

Exploiter le Deep Learning pour prévoir les vitesses de vent : Une approche liant topographie et données de réanalyse météorologique

Diego Grante*, Rachel Baile*
Christophe Paoli*

*Università di Corsica Pasquale Paoli
{grante_d, baile_r, paoli_c}@univ-corse.fr,

Résumé. Dans le contexte de mix énergétique, et afin d’améliorer la stabilité du réseau électrique, il est primordial pour un gestionnaire de réseau de connaître, le plus précisément possible, les quantités futures d’énergie produite notamment éoliennes et photovoltaïques. Afin d’améliorer les prévisions de la ressource éolienne, notre étude se base sur un réseau de neurones hybride, intégrant des données de réanalyse météorologique (ERA5) ainsi que des données topographiques géographiquement proche du lieu de prévision. Nous nous intéressons particulièrement à l’impact de l’évolution de la taille de la zone topographique considérée sur les performances de prévision du réseau. Nous montrons notamment dans cette étude que l’intégration des données de relief permet d’améliorer les prévisions par rapport au même réseau non hybridé quel que soit l’horizon de prévision considéré avec un gain sur la RMSE allant jusqu’à 17%.

1 Introduction

Dans le contexte actuel de transition énergétique, la prévision de la production d’énergie renouvelable à des horizons de temps de 1h à 24h est un enjeu majeur pour les gestionnaires de réseau électrique. En particulier, la production éolienne, fortement dépendante des conditions météorologiques, est soumise à une grande variabilité. Une estimation des vitesses de vent est essentielle pour garantir la stabilité du réseau et optimiser la planification de la production et de la distribution d’énergie Pinson (2013). Dans cette étude, nous proposons un modèle de prévision de vitesse de vent intégrant à la fois des données météorologiques de réanalyse (ERA5) mises à disposition par Hersbach et al. (2023) et des données topographiques sous forme d’images colorimétriques (Topographic Map) représentant le relief à proximité du point considéré, afin de capturer les interactions complexes entre la topographie et le comportement des vents. Ce modèle est considéré comme un modèle hybride. De nombreux articles utilisent des réseaux hybrides pour prédire les vitesses de vent, tels que Lv et al. (2023), qui exploitent la décomposition par modes variationnels (VMD) et des gated recurrent units (GRU), ou encore Abbasipour et al. (2021), qui basent des réseaux de neurones sur des ondelettes (WNN). En termes d’utilisation de la topographie, différentes approches existent. Lawan et al.

(2020) utilisent des données géographiques ponctuelles de latitude, longitude, altitude et rugosité de terrain. Eikeland et al. (2022) relèvent que l'utilisation du deep learning peut avoir d'autant plus d'intérêt dans des régions à la topographie accidentée. Inspirés par Dujardin et Lehning (2022), qui utilisent l'apprentissage profond et une haute résolution topographique pour générer des champs de vent, nous avons choisi d'intégrer les données topographiques en utilisant des couches convolutives afin de permettre au mieux l'extraction de caractéristiques de plus haut niveau par le réseau de neurones. Notre approche concilie un modèle d'extraction de caractéristiques topographiques et un modèle d'extraction de caractéristiques de séries temporelles. Nous analysons l'impact des tailles d'images topographiques (70x70, 110x110, et 150x150 pixels) sur les performances de prévision du modèle. Nous montrons que l'ajustement de cette granularité n'apporte pas d'amélioration notable en termes d'erreur quadratique moyenne (RMSE), mais que la prise en compte globale de la topographie permet de réduire l'erreur jusqu'à 17 % selon les horizons de prévision considérés. Dans la suite de ce document (section 2), nous aborderons la provenance, les caractéristiques et l'historique des données. Dans la section 3, nous discuterons de l'architecture du modèle et de la méthodologie expérimentale mise en oeuvre. Dans la section 4, nous traiterons des résultats obtenus.

2 Données

Les données constituent le pilier central d'un système d'apprentissage profond performant. La qualité, la quantité, et la diversité des données influencent directement la capacité du modèle à apprendre et à généraliser efficacement. Deux types de données ont été exploités afin de réaliser cette étude.

Les premières données sont des données météorologiques. Elles proviennent du Centre européen pour les prévisions météorologiques à moyen terme (ECMWF) compilées et mises à disposition par Hersbach et al. (2023). Ce jeu de données ERA5 est une réanalyse du climat. Cette réanalyse combine les observations passées (stations météorologiques, satellites, radiosondes, etc ...) avec des modèles pour générer des séries chronologiques cohérentes de multiples variables climatiques. Elle fournit une description complète du climat observé tel qu'il a évolué au cours des dernières décennies, sur des grilles 3D à des intervalles sous-quotidiens. La résolution spatiale est de $0.25^\circ \times 0.25^\circ$. Nous avons choisi le jeu de données présentant un seul niveau de pression ainsi que trois points : l'un à proximité de Corte, le deuxième près de Bastia, et le dernier près d'Ajaccio, afin de couvrir le massif et les rives Est et Ouest de la Corse, comme le montre la figure 1. Le point situé à proximité de Corte présente un fort relief à l'ouest et un relief plus léger à l'est, positionnant les relevés ERA5 dans une vallée. Pour les points de Bastia et d'Ajaccio, nous avons à la fois un relief montagneux et maritime à proximité. Parmi le large spectre de caractéristiques proposé dans le jeu de données ERA5, nous avons sélectionné des variables en lien avec la prévision de vitesse de vent (cf tableau 1) à l'instar de Baile et Muzy (2022) et Geng et al. (2020). Notre extrait du jeu de données couvre la période du 1er janvier 2013 au 31 décembre 2023 avec un pas de temps d'une heure.

Le deuxième jeu de données est constitué des données topographiques autour du point sélectionné. Ces données sont obtenues à partir de cartes avec une échelle colorimétrique en fonction de la hauteur du relief grâce à Topographic Map. Le format de ces données est une image, c'est-à-dire une matrice avec 3 canaux (RVB). La figure 2 illustre les images fournies au réseau de neurones, avec des résolutions de 70x70, 110x110, et 150x150 pixels, correspondant

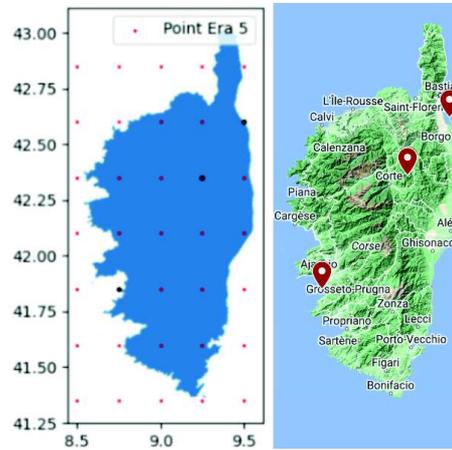


FIG. 1 – Localisation des points géographiques étudiés à partir du jeu de données ERA5

Nom	Unités	Description
100m u-component of wind	$m.s^{-1}$	Composante du vent à 100m vers l'est
100m v-component of wind	$m.s^{-1}$	Composante du vent à 100m vers le nord
10m u-component of wind	$m.s^{-1}$	Composante du vent à 10m vers l'est
10m v-component of wind	$m.s^{-1}$	Composante du vent à 10m vers le nord
2m dewpoint temperature	K	Température du point de rosée à 2m
2m temperature	K	Température à 2m
instantaneous 10m wind gust	$m.s^{-1}$	Rafale de vent à 10m
Mean sea level pressure	Pa	Pression au niveau de la mer
Surface pressure	Pa	Pression à la surface du point

TAB. 1 – Caractéristiques du jeu de données ERA 5 sélectionné

respectivement à des tailles de 24,5 km x 24,5 km, 38,5 km x 38,5 km, et 52,5 km x 52,5 km (échelle de 350 mètres par pixel).

3 Méthodologie

L'approche méthodologique adoptée dans cette étude repose sur une hybridation entre les connaissances géographiques contenues dans une image colorimétrique et les données de ré-analyse climatique, en les utilisant comme entrées d'un réseau de neurones hybride.

3.1 Description du modèle

Le réseau de neurones illustré dans la figure 3 accepte deux types de données en entrée, chacun traité par des sous-réseaux spécifiques. La première entrée concerne les données to-

Combiner topographie et météorologie pour prévoir les vitesses de vent

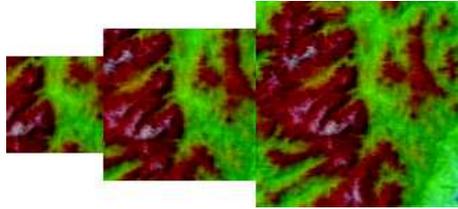


FIG. 2 – Exemple de données topographiques injectées dans le modèle. De gauche à droite, 70x70, 110x110 et 150x150 pixels (350 mètres par pixel)

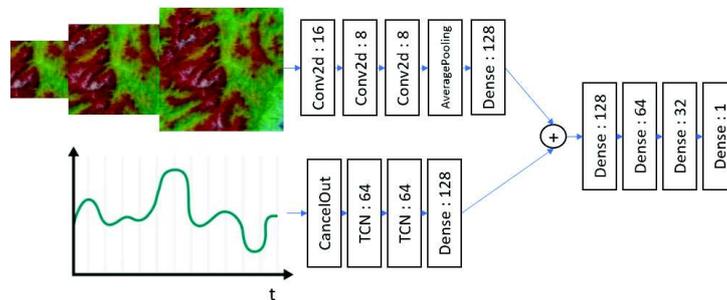


FIG. 3 – Architecture du modèle utilisé

pographiques, fournies sous forme d'images représentant le relief de la région. Ces images sont ensuite traitées par une série de réseaux convolutifs (Conv2d de la figure 3) destinés à en extraire des caractéristiques. Ces réseaux permettent d'obtenir des représentations de plus haut niveau, réduisant la grande dimension des données d'origine tout en conservant les informations essentielles pour la suite du traitement. La deuxième entrée correspond aux données temporelles, qui incluent l'ensemble des variables du tableau 1 sur une fenêtre temporelle de 18 pas de temps. Ce flux de données passe initialement par une couche de sélection automatique de caractéristiques proposée par Borisov et al. (2019), permettant d'attribuer un poids différent lors de l'entraînement à chaque caractéristique en fonction de sa pertinence pour la prévision finale. Cette pondération améliore la performance du modèle en accentuant l'influence des variables les plus déterminantes pour la prévision finale, tout en atténuant le risque de sur-apprentissage. De plus, elle confère une dimension explicative au réseau de neurones grâce à la mise en évidence des contributions respectives des différentes variables. Après la phase de sélection, les données temporelles traversent une série de couches successives, permettant de capturer les relations non linéaires entre les différentes variables atmosphériques. Ces couches sont des couches convolutives à dimension temporelle, appelées Temporal Convolution Networks (TCN, Yan et al. (2020)). Enfin, les vecteurs de caractéristiques issus des deux branches du modèle topographique et météorologique sont fusionnés en un seul vecteur, puis injectés dans une succession de couches entièrement connectées. Cette dernière phase a pour objectif d'apprendre les interactions complexes entre les facteurs topographiques et météorologiques,

et de fournir une prévision optimisée de la vitesse du vent. Les hyperparamètres, tels que le nombre de couches et le nombre d'unités par couche, ont été déterminés de manière empirique.

3.2 Expérimentation

Pour notre expérimentation, nous avons utilisé les données décrites dans le tableau 1 pour chacun des trois points indiqués sur la figure 1. Pour chaque image, nous utilisons la même série temporelle propre au point de données considéré et changeons l'horizon de prévision pour obtenir des résultats à 1h, 3h, 6h, 12h et 24h. Nous avons standardisé toutes les données en entrée du modèle. De plus, pour prendre en compte les variations saisonnières de la vitesse du vent, nous avons ajouté un indice temporel qui renseigne sur l'heure dans la journée et le jour dans l'année. La fonction ReLu est utilisée pour toutes les couches avec des fonctions d'activation, sauf celle de sortie qui a une fonction linéaire. Les deux premières couches de convolution (Conv2d) ont un *stride* égal à 2 et ont un filtre commun à toutes les couches convolutives de 3x3. La couche de pooling (AveragePooling) a un filtre de 2x2. Le jeu de données est divisé en trois parties : le jeu d'entraînement couvrant la période 2013-2021, le jeu de validation couvrant l'année 2022, et le jeu de test couvrant l'année 2023. Nous avons utilisé les composantes vectorielles du vent à 10 m : u_{10} étant la composante ouest-est et v_{10} étant la composante sud-nord, afin de créer une variable nommée amplitude (A), qui sera notre objectif de prévision.

$$A = \sqrt{u_{10}^2 + v_{10}^2}$$

Afin d'entraîner notre réseau de neurones, nous utilisons l'erreur quadratique moyenne (RMSE), qui traduit la distance euclidienne entre notre valeur prédite et notre valeur réelle.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (A_{i+h} - \hat{A}_{i+h})^2}$$

Avec n le nombre d'éléments dans la série temporelle du jeu de test, \hat{A}_{i+h} la prévision du modèle à l'instant i pour un horizon h .

Les séries temporelles utilisées en entrée de notre modèle comportent une fenêtre de 18 pas de temps telle qu'on considère les données de l'instant $i - 17$ à i . L'expérimentation est configurée avec 50 epochs, avec mécanisme d'early stopping. Le meilleur modèle est sélectionné sur la base de la plus faible RMSE obtenue pendant la phase de validation. Chaque expérimentation est indépendante, il y a un entraînement pour chaque site afin de ne pas avoir de biais de reconnaissance de site. Les résultats de notre expérience se basent sur le rapport entre la RMSE de d'un modèle dont on teste les performances (RMSE_t) et la RMSE obtenue par un modèle dit de référence (RMSE_r). Nous exprimons cette différence sous forme de gain (skill score) :

$$G = \left(1 - \frac{\text{RMSE}_t}{\text{RMSE}_r}\right) \times 100.$$

Combiner topographie et météorologie pour prévoir les vitesses de vent

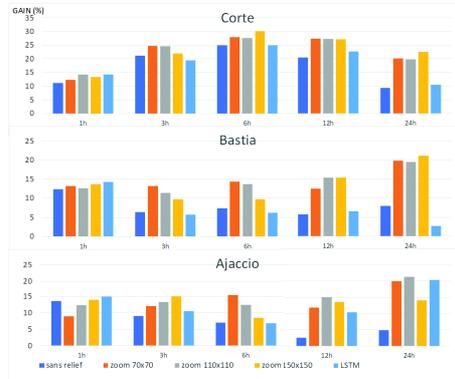


FIG. 4 – Valeurs du gain par rapport à la persistance pour trois niveaux de zoom topographique, sans les données topographiques et un LSTM simple avec cinq horizons de prévision, sur les sites de Corte, Bastia et Ajaccio.

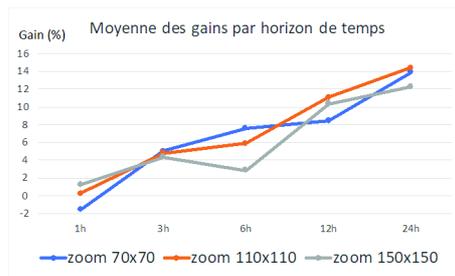


FIG. 5 – Évolution du gain (%) avec l'horizon de prévision (1h à 24h), calculé sur les RMSE moyennes d'Ajaccio, Bastia et Corte. Comparaison entre modèles sans données topographiques et avec images topographiques de différentes tailles (70x70, 110x110, 150x150 pixels).

4 Résultats

Dans la figure 4, nous avons comparé les RMSE obtenues à chaque horizon de prévision pour les quatre configurations du modèle : sans intégration d'image de données topographiques, avec une image de 70x70 pixels, une image de 110x110 pixels, et une image de 150x150 pixels, toujours centrées sur le point ERA5 étudié. Nous avons également inclus les résultats d'un LSTM. Le gain représenté pour valeur de référence la RMSE de la persistance. Nous observons que l'intégration des données topographiques améliore les performances du modèle, surtout pour les horizons de prévision plus longs. En effet, à partir d'un horizon de prévision de 3 heures, l'intégration des données topographiques apporte une amélioration significative des performances de prévision.

Pour mettre en valeur l'apport de la prise en compte des données topographiques, nous avons illustré dans la figure 5, le gain en pourcentage entre la RMSE obtenue en moyenne sur

Ajaccio, Bastia et Corte par notre modèle lorsque celui-ci intègre des données topographiques, et la RMSE obtenue en moyenne par notre modèle sans données topographiques intégrées. Nous nous sommes notamment intéressés à l'influence de la taille de l'image topographique sur la qualité de la prévision.

Tout d'abord, à l'horizon de 1 heure, le gain est quasiment nul, voire légèrement négatif, ce qui peut s'expliquer par une optimisation insuffisante des hyperparamètres pour cet horizon de temps ou par la simplicité relative de la prévision à court terme rendant plus difficile l'amélioration. Comme dans la figure 4, les résultats montrent également qu'il n'y a pas de différence de performance notable en fonction de la taille de l'image utilisée. Cependant, la différence de performance avec la configuration du modèle sans données topographiques est d'autant plus importante que l'horizon de prévision est élevé. Cela confirme que l'intégration des données topographiques, indépendamment de la taille de l'image, améliore les prévisions de vitesse de vent pour des horizons de prévision plus longs. Cette amélioration atteint un gain de 17% à Ajaccio à l'horizon de 24 heures.

5 Conclusion

Cette étude met en lumière l'intérêt de considérer les données topographiques dans le cadre de la prévision des vitesses de vent. L'intégration des données de relief, sous forme d'images colorimétriques, a permis une baisse de l'erreur quadratique moyenne obtenue pour des horizons de prévision allant de 3h à 24h, jusqu'à 17 pour cent pour l'un des points étudiés. La topographie locale est donc un facteur intéressant pour améliorer la qualité des prévisions de la vitesse du vent. Il sera intéressant de confirmer ce résultat sur des données de station météorologique mais aussi de comprendre pourquoi la modification de la taille de l'image n'a pas permis d'influencer de façon robuste les performances du modèle : la granularité de ces données pourra être interrogée et des données topographiques de nature différente pourront être testées. Enfin, des tests empiriques sont envisagés pour déterminer la méthode optimale de fusion des modèles. Nos travaux sont disponibles sur la plateforme GitHub à cette adresse Grante (2024).

Remerciements

Ce travail a été mené dans le cadre de l'ANR SAPHIR ANR-21-CE04-0014-03.

Références

- Abbasipour, M., M. A. Igder, et X. Liang (2021). A Novel Hybrid Neural Network-Based Day-Ahead Wind Speed Forecasting Technique. *IEEE Access* 9, 151142–151154, doi: 10.1109/ACCESS.2021.3126747.
- Baile, R. et J. F. Muzy (2022). Leveraging data from nearby stations to improve short-term wind speed forecasts. *Elsevier*.

- Borisov, V., J. Haug, et G. Kasneci (2019). Cancelout : A layer for feature selection in deep neural networks. In *Artificial Neural Networks and Machine Learning – ICANN 2019 : Munich*. Springer-Verlag, doi: 10.1007/978-3-030-30484-3_6.
- Dujardin, J. et M. Lehning (2022). Wind-topo : Downscaling near-surface wind fields to high-resolution topography in highly complex terrain with deep learning. *Quarterly Journal of the Royal Meteorological Society* 148, 1368–1388, doi: 10.1002/qj.4265.
- Eikeland, O., F. Hovem, T. Olsen, M. Chiesa, et F. Bianchi (2022). Probabilistic forecasts of wind power generation in regions with complex topography using deep learning methods : An arctic case. *Energy Conversion and Management : X* 15, 100239.
- Geng, D., H. Zhang, et H. Wu (2020). Short-term wind speed prediction based on principal component analysis and lstm. *Applied Sciences* 10(13), 4416, doi: 10.3390/app10134416.
- Grante (2024). Egc2025. https://github.com/Itexe2B/egc_2025.
- Hersbach, H., B. Bell, P. Berrisford, G. Biavati, A. Horányi, J. Muñoz Sabater, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee, et J.-N. Thépaut (2023). Era5 hourly data on pressure levels from 1940 to present.
- Lawan, S., W. Abidin, et T. Masri (2020). Implementation of a topographic artificial neural network wind speed prediction model for assessing onshore wind power potential in sibu, sarawak. *The Egyptian Journal of Remote Sensing and Space Science* 23(1), 21–34.
- Lv, S., L. Wang, et S. Wang (2023). A hybrid neural network model for short-term wind speed forecasting. *Energies* 16(4), 1841, doi: 10.3390/en16041841.
- Pinson, P. (2013). Wind Energy : Forecasting Challenges for Its Operational Management. *Statistical Science* 28(4), doi: 10.1214/13-STS445.
- Topographic Map. Topographic map - carte topographique du monde. <https://fr-fr.topographic-map.com/>. Consulté le 09 octobre 2024.
- Yan, J., L. Mu, L. Wang, R. Ranjan, et A. Y. Zomaya (2020). Temporal Convolutional Networks for the Advance Prediction of ENSO. *Scientific Reports* 10(1), 8055, doi: 10.1038/s41598-020-65070-5.

Summary

In the context of energy mix, and in order to improve the stability of the power grid, it is essential for a grid operator to know, as precisely as possible, the future quantities of energy produced, particularly from wind and solar sources. To improve wind energy resource forecasts, our study is based on a hybrid neural network, integrating meteorological reanalysis data (ERA5) as well as topographical data geographically close to the prediction site. We are particularly interested in the impact of the size of the considered topographical area on the network's forecasting performance. In this study, we specifically show that the integration of relief data improves predictions compared to the same non-hybrid network, regardless of the forecast horizon considered, with a gain in RMSE of up to 17%.