

Modélisation et Prévision de la Production Électrique en Corse à l'Aide du Modèle ELM : Challenge Data Viz

Cyril VOYANT *

12/10/2024

1 Introduction

La gestion de l'énergie, notamment pour des îles comme la Corse, est un défi majeur en raison de leur isolement et de leur dépendance aux énergies renouvelables. La Corse n'étant pas raccordée au réseau électrique continental et sans accès à l'énergie nucléaire, elle dépend fortement des sources locales comme l'hydroélectricité, le solaire, et le thermique pour répondre à la demande. De plus, des prévisions précises sont essentielles pour éviter une surproduction coûteuse ou des déficits d'énergie.

L'objectif principal de cette étude est de proposer une méthode de prévision à court terme de la production d'électricité en Corse à l'aide d'un modèle *Extreme Learning Machine* (ELM). Ce modèle vise à prédire la production pour plusieurs horizons temporels (10 horizons distincts) à partir de données historiques, et à comparer ces prévisions à un modèle de persistance, qui est souvent utilisé comme base pour les prévisions énergétiques. Nous explorerons les gains de précision par rapport à ce modèle de référence, tout en détaillant l'impact potentiel sur les coûts et les émissions de gaz à effet de serre.

2 Périmètre de l'étude

Cette étude se concentre sur les données de production énergétique en Corse pour plusieurs types de production : thermique, hydroélectrique, solaire, éolienne, biomasse, et importations au pas horaire (mise à disposition par edf via <https://opendata-corse.edf.fr/>). Ces données sont obtenues à partir de relevés horaires sur une période donnée. Les prévisions sont réalisées pour des horizons temporels allant de 1 à 10 heures. Pour chaque horizon, nous évaluerons les erreurs de prévision par rapport aux valeurs réelles, avec un intérêt particulier porté à l'amélioration par rapport à un modèle de persistance.

Le modèle de persistance est défini comme suit :

$$\hat{y}_{t+h} = y_t$$

où \hat{y}_{t+h} est la prédiction à l'instant $t + h$, et y_t est la valeur réelle à l'instant t . Une amélioration de 10% par rapport à ce modèle est représenté par un gain de 0,1.

Pour ce défi, dix courbes sont disponibles, chacune correspondant à un horizon temporel spécifique et présentant les mesures observées ainsi que les prévisions associées. Il est possible d'effectuer un déplacement temporel interactif à l'aide de la souris, étant donné que les abscisses des sous-graphes (subplots) sont liées entre elles. De plus, les métriques d'erreur standard ainsi que le gain généré par l'application de la méthode Extreme Learning Machine (ELM) sont visualisables pour évaluer la performance du modèle.

2.1 Le Challenge Data Viz

Cette étude s'inscrit également dans le cadre du *Challenge Data Viz*, une compétition nationale centrée sur la visualisation et l'analyse des données. Pour cette édition, l'enjeu est d'améliorer les méthodes de prévision afin de mieux gérer les ressources énergétiques et de minimiser les impacts écologiques et financiers. En optimisant les prévisions pour une production électrique plus stable, on peut notamment limiter le recours à des solutions coûteuses et polluantes, comme l'utilisation de turbines thermiques.

*Directeur de recherche MINES Paris - PSL (HDR, Dr) Département Énergétique et Procédés, Centre Observation, Impacts, Energy (O.I.E.) MINES Paris - PSL Campus Pierre Laffitte 1 Rue Claude Daunesse - CS 10207 F-06904 Sophia Antipolis Cedex

3 Intérêt des Prévisions de Qualité

L'amélioration de la précision des prévisions énergétiques présente des avantages considérables :

- **Efficacité énergétique** : Des prévisions précises permettent d'optimiser la production en fonction de la demande anticipée, évitant ainsi la production excédentaire ou insuffisante.
- **Réduction des coûts** : Chaque pourcent d'amélioration en précision peut générer une économie d'environ 0.5% en coûts d'exploitation, notamment en réduisant le recours à des sources d'énergie plus coûteuses comme les turbines thermiques.
- **Impact écologique** : Une meilleure gestion de la production réduit le recours à des sources énergétiques polluantes, contribuant ainsi à diminuer les émissions de gaz à effet de serre et à favoriser les énergies renouvelables.

Pour illustrer l'impact de ces améliorations, une prévision 1% plus précise équivaut à environ 0.5% de réduction des coûts et à une diminution des émissions polluantes, puisque cela limite l'activation des turbines thermiques en période de forte demande.

4 Prétraitement des Données

Le prétraitement des données est une étape cruciale pour garantir la qualité des prévisions. Les données utilisées proviennent de plusieurs sources de production (thermique, hydraulique, solaire, etc.) ainsi que d'autres variables telles que l'heure et les importations d'électricité. Ces données sont organisées sous forme de matrice d'entrées X et de sorties Y , et un fenêtrage est appliqué pour structurer les observations temporelles.

4.1 Transformation des Heures

Pour capturer la périodicité quotidienne, l'heure est transformée en utilisant des fonctions trigonométriques sinus et cosinus, comme suit :

$$\begin{aligned}\text{Heures_sin} &= \sin\left(2\pi \frac{\text{Heures} + 1}{24}\right) \\ \text{Heures_cos} &= \cos\left(2\pi \frac{\text{Heures} + 1}{24}\right)\end{aligned}$$

Ces transformations permettent de représenter les heures dans un format cyclique, essentiel pour des modèles d'apprentissage automatique qui doivent prendre en compte les variations temporelles.

5 Création des Matrices d'Entrée et de Sortie

Les données sont organisées en matrices d'entrée X et de sortie Y . L'entrée X contient les observations sur une fenêtre de 48 heures pour chaque variable (production totale, solaire, thermique, etc.). La sortie Y , quant à elle, représente les valeurs à prédire pour chaque type de production, à des horizons de temps donnés. Les matrices sont préallouées pour une efficacité optimale lors de l'entraînement du modèle.

$$X \in R^{n \times (48 \cdot d)}, \quad Y \in R^{n \times (d-2)}$$

où n est le nombre d'observations et d est le nombre de variables.

6 Modèle ELM : Théorie et Entraînement

L'Extreme Learning Machine (ELM) est un modèle de réseau de neurones à une seule couche cachée, où les poids d'entrée et les biais sont générés aléatoirement, tandis que les poids de sortie sont calculés en résolvant un problème de régression linéaire. Ce modèle est particulièrement rapide à entraîner et bien adapté pour les grandes quantités de données.

6.1 Formulation Mathématique

Le modèle ELM est défini par l'équation suivante pour la sortie prédite \hat{Y} :

$$\hat{Y} = HW$$

où H est la matrice des sorties de la couche cachée et W représente les poids de sortie. La matrice H est obtenue via une fonction d'activation non-linéaire appliquée à l'entrée X , soit :

$$H = \phi(XW_{in} + b)$$

où W_{in} représente les poids d'entrée, b le biais, et ϕ est la fonction d'activation choisie (ReLU dans notre cas, définie par $\phi(x) = \max(0, x)$).

Les poids de sortie W sont ensuite déterminés en résolvant l'équation suivante :

$$W = (H^T H)^{-1} H^T Y$$

Cette approche garantit une solution optimale en termes de moindres carrés.

6.2 Algorithme d'Entraînement

L'entraînement consiste à :

- Générer aléatoirement les poids d'entrée W_{in} et les biais b .
- Calculer la matrice H en appliquant la fonction d'activation ReLU.
- Résoudre la régression linéaire pour obtenir les poids de sortie W .
- Évaluer la performance sur les données de test en utilisant les métriques d'erreur.

7 Évaluation des Performances

Les performances du modèle sont mesurées à l'aide de plusieurs métriques :

- **RMSE** (Root Mean Square Error), donnée par :

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

- **MAE** (Mean Absolute Error) :

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- **Coefficient de détermination R^2** , calculé comme :

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Ces métriques permettent de comparer la performance du modèle ELM avec celle du modèle de persistance, en particulier pour les horizons de temps prédéfinis.

8 Conclusion

Cette étude montre que le modèle ELM offre des gains de performance significatifs par rapport au modèle de persistance, surtout pour des horizons de prédiction courts. Ces résultats mettent en évidence le potentiel des modèles d'apprentissage automatique pour améliorer la gestion de la production énergétique, en particulier dans des contextes insulaires où l'équilibre entre offre et demande est plus fragile.