, en les utilisant pour définir une loi de probabilité multinomiale, il est possible de déterminer pour chaque scénario le régime ayant généré la trajectoire initiale. Le régime suivant est alors déterminé par tirage d'une autre loi binomiale définie cette fois par une ligne de la matrice de transition Γ . L'observation générée sera donc la moyenne du régime tiré auquel est ajouté un bruit défini par la matrice de variance covariance dudit régime. L'opération est à répéter jusqu'à l'obtention de scénarios de la durée voulue.

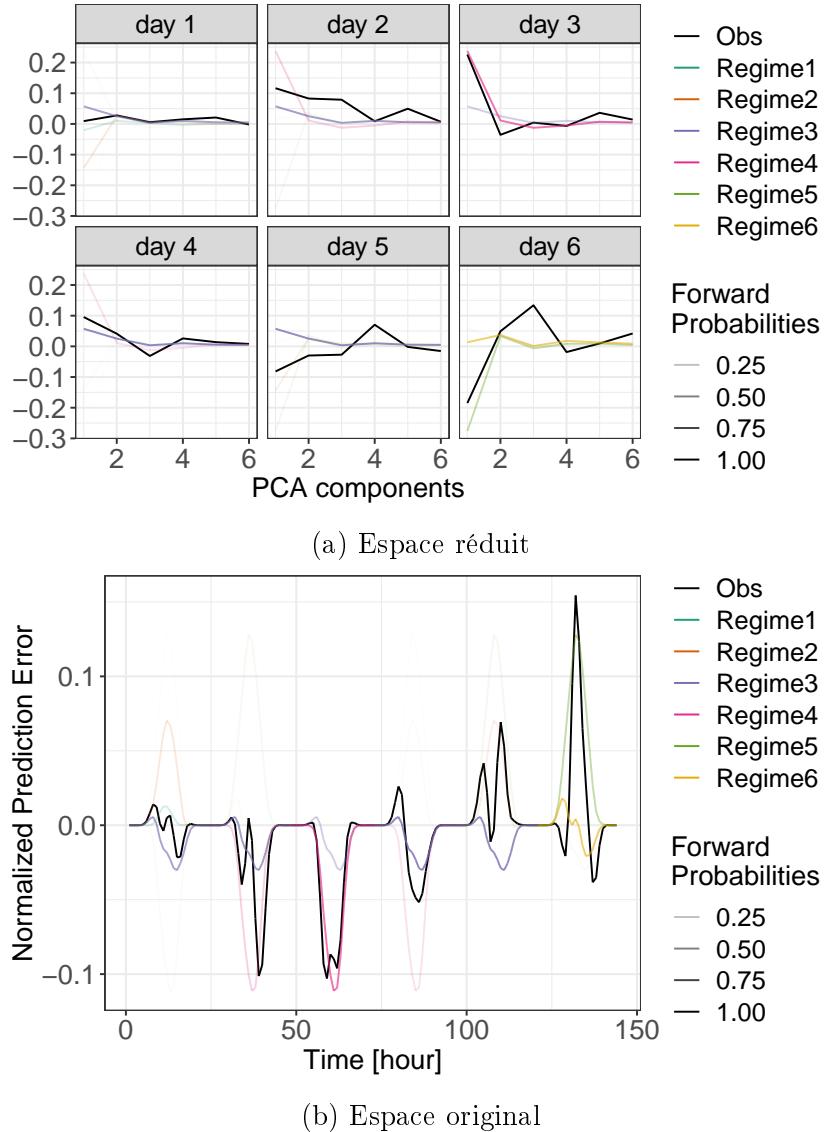


FIGURE 5.10 – Probabilités d'appartenance de 6 jours consécutifs

ANNEXE A

Première version de la publication
concernant les erreurs de prévisions
éoliennes

Modeling of the wind power forecast errors and associated optimal storage strategy

Abstract—Production forecast errors are the main hurdle to integrate variable renewable energies into electrical power systems. Regardless of the technique, these errors are inherent in the forecast exercise, although their magnitude significantly vary depending on the method and the horizon. As power systems have to balance out these errors, their dynamic and stochastic modeling is valuable for the real time operation. This study proposes a Markov Switching Auto Regressive – MS-AR – approach. The strength of such a model is to be able to identify weather types according to the reliability of the forecast. These types are captured with a hidden state whose evolution is controlled by a transition matrix. The autocorrelation and variance parameters of the AR models are then different from one state to another. After having validated its statistical relevance, this model is used to solve the problem of the optimal management of a storage associated with a wind power plant when this virtual power plant must respect a production commitment. The resolution is carried out by stochastic dynamic programming while comparing the proposed MS-AR with several other models of forecast errors. This illustrative problem highlights the improvements made by a fine modeling of forecast errors.

Keywords : forecast errors, wind energy, Markov Switching Auto Regressive, Stochastic Dynamic Programming

NOMENCLATURE

ΔP	forecast error of the wind power	W
ΔT	time step	h
\tilde{P}	Wind power generation forecast	W
E_{sto}	rated capacity of the storage device	Wh
f	cost function of the storage management problem –	
f_{dyn}	dynamic equation of the system	–
P	actual wind power generation	W
P_{loss}	power of the losses int the storage device	W
P_{sto}	power exchanged with the storage device	W
SoE	State of Energy of the storage device	–
x, X	state variable of the system, set of the discretised values of the state variable	
AR	Auto Regressive	
MS-AR	Markov Switching Auto Regressive	

I. INTRODUCTION

The integration of variable renewable energies into electrical systems is mainly hampered by the difficulty of forecasting their electricity production [8], [30]. This low predictability compels the power grid as a whole to compensate for their fluctuations in real time. This may involve adjusting production [33] and consumption – Demand Side Management [28]– or using storage [9]. This global problem,

which involves all the players in the electricity networks, is already sensitive when it comes to planning operations in advance on the basis of forecasts. Dynamically adapting the planned schedule is therefore all the more difficult when the deadline arrives and actual production is accessible.

Because of the burden of forecast errors, a major share of the literature is therefore devoted to forecasting renewable energy production. Several techniques are implemented and their complementarities make it possible to refine the forecast progressively when more and more information becomes available while approaching the deadline [18], [25]. These include global numerical weather prediction – NWP – models [26] that provide a quality forecast with a horizon of several days. Statistical models allow this forecast to be improved using time series [6] or neural networks [15], [27]. Finally, very short term forecasting can be done by means of imagery, satellite or by fisheye camera.

Nevertheless and however good they may be, these forecasts are necessarily tainted by an irreducible error. This error is inherent in the weather forecasting task. Recent models therefore provide information on the reliability of their forecast, in the form of an error range or ensemble forecast [25]. First of all the characterization of these errors [5] is important to help the progress of forecasting models. Moreover, this characterization is useful for electricity networks in order to anticipate sufficient operating reserves and infrastructure [10], [31].

However, error modeling must go beyond a statistical description [29]. Indeed, their dynamic behavior is also crucial : how does the error evolve over time, will it be prolonged or not [10] ? If such information are available, this will be most helpful to decide how the error should be counterbalanced. Calling upon storage resources is the easiest bet as long as their capacity permits but starting up a backup power plant when it is mandatory needs to be anticipated. Moreover, it seems obvious that error modeling must be stochastic modeling [16], [23].

This study focuses on wind energy forecast errors. Previous research on this issue has extensively used Auto Regressive Moving Average models – ARMA [20], [10], [31]. Indeed, one may consider as natural that the forecast error signal will most probably follow the same dynamics as the signal to be predicted. It is therefore relevant to rely on ARMA models which are widely used to predict wind power. However, these models do not capture the diversity of regimes that wind generation may encounter. Like any weather variable, it is driven by weather types that can radically change the behavior of this variable. Several

studies devoted to wind speed forecasts have made use of this idea by introducing Markov Switching Auto Regressive MS-AR models [3], [34], [22]. The main idea of these models is that the parameters of the AR model are not unique but are determined by a hidden state whose evolution follows a Markov chain. The characteristics of the signal can consequently be very different from one time step to the next one. This approach has proven its effectiveness in capturing different wind regimes. However, weather types have also an impact on forecast errors. Some types of weather produce easily predictable wind conditions – hence small forecast errors – while others are much more chaotic and produce large errors.

The objective of this study is therefore to propose a modeling of wind production forecast errors by the mean of a Markov Switching Auto Regressive model, as well as to highlight its relevance. Section II will be dedicated to the presentation of these models and their validation by several statistical criteria. Other simpler models will also be introduced so as to allow comparison. Section III presents a representative application of such a forecast error model : the optimal management of a storage associated with a wind power plant that must meet a generation commitment. This optimal management problem will be solved using stochastic dynamic programming. As this algorithm relies on a stochastic modeling of the future, the resolution will be carried out several time on the basis of various forecast errors models in order to validate the added value of an MS-AR model. This resolution method will be carried out on the case study presented section IV. The section V will finally outline the results of this case study and how a better modeling of the forecast errors may improve the overall performance.

II. MODELING OF THE WIND POWER FORECAST ERRORS

A. Data description

Studying the forecast errors requires datasets with informations on both wind power actual realization and its associated forecast. While actual production can be directly measured, the production forecast first takes the form of a wind speed forecast delivered by meteorologists weather models. This forecast must then be transformed into a production forecast. An open dataset of the production forecast – provided by the wind generators to the local grid operator – and the realization is here provided by the Bonneville Power Administration [2]. The time series consists of 9 years of data, from 2009 to the present, with one input every 5 minutes. Years from 2009 to 2017 have been used as a train dataset and the period from January 1st 2018 to May 2nd 2018 as a test dataset. It includes aggregated forecast and realization over BPA balancing authority area. In order to improve the generality, this time series has been normalized according to the installed capacity of wind power generation [2] and reduced to an hourly time step by averaging. The convention used to define prediction errors is

$$\Delta P = P - \tilde{P}$$

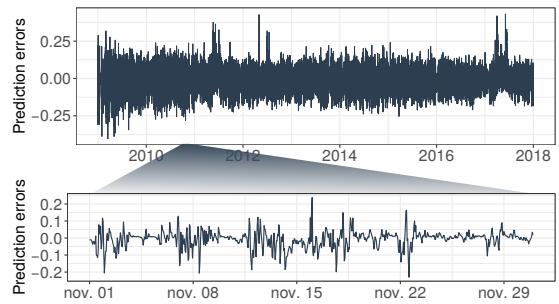


Fig. 1: Normalized prediction errors for 2009 - 2017 and a zoom on November 2010

where \tilde{P} stands for the forecast and P for the actual realization. The time series is shown in Figure 1.

B. Models descriptions

In order to model forecast errors, MS-AR models are here defined. The autoregressive AR models are also recalled, on the one hand because they are at the genesis of MS-AR, on the other hand because they will be used thereafter to provide a benchmark for modeling performance. Along this section, the generic notation Y_t is used to denote the observation at time t in any model. All of the models here introduced are then identified on the previously described time series of prediction errors (ΔP_t).

a) Auto-Regressive (AR): AR(p) processes model the next time step as a linear combination (1) of the p previous time steps plus a white noise ϵ with unit variance and an intercept a_0 . ϵ is also called *innovation*.

$$Y_t = a_0 + a_1 Y_{t-1} + \dots + a_p Y_{t-p} + \sigma \epsilon_t \quad (1)$$

Conventional statistical models such as the AR and their extensions to SARIMA are commonly used in wind speed and power forecasting. However, with successive improvements in forecasts accuracy and as stated by [14], these models are now more considered as reference models.

b) Markov Switching Auto Regressive (MS-AR): MS-AR models allow to model a time series by a mixture of several AR processes. These models have been introduced initially by [11] to capture markedly different regimes in economy related time series. As weather related time series also exhibit weather regimes, typically cyclonic and anti-cyclonic (e.g. Figure 1) such models seems particularly adapted for wind power forecast errors. In particular, these unsupervised models demonstrated their ability to associate weather types with their latent state, by definition unobservable. The switches between the different regimes are described by a latent variable π whose dynamic is driven by a Markov chain. Therefore the Markov fundamental property (see (2)) is respected.

$$\mathbb{P}(\pi_{t+1} | \pi_t, \pi_{t-1}, \dots, \pi_1) = \mathbb{P}(\pi_{t+1} | \pi_t) \quad (2)$$

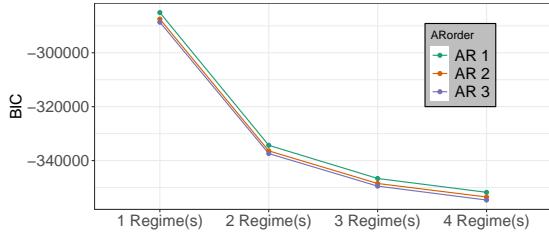


Fig. 2: Bayesian Information Criteria for different model parameters

Let us denote M is the number of possible states of the latent variable. The transition probabilities are defined by a $M \times M$ matrix denoted Γ . An element γ_{ij} represent the probability to switch from π_i to state π_j . Another fundamental property of the MS-AR model is that the observations are conditionally independent regarding to the current latent state and the p previous observations. In other words, at time t , given the latent state π_t ,

$$Y_t = a_0^{(\pi_t)} + a_1^{(\pi_t)} Y_{t-1} + \dots + a_p^{(\pi_t)} Y_{t-p} + \sigma^{(\pi_t)} \epsilon_t$$

In practice, given a time series, the MS-AR model is calibrated by maximum likelihood. Furthermore, the Viterbi algorithm allow to estimate for all t the smoothing probability $\mathbb{P}(\pi_t | Y_1, \dots, Y_T)$ and deduce the most likely sequences of latent states. Usually, the number of regimes M and the order of the AR processes p are chosen according to the Bayes Information Criterion (BIC). This criterion computes a compromise between a model with a high likelihood and low number of parameters. In practice, the model with the smallest BIC should be retained.

C. Results

Identifying a MS-AR model first requires to decide an order p for the AR models and the number of hidden states M . Table I reports BIC values for MS-AR models with $p = 1$ to 3 and M varying from 1 to 4. Remark that $M = 1$ corresponds to the AR model.

TABLE I: BIC Values obtained for different MS-AR parameters

	AR(1)	AR(2)	AR(3)
M = 1	-285080	-287490	-288646
M = 2	-334339	-336405	-337408
M = 3	-346642	-348507	-349488
M = 4	-351787	-353521	-354654

A first effect shown in table I is that the larger M the smaller the BIC value. The improvement brought by the number of possible hidden states can be contextualized by the notion of weather type. This is a meteorological phenomenon that cannot but poorly be described in terms of a finite number of states. Nevertheless, the interpretability of the results quickly becomes impractical when $M \geq 4$. For this reason the 3 state hidden model is chosen later, despite the slight improvement made by the MS(4). **RLGL : quasi**

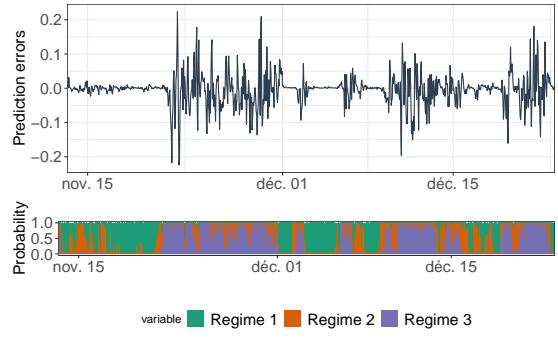


Fig. 3: Exemple of forward probabilities for a given sequence of the time series.

plus d'amélioration avec l'AR(3) et le MS(4) Secondly the decrease of BIC while increasing the auto regressive order p brings a much smaller improvement. In a MS-AR model, increasing the AR order introduces a significant number of new parameters because of the several hidden states. Moreover a longer memory can be computationally expensive in some practical applications. As will be discussed in the following sections, these models are likely to be used to solve stochastic optimization problems. Some of the algorithms used may be sensitive to the dimension of the problem. One can think in particular of the case of dynamic programming and its curse of dimensionality. As adding an extra memory therefore implies adding a new dimension, an AR(2) model will therefore be preferred for the remainder of the study. The previous studies dedicated to the modeling of wind power forecast errors be the mean of AR models [10] – which did not take into account hidden states – were first order models to the best of our knowledge.

The remainder of this section will therefore consider the AR(2) and MS(3)-AR(2) models. The AR(1) and MS-AR(1) versions will also be used in the following sections to compare the impact that different modeling of various complexities may have on the final applications. Model parameters are shown in table II. The transition matrix is diagonally dominant which entails that the regimes are relatively stable. The mean duration of sojourn in the regimes 1, 2, and 3 are respectively 10 hours, 7 hours 38 minutes and 16 hours 40 minutes. Indeed, weather conditions are also relatively stable on a hourly scale. To have almost impossible transitions between states 1 and 3 is to say that one does not abruptly move from one weather type where predictions are reliable to one where they would be very uncertain. This seems consistent with an evolution in weather conditions and supports the idea that these states are associated with weather types. The observed time series inside the various regimes have different innovation variances. That means that they are associated with weather types characterized by their predictability (strongly, moderately or weakly in this case). This feature is illustrated in figure 3 where a sample of the time series and its associated forward probabilities are shown. One also notices that the parameters of the AR(2) are

TABLE II: Fitted parameters of the MS(3)-AR(2) and AR(2)

Reg.	Transition matrix			AR parameters				
	1	2	3	a_0	a_1	a_2	σ	
1	0.90	0.10	$8e^{-6}$	$5e^{-4}$	0.64	-0.08	$9e^{-6}$	
2	0.04	0.87	0.08	$5e^{-4}$	0.73	-0.12	$3e^{-4}$	
3	$5e^{-13}$	0.06	0.94	$-4e^{-3}$	0.67	-0.2	$3e^{-3}$	
AR(2)								
1		1.0	0.0	0.0	$-1.5e^{-3}$	0.68	-0.17	$-4e^{-3}$

very similar to those of the third regime of the MS(3)-AR(2). Therefore these two models may in some circumstances behave almost identically.

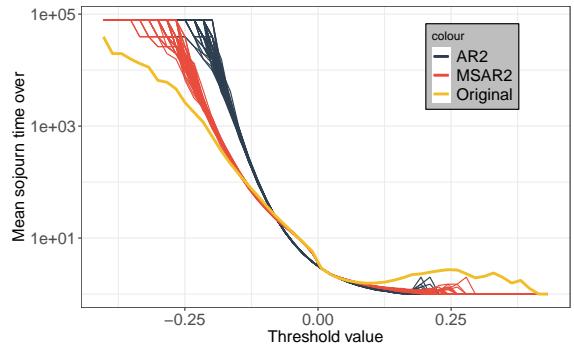
Both AR and MS-AR models allow to easily generate synthetic data. This feature can be used in order to infer how some specific statistics characteristics are captured by comparing synthetic and original time series. Figure 4 show the mean sojourn time over threshold values and the number of up-crossings for synthetic data generated by the previous models and the original series. On both the MS-AR model show significant improvements compared to the AR model. Figure 4b shows that the AR model overestimate variability for small errors and whereas the MS-AR overlays the original series's up-crossings. One can note that, for both models, the 50 scenarios have very little variation between one another. Figure 4a shows that both models fail to reproduce extreme prediction error. In this case, the distinction between scenarios is very clear, especially for extreme values. This is due to the fact that very slight variations in the maxima reached will by definition have a very strong impact on average sojourn time. However the MS-AR brings significant improvements in both capturing the dynamic of small prediction errors and allowing for higher absolute errors to be reached.

According to [19] and [13] a classical method to estimate the reliability of forecast errors is to compare the following metrics : Mean Absolute Error (MAE), the model bias defined as the average error (BIAS) and the Root Mean Squared Error (RMSE). More specifically, [19] propose to use the improvement score to compare two models which is defined in (3) where EC stands for one of the metrics mentioned before and *new* and *ref* point out respectively to the new model and the reference model.

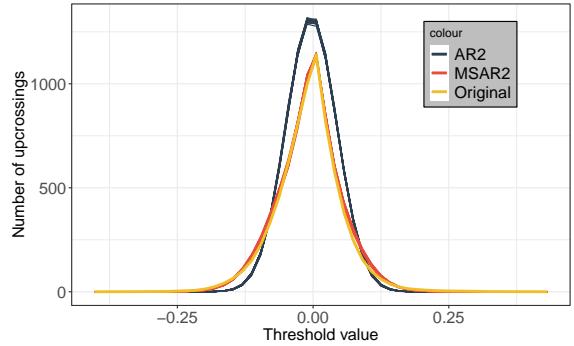
$$ISC(new, ref) = \frac{EC_{ref} - EC_{new}}{EC_{ref}} \quad (3)$$

These results, calculated on the test dataset, are presented in table III. One can notice that the MS-AR improves significantly on both MAE and RMSE scores while enhancing BIAS. However BIAS magnitude shows more that both models do perform well on this metric. After a few forecasting steps, the Markov chain tends to its stationary distribution and therefore the positive impact of the initial probabilities to belong in a particular state has vanished. This explains why overall performances on longer forecast horizons tends to decrease.

Figure 5 presents some sample scenarios generated by



(a) Mean sojourn time over threshold values



(b) Upcrossings of threshold values

Fig. 4: statistical characterization of the real series as well as 50 synthetic series of the same length generated by the AR(2) and MS(3)-AR(2) models

TABLE III: Different scores per scenario obtained by the AR(2) and MS(3)-AR(2) on the test dataset with 100 scenarios generated per time steps

Metric	12H forecast horizon			24H forecast horizon		
	AR	MS-AR	ISC	AR	MS-AR	ISC
MAE	0.0450	0.0408	[9.3%]	0.0489	0.0452	[7.6%]
RMSE	0.0590	0.0544	[7.8%]	0.0632	0.0598	[5.4%]
BIAS	0.0049	0.0051	[-4.1%]	0.0054	0.0055	[-1.9%]

the AR(2) and MS(3)-AR(2) models. In the MS(3)-AR(2) case, the first hidden state of every scenarios are determined through random draws from a multinomial distribution characterized by the forward probability $\mathbb{P}(\pi_t|Y_1 \dots Y_t)$ of the last known observation. Transitions from one hidden state to another at every time steps are then fully characterized by the transition matrix Γ . For both models, observations are generated through an auto regressive process. Figures 5d shows that, through the forward probabilities, the MS-AR model acknowledges that it is in a regime with small prediction errors (as illustrated by the 90% simulation interval). On the other hand, Figure 5c points out that the AR model is not able to take this information into account and the dispersion of scenarios remains the same all along. Figure 5b shows that when initialized in a less predictable weather type, the MS(3)-AR(2) produces scenarios that are more variable

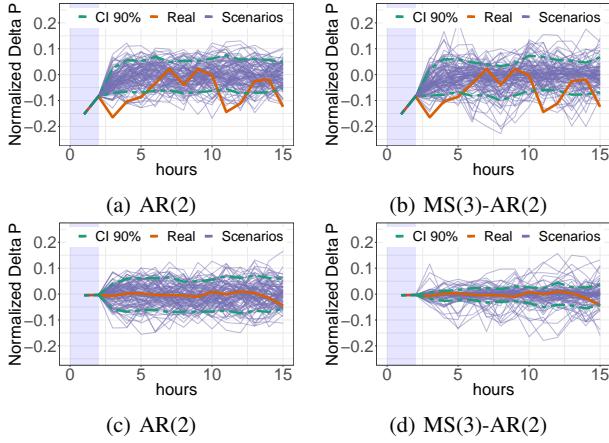


Fig. 5: 50 15-hour scenarios generated by the AR(2) and MS(3)-AR(2) models initialized on a sample of the test data and their associated 90% simulation interval

and will occasionally reach more extreme values than those reached by the AR(2) (see Figure 5a).

This section presented the MS-AR model that is here suggested to describe wind generation forecast errors. After having discussed the relevant orders for the AR model and for the number of hidden states, several statistical scores were used to validate that these models allow to reflect accurately the behaviour of forecast errors with a significant improvement in comparison with usual AR models. However the reader's attention should be drawn to the fact that the value of the model coefficients are very largely sensitive to the series on which it is identified. An illustration of this sensitivity is presented in Appendix A, where the same model is applied to another series [1]. In order to go beyond the validation of the statistical relevance of the proposed model, the following sections propose to use them to solve a canonical problem in order to evaluate the improvement on the final performance that can be brought about by the modeling of forecast errors.

III. OPTIMAL STORAGE STRATEGIES

Because of the forecast errors of wind power generation – as well as generally speaking variable renewable energies – their management within power systems is made challenging in many different ways. Examples include their integration into energy markets [21], [12], compliance with grid constraints for voltage and frequency [17], supply of reserves [32], *et cetera*. As all of these contexts come under the hazard of forecast errors, using models to anticipate them can therefore be useful in many different ways.

In many cases, combining the renewable power plant with a storage unit is a particularly relevant solution to overcome the intrinsic restrictions of variable renewable energies [35], [36]. The issue of optimal management of this storage – how much power must be exchanged with the storage unit at any one time – is then a problem that has two characteristics that require the use of a forecast error model.

- The problem is dynamic : using a storage unit introduces a temporal coupling between successive moments due to the integrating behaviour of the storage. This behaviour can be described by a dynamic equation – which is considered deterministic – over a time step ΔT :

$$SoE(t + \Delta T) = SoE(t) + \frac{\Delta T \cdot (P_{sto} - P_{loss})}{E_{sto}} \quad (4)$$

where SoE designates the state of energy of the storage unit – which is bounded between 0 and 1, P_{sto} stands for the setpoint of the power exchanged with the storage, P_{loss} indicates the internal losses of the device and E_{sto} refers to the storage capacity, *i.e.* the maximum energy that can be stored into it. This temporal coupling will therefore make it necessary to anticipate future moments, hence the need for a model describing the evolution of the error.

- The problem is stochastic : since forecast errors are inherently imperfectly controlled, the model used to anticipate them must be a stochastic model. Only then will it be possible for each agent to determine what could be the best decision to anticipate an uncertain future, depending of his own risk policy.

A. Optimal management of a storage

The scope of this study aims to highlight how a better modelling of forecast errors can improve the final performance of the renewable plant. In this section, the problem of managing a storage associated with a wind farm will therefore be used for this purpose. This problem is at this stage formulated in a generic way and can be adapted to as many situations as possible. The resolution method used is also chosen to require few assumptions and to be easily adaptable. In the next section, this method will be applied to a basic and representative case study to quantify the contributions of the forecast error model.

A wind power plant associated with a storage unit is therefore considered. The goal of this virtual power plant is to minimize a cost over time. This cost function is the sum over time of instant costs. The optimal management problem can therefore be stated in the following form :

$$\min_{P_{sto}(t)} \mathbb{E} \left\{ \sum_{\tau=t}^{\infty} f(x(\tau), P_{sto}(\tau), \Delta P(\tau)) \right\} \quad (5a)$$

such that, $\forall t, \forall \tau,$

$$P_{sto}^b \leq P_{sto} \leq P_{sto}^{\sharp} \quad (5b)$$

$$0 \leq SoE \leq 1 \quad (5c)$$

where f refers to the cost function which allows to evaluate the instantaneous cost at each moment. This depends on the system state x , the command P_{sto} and the forecast error with random behavior ΔP .

Note : the capacity to shed producible is not taken into account in this case. In an actual situation, the renewable plant operator can deoptimize the conversion efficiency – via the wind turbine blade pitch or the PV panel operating

point – which would provide an additional decision variable. Although this may well be taken into account in the resolution method that is presented, this possibility will not be exploited later on because this study focuses on the impact of the forecast error model. Indeed, taking load shedding into account would have several consequences which would make more difficult the interpretation of the results. First it would introduce an asymmetry between the cost linked to positive and negative errors. Secondly the calculation of the commitment would no longer be equal to the expectation of the forecast.

B. Resolution using Stochastic Dynamic Programming

Several methods may be considered to overcome such a problem. Their required properties are to be able to support the stochastic and multitemporal characters of the problem. Methods like Model Predictive Control could be considered. Nevertheless, such methods would require that the calculation of the optimal decision should be carried out at each moment according to the present situation, which would potentially require a substantial real-time calculation burden.

In order to support the stochastic and temporal coupling characteristics of the problem while minimizing the real-time computation cost, the resolution of the problem (5) is performed here by the use of stochastic dynamic programming [4]. This algorithm allows to establish an optimal strategy which describes the best decision to take for any configuration of the state vector. The result obtained is therefore not only the decision to be taken in the current situation, but the optimal decisions for all possible configurations. The real-time use of this strategy then consists of a simple interpolation of the matrix describing the optimal strategy.

Stochastic dynamic programming is based on the resolution of the Bellman equation. This allows the calculation of the costs associated with each configuration of the state vector when the optimal decision is applied. It is calculated from the final state of the system at the T horizon and going back in time.

$$V(T, X) = 0 \quad (6a)$$

$\forall t < T, \forall x \in X,$

$$V(t, X) = \min_{P_{sto}} \underbrace{f(x, P_{sto})}_{\text{instantaneous cost}} + \underbrace{\mathbb{E}_{\Delta P} \left(V(t + \Delta T, f_{dyn}(x, P_{sto})) \right)}_{\text{expectation of the future cost}} \quad (6b)$$

where

- The horizon T of the problem is not associated with any particular value. This final value is therefore initialized to a zero. However, no horizon value could be preferable in the context of managing storage associated with a wind power plant. Instead it would be desirable for the problem to have an infinite horizon rather than become myopic beyond a given time frame.

The resolution is thus iterated back in time until the optimal strategy converges, so that it does not change from one iteration to the next. We then obtain a strategy considering an infinite horizon of optimization.

- f_{dyn} represents the dynamic function of the system. This links the current state and the current control to the future state of the system :

$$x(t + \Delta T) = f_{dyn}(x(t), P_{sto}(t)) \quad (7)$$

In the present case, this dynamic function includes not only the deterministic component of the equation 4, but also a random component due to the evolution of the forecast error which cannot be perfectly anticipated. Therefore all quantities that are involved in the forecast error model must be included in the state vector in order to be able to evaluate the expectation of the forecast error at the next time step. As this study compares several models, the composition of the state vector will then differ from one resolution to another. Prediction error models described as uniform noise, persistence or a first-order AR model will have the state vector :

$$x = \begin{pmatrix} SoE \\ \Delta P \end{pmatrix} \quad (8)$$

In the case of the resolutions involving a second order AR model, all possible configurations of the forecast error at the two previous time steps must be enumerated. The forecast error ΔP then appears twice in their state vector, a first time for the current time step forecast error and a second time for the forecast error at the previous time step. Finally resolutions based on MS-AR models add the hidden state into their state vector :

$$x = \begin{pmatrix} SoE \\ \Delta P \\ \pi \end{pmatrix} \quad (9)$$

- the f notation represents the instant costs of the problem. This generic notation emphasizes that the method which is here used to solve the optimal management problem is almost fully independent from the formulation adopted to estimate instantaneous costs. Any formulation depending on the system status x and its command P_{sto} can be adopted. The section IV will present the example of formulations adopted for the application case. During the optimal strategy calculation, the cost f is evaluated for each configuration of the x state vector, hence the disappearance of the time index.

As described during section III, the resolution of this problem is based on a model evaluation the expectancy of the future cost, depending on the stochastic variables of the problem. To compare the added value of a fine forecast error model, the resolution will therefore be carried out several times using the following models.

- a uniform distribution of the errors : all possible values are supposed to be equiprobable.

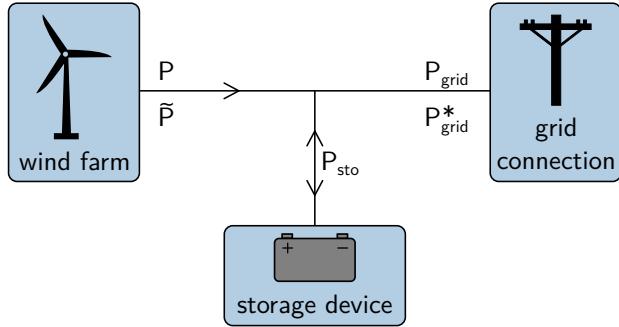


Fig. 6: Situation being considered for the case study.

- a persistence of the previous error : the error at the previous time step is supposed to be equal to the current value.
- AR(1) and AR(2) models.
- MS(3)-AR(1) and MS(3)-AR(2) models.

Finally a last resolution is carried out where the future is not anticipated. This resolution is blinded and does not take into account the future forecast errors.

IV. CASE STUDY

In order to present an application of the optimal strategy described in the previous section, a case study is introduced here. The objective of this case study is to present as straightforwardly as possible the impact that the forecast error model can have on the decisions to be made. A situation as simple as it can is therefore under consideration. A wind power plant is considered in association with a storage unit. This virtual power plant is supposed to be compelled to a production commitment. It can arise from a participation in an energy market, a constraint linked to the electricity network, a regulatory obligation, *et cetera*. In the case where the cost functions of the problem are symmetrical and the possibility of load shedding is not taken into account, the optimal commitment of the plant is then equal to the expectation of the forecast :

$$P_{grid}^* = \tilde{P} \quad (10)$$

The objective function to be minimized encompass two terms :

- losses within the storage system. Many technologies of components can be considered in association with a renewable power plant depending on the reactivity and amount of stored energy required. Nevertheless, losses are inherent in each exchange of energy while it is charged or discharged. The case of lithium-ion batteries is considered here, as this technology is used in a wide variety of applications. A quadratic model is adopted to describe the losses :

$$P_{loss} = a_{loss} \cdot P_{sto}^2 \quad (11)$$

Here it is necessary to decide how to quantify this lost power in an objective function. In fine this choice depends on the case study and the business model

TABLE IV: Values of the parameters and coefficients used within the case study **RLGL : ATTENTION il faut le mettre à jour**

error model	myopic	uniform	AR(1)
kWh _p /day	32.6	20.4	12.5
error model	AR(2)	MS-AR(1)	MS-AR(2)
kWh _p /day	12.4	12.2	12.1

considered. We choose here to adopt a quantification in an equivalent primary energy in order to keep as much as possible the generality of this study. The wind power plant is therefore considered to be characterized by a yield over life cycle η_{wind} , i.e. the total energy it can produce – depending on site weather conditions – divided by the primary energy that was required for its construction. The losses are therefore a degradation of this efficiency and an associated loss of primary energy :

$$C_{loss} = \frac{\Delta T \cdot P_{loss}}{\eta_{wind}} \quad (12)$$

- the penalty for deviations from the commitment. In the various situations where the production commitment obligation is present, a penalty is associated with non-compliance. In a similar way to the quantification of the cost of losses, the actual quantification of a deviation from the commitment will depend on the final application case. In this study, therefore, we only consider the case of a quadratic penalty for the mismatch :

$$C_{mis} = a_{mis} \cdot (\Delta P - P_{sto})^2 \quad (13)$$

in the case where the commitment is set as equal to the forecast expectancy.

V. RESULTS AND DISCUSSION

The problem that is described in the section IV is solved several times using the Bellman equation (6) using the different forecast error models described in the section II. This section will first present the observable consequences on the optimal strategies thus obtained. In a second step these optimal strategies will be applied to the time series of forecast errors in order to compare temporal behaviors and calculate the obtained performances.

A. Optimal storage strategies

The result of the resolution of the Bellman equation is a response surface associating the optimal storage power with every configuration of the state vector – discretized on a grid fine enough to guarantee a satisfactory interpolation between two grid points. These response surfaces are therefore of the same dimension as the state vector. The Figure 7 represents some of the calculated strategies. The interpretation of these illustrations is that if the system is in a configuration where the energy state is x on the abscissa and the forecast error is

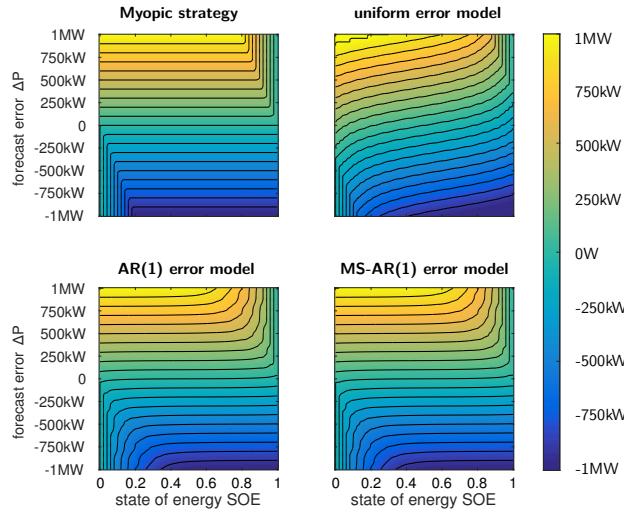


Fig. 7: Cross-sections of optimal storage strategies obtained with different forecast error models for $E_{sto} = 5$ MWh. In the case of the *Markov Switching Auto Regressive* model, the cross section is in the $\pi = 3$ hidden state.

y on the ordinate, then the storage power described by the optimal surface must be applied.

- The first panel in this figure represents the optimal strategy when no anticipation is made. The storage then tries to perfectly compensate for the forecast error until it is too much or too little charged. The isopowers curves are therefore perfectly horizontal, until the storage can no longer provide.
- On the second panel, the forecast error is modelled by uniform noise. All error values are therefore equiprobable at the next instant, regardless of the current state. This entails a strong forward-looking behaviour of the optimal strategy. Indeed even in a configuration of the state vector where the error is low, it is nevertheless necessary to anticipate that very strong errors can occur in the next time steps. As a result, the isopowers curves are very steep, which means that the forecast error is never perfectly compensated, but only attenuated.
- On the third panel, the optimal strategy is determined using an AR(1) model. Although rudimentary – this model uses only autocorrelation and a standard deviation – it allows much better anticipation of future errors. Especially when the errors are of small amplitude, almost total compensation is possible because it can be reliably anticipated that the error will remain of small amplitude during the next few time steps.
- Finally, on the last panel of the figure, the optimal strategy is determined using a MS-AR(1) model. The overall behaviour is therefore very similar to that of the strategy based on an AR(1). However, this is a cut for only one of the three hidden states of the model. Indeed, since the MS-AR model introduces the

possibility to switch from one hidden state to another, the strategies corresponding to these three states differ finely according to the standard deviation and the correlation of errors from one hidden state to another. When being used in real time, the probabilities of belonging to a hidden state are reconstructed according to the observations available up to now – for example by a Viterbi algorithm. The storage power decision is then the weighted average of the decisions for each of the 3 hidden states.

Note : although persistence is usually an excellent way of anticipating weather phenomena with ease, this model here leads to dreadful results. Indeed, within the context of the resolution here presented, the persistence suggests that the value of the forecast error that is currently observed would persist until the end of time. In such a case, any storage system regardless of its capacity will eventually be saturated. The optimal storage strategy when this model is used is therefore to do nothing. This case is mentioned here to draw the reader's attention to this adverse effect.

B. Application to a forecast error time serie

The previously described storage strategies are applied to the time series of the wind power forecast error which has been presented during section II to identify the various models. The initial state of energy of the storage is set to $SoE(t = 0) = 0.5$. At each time step of this simulation, the state vector is formed, then the optimal strategy is interpolated to determine the storage power. The dynamic equation of the system is then applied to progress to the next time step. In the case of MS-AR models, the decision being applied is the average of the optimal decisions for each hidden state, weighted by the probabilities of belonging to each state. These forward probabilities are established on the basis of the observations available so far. The Figure 3 represents an example of temporal evolution of these probabilities. The succession of regimes is clearly observable with the naked eye : periods of very small errors are followed by periods when errors are large and chaotic. This succession is translated by the evolution of the probabilities of belonging to each hidden state, represented in the second panel of the figure. The state 1 – in red – is most likely when the errors have a very small amplitudes. The state 3 – in blue – appears to reflect high amplitude and low correlation errors. Intermediate situations are taken into account by a high probability of belonging to regime 2 – in green.

On the Figure 8, the temporal evolution of the stored power and the corresponding energy state are represented over a 400 hour sample. The forecast error is shown in blue. The stored power tries to compensate it as much as possible.

- The basic strategy – in red – does not anticipate the future. Therefore it offers perfect compensation but quickly causes storage saturation and then becomes almost useless.
- The second strategy is based on a uniform noise model of the forecast error. It always seeks to maintain an energy state close to 0.5. This is due to the overem-

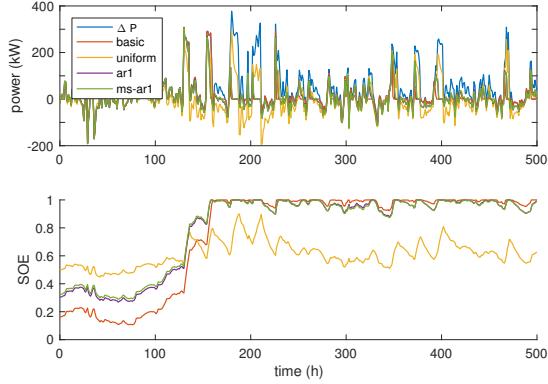


Fig. 8: Top : time trajectories of the forecast error and associated stored power according to various strategies. Bottom : corresponding state of energy of the storage device

TABLE V: Comparison of the total costs obtained by the different management strategies, expressed in primary energy per day

error model	myopic	uniform	AR(1)
kWh _p /day	32.6	20.4	12.5
error model	AR(2)	MS-AR(1)	MS-AR(2)
kWh _p /day	12.4	12.2	12.1

phasis this strategy places on extreme errors. The compensation for current errors is therefore downgraded because of this exaggerated anticipation. It can thus be seen that the stored power deviates significantly from the forecast error to unload the storage.

- Autoregressive models – with or without hidden state – exhibit an intermediate behavior that allows them to largely compensate for forecast errors while controlling changes in the energy state of the storage. The differences between models with and without hidden state are very fine. They are notable when total costs are calculated over the course of a year.

The table V summarizes the total cost evolution according to the different optimal storage strategies. This total cost corresponds to the problem 5 computed throughout the forecast error series. These results first show the reduction of the total cost with each refinement of the model used to describe forecast errors. This highlights that the storage strategy can be significantly improved when it is based on a relevant model. However, it should be noted that the most significant improvement in the results appears as soon as an AR(1) model is used. The transition to an AR(2) and then to MS-AR does continue to improve the final result, but in a much more marginal way. The performance of such a simple model can therefore be underlined here.

VI. CONCLUSION AND FURTHER WORKS

The scope of this paper was the modeling of wind power forecast errors. Such errors are inherent in any forecast,

regardless of the method used to generate it. Therefore a modeling of forecast errors is necessary in any system that relies on a forecast to establish its planning, in order to compensate them in real time. In particular, such a model must support the temporal evolution of errors – dynamic modeling – and the intrinsic uncertainty in the forecast error signal – stochastic modeling.

The temporal autocorrelation of wind generation forecast errors was first pointed out. Moreover, the succession of regimes where errors are of small or large amplitude has been observed. This motivated the identification of a Markov Switching AutoRegressive model for forecast errors. The coefficients of such a model depend on a hidden state that is driven by a Markov chain. The relevance of this model was initially shown on the basis of several statistical scores.

In a second part, this MS-AR model was used to solve an optimal storage management problem within the context of an elementary and representative problem : the management of a battery associated with a wind power plant subject to a production commitment. As the storage must mitigate the forecast errors as best it can, the forecast error modeling is integrated into the resolution using stochastic dynamic programming algorithm. Several error models were also compared to the MS-AR models : autoregressive, uniform and myopic. The contribution of a good error modeling was highlighted, the management strategy based on an MS-AR model having improved the overall performance.

As an outlook to this study, significant improvements remain open in the MS-AR modeling of forecast errors. For instance using different models for different periods of the year seems a natural way to deal with the seasonality of wind energy forecast errors. However, this seasonality could also be integrated into a single model using non-homogeneous transition probabilities. That is, the probabilities of switching from one state to another would not be constant but would depend on an observable variable, such as the calendar date for example. It would then be possible to create states that only appear at specific seasons of the year.

Furthermore taking spatial correlations into account in forecast errors would be a particularly interesting development for electricity network management. Indeed, geographical effects are crucial to cope with the limitations of the power grid within which many wind farms are connected. The grid operator must then ensure that a forecast error on the production injected at one location can only be compensated by another remote equipment within the operating limits of the network. These spatial effects are starting to be taken into account when predicting wind conditions [24], [7].

Finally, the study presented here on the optimal storage strategy should be extended by a study of the optimal sizing of the battery. Sizing and management strategy are indeed mutually dependent : in similar situations, the optimal decision cannot be the same for two batteries having different capacities. The present paper can therefore be a prerequisite for the introduction of forecast error models within a co-optimization approach over life cycle of the design and management, integrating the effects of aging of the storage

device.

REFERENCES

- [1] Actual and forecast wind energy feed-in - TenneT (Available at <https://www.tennet.eu/electricity-market/transparency-pages/transparency-germany/network-figures/actual-and-forecast-wind-energy-feed-in/>).
- [2] BPA : Balancing Authority Load & Total Wind Generation. Available at <https://transmission.bpa.gov/Business/Operations/Wind/>.
- [3] Pierre Ailliot and Valérie Monbet. Markov-switching autoregressive models for wind time series. *Environmental Modelling & Software*, 30 :92–101, April 2012.
- [4] Dimitri Bertsekas. *Dynamic programming and optimal control*, volume 1. Athena scientific Belmont, MA, 1995.
- [5] H. Bludszuweit, J.A. Dominguez-Navarro, and A. Llombart. Statistical Analysis of Wind Power Forecast Error. *IEEE Transactions on Power Systems*, 23(3) :983–991, August 2008.
- [6] Barbara G Brown, Richard W Katz, and Allan H Murphy. Time series models to simulate and forecast wind speed and wind power. *Journal of climate and applied meteorology*, 23(8) :1184–1195, 1984.
- [7] Ahmed Aziz Ezzat, Mikyoung Jun, and Yu Ding. Spatio-Temporal Asymmetry of Local Wind Fields and Its Impact on Short-Term Wind Forecasting. *IEEE Transactions on Sustainable Energy*, 9(3) :1437–1447, July 2018.
- [8] A. Fabbri, T. GomezSanRoman, J. RivierAbbad, and V.H. Mendez-Quezada. Assessment of the Cost Associated With Wind Generation Prediction Errors in a Liberalized Electricity Market. *IEEE Transactions on Power Systems*, 20(3) :1440–1446, August 2005.
- [9] Nicolas Gast, Jean-Yves Le Boudec, Alexandre Proutière, and Dan-Cristian Tomozei. Impact of storage on the efficiency and prices in real-time electricity markets, page 15. ACM Press, 2013.
- [10] Pierre Haessig, Bernard Multon, Hamid Ben Ahmed, Stéphane Lascaud, and Pascal Bondon. Energy storage sizing for wind power : impact of the autocorrelation of day-ahead forecast errors : Energy storage sizing for wind power. *Wind Energy*, pages n/a–n/a, October 2013.
- [11] James D. Hamilton. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57(2) :357, March 1989.
- [12] Li Han, Carlos E Romero, Xuesong Wang, and Liping Shi. Economic dispatch considering the wind power forecast error. *IET Generation, Transmission & Distribution*, 12(12) :2861–2870, July 2018.
- [13] Ian T. Jolliffe and David B. Stephenson, editors. *Forecast Verification : A Practitioner’s Guide in Atmospheric Science*. John Wiley & Sons, Ltd, Chichester, UK, December 2011.
- [14] Jaesung Jung and Robert P. Broadwater. Current status and future advances for wind speed and power forecasting. *Renewable and Sustainable Energy Reviews*, 31 :762–777, March 2014.
- [15] Georges Kariniotakis, G. Stavrakakis, and E Nogaret. Wind power forecasting using advanced neural networks models. *IEEE Transactions on Energy Conversion*, 11(4) :762 – 767, 1996.
- [16] Matthias Lange. On the Uncertainty of Wind Power Predictions—Analysis of the Forecast Accuracy and Statistical Distribution of Errors. *Journal of Solar Energy Engineering*, 127(2) :177, 2005.
- [17] J.A. Peças Lopes, N. Hatziairgyiou, J. Mutale, P. Djapic, and N. Jenkins. Integrating distributed generation into electric power systems : A review of drivers, challenges and opportunities. *Electric Power Systems Research*, 77(9) :1189–1203, July 2007.
- [18] P. Louka, G. Galanis, N. Siebert, G. Kariniotakis, P. Katsafados, I. Pytharoulis, and G. Kallos. Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. *Journal of Wind Engineering and Industrial Aerodynamics*, 96(12) :2348–2362, December 2008.
- [19] Henrik Madsen, Pierre Pinson, Georges Kariniotakis, Henrik Aa Nielsen, and Torben Skov Nielsen. A Protocol for Standardizing the performance evaluation of short term wind power prediction models, page 18.
- [20] Phillip E. Mello, Ning Lu, and Yuri Makarov. An optimized autoregressive forecast error generator for wind and load uncertainty study : An optimized autoregressive forecast error generator. *Wind Energy*, 14(8) :967–976, November 2011.
- [21] Pierre Pinson, Christophe Chevallier, and George N. Kariniotakis. Trading Wind Generation From Short-Term Probabilistic Forecasts of Wind Power. *IEEE Transactions on Power Systems*, 22(3) :1148–1156, August 2007.
- [22] Pierre Pinson and Henrik Madsen. Adaptive modelling and forecasting of offshore wind power fluctuations with Markov-switching autoregressive models. *Journal of Forecasting*, 31(4) :281–313, July 2012.
- [23] Pierre Pinson, Henrik Madsen, Henrik Aa. Nielsen, George Papaefthymiou, and Bernd Klöckl. From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, 12(1) :51–62, January 2009.
- [24] Teng Qijun, Wang Chengfu, Liang Jun, and Liang Zhengtang. Research on modeling spatiotemporal correlation of wind power forecast error on multiple wind farms based on Copula theory. pages 447–450. IEEE, September 2017.
- [25] Ye Ren, P.N. Suganthan, and N. Srikanth. Ensemble methods for wind and solar power forecasting—A state-of-the-art review. *Renewable and Sustainable Energy Reviews*, 50 :82–91, October 2015.
- [26] Y. Seity, P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson. The AROME-France Convective-Scale Operational Model. *Monthly Weather Review*, 139(3) :976–991, March 2011.
- [27] Zhichao Shi, Hao Liang, and Venkata Dinavahi. Direct Interval Forecast of Uncertain Wind Power Based on Recurrent Neural Networks. *IEEE Transactions on Sustainable Energy*, 9(3) :1177–1187, July 2018.
- [28] Goran Strbac. Demand side management : Benefits and challenges. *Energy Policy*, 36(12) :4419–4426, December 2008.
- [29] Lennart Söder. Simulation of Wind Speed Forecast Errors for Operation Planning of Multi-Area Power Systems. *8th International Conference on Probabilistic Methods Applied to Power Systems*, page 7, 2004.
- [30] Falko Ueckerdt, Lion Hirth, Gunnar Luderer, and Ottmar Edenhofer. System LCOE : What are the costs of variable renewables ? *Energy*, 63 :61–75, December 2013.
- [31] Chengfu Wang, Zhengtang Liang, Jun Liang, Qijun Teng, Xiaoming Dong, and Zhaoqing Wang. Modeling the temporal correlation of hourly day-ahead short-term wind power forecast error for optimal sizing energy storage system. *International Journal of Electrical Power & Energy Systems*, 98 :373–381, June 2018.
- [32] Fengyu Wang and Kory W. Hedman. Dynamic Reserve Zones for Day-Ahead Unit Commitment With Renewable Resources. *IEEE Transactions on Power Systems*, 30(2) :612–620, March 2015.
- [33] C.K. Woo, I. Horowitz, J. Moore, and A. Pacheco. The impact of wind generation on the electricity spot-market price level and variance : The Texas experience. *Energy Policy*, 39(7) :3939–3944, July 2011.
- [34] Megan Yoder, Amanda S. Hering, William C. Navidi, and Kristin Larson. Short-term forecasting of categorical changes in wind power with Markov chain models : Forecasting categorical changes in wind power. *Wind Energy*, pages n/a–n/a, June 2013.
- [35] Haoran Zhao, Qiuwei Wu, Shuju Hu, Honghua Xu, and Claus Nygaard Rasmussen. Review of energy storage system for wind power integration support. *Applied Energy*, 137 :545–553, January 2015.
- [36] Yu Zheng, Junhua Zhao, Yue Song, Fengji Luo, Ke Meng, Jing Qiu, and David John Hill. Optimal Operation of Battery Energy Storage System Considering Distribution System Uncertainty. *IEEE Transactions on Sustainable Energy*, 9(3) :1051–1060, July 2018.

APPENDIX

IDENTIFICATION OF THE MS-AR MODEL ON ANOTHER TIME SERIES

As it has been emphasized in section II, the parameters of both the AR and MS-AR processes are highly dependent on the considered time series. To highlight this, table VI shows the parameters obtained on another time series of prediction errors provided by Tennet [1].

One can notice by looking at the variances σ that the regimes seems again to match with weather type characterized by their predictability. Moreover the transition matrix

TABLE VI: Fitted parameters of the MS(3)-AR(2) on another time series provided by Tennet

MS(3) - AR(2)						
Reg.	Transition matrix			AR parameters		
	1	2	3	a_0	a_1	a_2
1	0.93	0.07	$2e^{-4}$	$5e^{-4}$	1.34	-0.41
2	0.03	0.92	0.05	$4e^{-4}$	1.35	-0.42
3	$9e^{-4}$	0.27	0.73	$-2e^{-3}$	1.21	-0.34
						$2e^{-3}$

show that the hidden states are persistent with respectively a mean sojourn time of 14 hours 16 minutes, 12 hours 30 minutes and 3 hours 42 minutes. However the autoregressive coefficients are nothing alike. This means that the parameters shown in this paper are not meant to be transposed as such to another time series. Regardless of the fact that one needs to fit the model himself to use it for its own purposes, it can be seen as a benefit. Indeed, this allow the model to take into account specificities of the considered time series.

Bibliographie

- [1] W. Zucchini, I. L. MacDonald, and R. Langrock, *Hidden Markov Models for Time Series : An Introduction Using R, Second Edition.* CRC Press, Dec. 2017, google-Books-ID : KIWzDAAQBAJ.
- [2] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
- [3] D. Jurafsky and J. H. Martin, “Speech and Language Processing,” Aug. 2017. [Online]. Available : <https://web.stanford.edu/~jurafsky/slp3/>
- [4] J. A. Hartigan and M. A. Wong, “Algorithm AS 136 : A K-Means Clustering Algorithm,” *Applied Statistics*, vol. 28, no. 1, p. 100, 1979. [Online]. Available : <https://www.jstor.org/stable/10.2307/2346830?origin=crossref>
- [5] C. Bouveyron and C. Brunet-Saumard, “Model-based clustering of high-dimensional data : A review,” *Computational Statistics & Data Analysis*, vol. 71, pp. 52–78, Mar. 2014. [Online]. Available : <http://linkinghub.elsevier.com/retrieve/pii/S0167947312004422>
- [6] T. Gneiting, F. Balabdaoui, and A. E. Raftery, “Probabilistic Forecasts, Calibration and Sharpness,” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 69, 2007. [Online]. Available : <http://www.jstor.org/stable/4623266>
- [7] P. Pinson and J. Tastu, “Discrimination ability of the Energy score,” Technical University of Denmark (DTU), Technical Report 15, 2013.
- [8] R. L. G. Latimier, “Gestion et dimensionnement d’une flotte de véhicules électriques associée à une centrale photovoltaïque : co-optimisation stochastique et distribuée,” PhD Thesis, Université Paris-Saclay, 2016.