

RAPPORT DE STAGE

Prévision Énergétique Multi-Sources avec Réconciliation Optimale via une Approche MIMO-MH et Extreme Learning Machine (ELM)

Yoan JHEELAN

*Master 1 Energie - Matériaux, Université
Paris-Saclay*
yoan.jheelan@etu-upsaclay.fr

Prof. Elena MAGLIARO

Chercheur à Mines Paris – PSL (OIE)
elena.magliaro@minesparis.psl.eu

Prof. Cyril VOYANT

*Directeur de recherche MINES Paris -
PSL (OIE)*
cyril.voyant@minesparis.psl.eu

Prof. Andrea MICHIORRI

*Chercheur à Mines Paris – PSL
(PERSEE)*
andrea.michiorri@minesparis.psl.eu

Remerciement

Tout d'abord, je tiens à exprimer ma plus profonde gratitude envers mes professeurs et superviseurs du Centre OIE / PERSEE de Mines Paris - PSL, en particulier à Cyril VOYANT, Elena MAGLIARO et Andrea MICHIORRI. Votre accompagnement et votre soutien indéfectible durant ce stage ont été déterminants pour mon épanouissement académique, professionnel et personnel. Vos retours exigeants et votre implication ont guidé chaque étape de ce stage. Cette immersion au sein de votre équipe m'a permis d'appréhender la complexité de la prévision de la production d'énergie : en partant de la modélisation des modèles MIMO (Multi-Input Multi-Output) avec ELM (Extreme Learning Machine) puis MIMO MH (Multi-Horizon) à la réconciliation optimale des prévisions, en passant par l'analyse des données et une revue bibliographique. Vos conseils avisés sur mon projet professionnel ont été tout aussi précieux. Pour tout cela, et bien plus encore, merci infiniment. J'espère sincèrement que cette collaboration n'est que la première d'une longue série.

Un immense merci à toute l'équipe de l'OIE : à Benoît GSCHWIND, pour sa disponibilité et ses judicieux conseils ; à Rodrigo AMARO E SILVA, pour ses nombreux conseils et recommandations ; à Sandra HASSAN ; à Raphaël JOLIVET ; à Yves-Marie SAINT-DRENAN. Merci également aux doctorants : à Candice BANES ; à Amar MEDDAHI ; à Vadim BECQUET ; à Susanne WEYAND ; à Alan JULIEN ; et aux post-doctorants : Gabriel CHESNOIU ; Alejandra CUE GONZALEZ ; à Andriamahefason RAJAONISON ; à Mathilde MARCHAND-LASSERRE ; à Alessia BOCCALATTE, pour leurs conseils, éclairages méthodologiques, leur patience et leur convivialité, ainsi que pour les randonnées, qui a rendu ce séjour si enrichissant.

Je tiens à remercier Augustin TOURON et Simon POUREAU pour leurs précieuses informations sur la question de ce qu'est un bon outil de prévision production-consommation multi-énergie selon les acteurs de la prévisions multi-énergie en France.

Enfin, cette expérience n'aurait pu aboutir sans le directeur du laboratoire, le professeur Thierry RANCHIN et le titulaire de la Chaire SciDoSol, le professeur Philippe BLANC, ainsi qu'une fois encore le professeur Cyril VOYANT. Je vous remercie de m'avoir accueilli au sein de votre laboratoire en tant que stagiaire et permis de contribuer au développement d'un modèle pour la prévision de la production d'énergie adaptée aux contraintes des gestionnaires de réseau. Je suis honoré d'avoir participé à cette aventure au sein des laboratoires OIE / PERSEE, où l'innovation sert une industrie plus agile, responsable et efficace. Pour tout cela, et bien plus encore, merci.

Résumé/Abstract

Ce rapport constitue une première synthèse d'étape, à mi-parcours d'une étude plus large qui sera finalisée fin août 2025. Le stage a été prolongé par un contrat CDD de deux mois financé dans le cadre de la chaire SciDoSol (SCIences des DOnnées appliquées à l'énergie SOLaire), soutenue par TotalEnergies, Somfy, RTE, TSE et CalibSun. L'objectif global du projet est de développer un outil de prévision énergétique multi-sources et multi-horizons combinant MIMO-MH, Extreme Learning Machine (ELM) et réconciliation optimale, en réponse aux besoins opérationnels exprimés par des acteurs comme EDF et RTE. Les résultats obtenus à ce stade montrent une précision élevée et des gains significatifs par rapport aux méthodes de référence, avec validation croisée entre implémentations Matlab et Python. Les données issues du système électrique insulaire corse (2016–2022) constituent un jeu de données réaliste et complexe, bien adapté à l'expérimentation de modèles avancés. L'étude s'inscrit pleinement dans les enjeux de transition énergétique, prévision fiable et open science, avec une attention particulière portée à la reproductibilité et à la robustesse méthodologique. Les prochaines étapes porteront sur l'intégration complète de la réconciliation hiérarchique optimale et la mise à disposition open source de l'outil pour les membres de la chaire et, plus largement, la communauté scientifique. Une valorisation scientifique et technique est prévue fin 2025, et un nouveau sujet de stage est d'ores et déjà prévu pour 2026, afin d'approfondir les perspectives ouvertes par cette première phase.

This report provides a preliminary synthesis, representing the mid-point of a broader study that will be completed by the end of August 2025. The internship has been extended through a two-month fixed-term contract (CDD) funded by the SciDoSol Chair (Data Science Applied to Solar Energy), supported by TotalEnergies, Somfy, RTE, TSE, and CalibSun. The main objective of the project is to develop a multi-source, multi-horizon energy forecasting tool that combines MIMO-MH, Extreme Learning Machine (ELM), and optimal reconciliation, addressing operational needs expressed by key stakeholders such as EDF and RTE. Results obtained so far demonstrate high accuracy and significant improvements over baseline methods, with consistent performance validated across both Matlab and Python implementations. The dataset, based on the Corsican island power system (2016–2022), offers a realistic and complex context well suited to testing advanced modeling approaches. The study fully aligns with the challenges of energy transition, reliable forecasting, and open science, with particular focus on methodological robustness and reproducibility. The next steps will involve the full integration of hierarchical optimal reconciliation and the open-source release of the tool for use by SciDoSol partners and the broader scientific community. Scientific and technical dissemination is planned for late 2025, and a follow-up internship is already scheduled for 2026 to further extend the work initiated in this phase.

Table des matières

Remerciement	1
Résumé	2
Table des figures	4
Liste des tableaux	4
1 Introduction	6
1.1 Contexte	6
1.2 Contexte Institutionnel du Stage	6
1.2.1 Centre Observation, Impacts, Énergie (OIE)	6
1.2.2 Centre PERSEE (Procédés, Energies Renouvelables et Systèmes Energétiques)	7
1.2.3 Chaire SciDoSol (SCIences des DOnnées appliquées à l'énergie SOLaire)	7
1.3 Un Bon Outil de Prévision Selon les Acteurs Nationaux	7
1.3.1 Point de Vue du Réseau de Transport d'Électricité (RTE)	7
1.3.2 Point de Vue de EDF	8
1.3.3 Cohérence et Impact de la Thématique Explorée Durant ce Stage	8
2 Objectifs et Timeline	8
3 Méthodes et Théories	9
3.1 Extreme Learning Machine (ELM)	10
3.2 Multi Input Multi Output (MIMO)	11
3.3 Extension MIMO-MH : Multi-Horizon	12
3.4 Réconciliation des Prévisions Hiérarchiques	13
3.5 Méthodes Naïves et de Référence	14
3.6 Métriques d'Evaluation	14
4 Données Utilisées	15
4.1 Nature des données	15
4.2 Contexte énergétique corse	15
4.3 Analyse des données	16
4.4 Statistiques Descriptives	16
4.5 Saisonnalités et Tendances	18
4.6 Dépendances Statistiques et variables explicatives	20
4.6.1 Principaux résultats	20
4.6.2 Analyse de significativité	21
4.7 Résumé et perspectives	21
5 Résultats et Discussion	22
5.1 Corrélation entre variables	22
5.2 Résultat à court terme (horizon 1 h)	23
5.3 Réconciliation	26
5.4 Résultats	26
5.4.1 nRMSE	28
5.4.2 nMAE	30
5.4.3 nMBE	32
5.4.4 R^2	34
5.5 Résumé des résultats	35
5.5.1 Comparaison par source	35
5.5.2 Comparaison des modèles	36
5.6 Comparaison avec SISO et LSTM	36
5.7 Temps d'exécution	37
6 Perspectives	38
7 Conclusion	38

A Moindre carré	40
A.1 Pseudo inverse	40
A.1.1 Définition de la pseudo-inverse	40
A.1.2 Construction via décomposition en valeurs singulières (SVD)	40
A.1.3 Vérification des conditions de Moore-Penrose	40
B Résultats des huits moyens de production	41
B.1 Bioénergies	41
B.2 Eolienne	42
B.3 Hydraulique	43
B.4 Importations	44
B.5 Micro-Hydraulique	45
B.6 Solaire	46
B.7 Thermique	47
B.8 Total	48

Table des figures

1 Structuration de l'activité du centre OIE	7
2 Diagramme de Gantt	9
3 Schéma de la structure du réseau de ELM	10
4 Schéma de la structure du modèle MIMO	11
5 Schéma des modèles SISO, MIMO, MIMO-MH	13
6 Répartition de la production énergétique horaire moyenne globale par source d'énergie	17
7 Production énergétique horaire moyenne par source d'énergie (2016-2022)	18
8 Profils journaliers moyens de production d'énergie par source en Corse (2016-2022)	19
9 Violin plot de la production totale (MWh) par saison de 2016 à 2022	20
10 Heatmap de la corrélation de Spearman des sources d'énergies	21
11 Correlation de spearman	22
12 pacf _{solar, photovoltaic}	23
13 Prévision énergétique multi-sources via une approche MIMO sur python (à gauche) et Matlab (à droite)	24
14 Prévisions des variables énergétiques en Corse pour l'horizon 1h sur Python via MIMO et comparaison avec la persistance horizon et la persistance quotidienne	25
15 Amélioration apportée par la réconciliation	26
16 Benchmarks des métriques normalisées (moyenne des variables énergétiques)	27
17 Benchmarks nRMSE	28
18 Benchmarks par variable (nRMSE)	29
19 Benchmarks nMAE	30
20 Benchmarks par variable (nMAE)	31
21 Benchmarks nMBE	32
22 Benchmarks par variable (nMBE)	33
23 Benchmarks R^2	34
24 Benchmarks par variable (r2)	35
25 Temps d'exécution de chaque models	37

Liste des tableaux

1 Métriques d'Évaluation	15
2 Résumé global de la production horaire moyenne (2016-2022)	17
3 Résumé par année de la production horaire moyenne	17
4 Résumé par saison de la production horaire moyenne	19
5 Comparaison des performances du modèle MIMO entre Matlab / Python à l'horizon 1h	24
6 Comparaison synthétique des performances de trois approches pour $h + 5$ (nRMSE)	37
7 Comparaison des performances pour la production de bioénergies (nRMSE)	41
8 Comparaison des performances la production éolienne (nRMSE)	42
9 Comparaison des performances pour la production hydraulique (nRMSE)	43
10 Comparaison des performances pour les importations (nRMSE)	44

11	Comparaison des performances pour la production micro-hydraulique (nRMSE)	45
12	Comparaison des performances pour la production solaire photovoltaïque (nRMSE)	46
13	Comparaison des performances pour la production thermique (nRMSE)	47
14	Comparaison des performances pour la production totale (nRMSE)	48

1 Introduction

Cette section introduit le cadre général de l'étude, les motivations et les enjeux liés à la prévision énergétique multi-sources en contexte insulaire.

1.1 Contexte

L'utilisation des énergies renouvelables est essentielle pour sortir de la dépendance des énergies fossiles et augmenter la part d'énergie décarboné du mix énergétique français. Ce qui permettrait de réduire les émissions de gaz à effet de serre et suivre les scénario de projections climatiques les moins pessimistes du Groupe Intergouvernemental d'Experts sur l'Evolution du Climat (GIEC) ou encore suivre les scénarios de mix de production permettant d'atteindre la neutralité carbone à l'horizon 2050 de RTE [1]. Cependant, la nature intermittente des énergies renouvelables pose plusieurs défis, notamment en matière de fiabilité, de qualité de l'énergie et d'équilibre entre l'offre et la demande[2]. Dans ce contexte, la prévision de la production d'électricité issue de sources renouvelables, telles que l'énergie éolienne et solaire, devient essentielle pour le fonctionnement efficace et continu du réseau électrique [3]. L'approche Multi-Input Multi-Output Multi-Horizon (MIMO-MH) avec réconciliation et un Extrême Learning Machine (ELM) permet de synchroniser les prévisions des différentes sources (solaire, thermique, hydraulique, imports, etc.) pour garantir la cohérence avec la consommation nette (équivalente à la demande du réseau), tout en capturant les interactions et les contraintes physiques globales (équilibre offre-demande, import/export, autoconsommation) [4]. Contrairement aux modèles Single-Input Single-Output (SISO) qui traitent chaque source isolément, MIMO tire parti des corrélations entre sources et de la variabilité partagée, ce qui améliore la précision agrégée et réduit les écarts totaux (via l'effet de compensation entre erreurs). En parallèle, ELM apporte un apprentissage rapide, une solution fermée analytique et une faible charge computationnelle, le rendant idéal pour l'adaptation en temps quasi réel. Cette approche optimise la prévision de la demande finale (consommation nette), indispensable pour le dispatch, la gestion des importations et la stabilité du réseau, tout en offrant une solution robuste et peu coûteuse pour des systèmes multi-énergies fortement variables et sujets à l'autoconsommation.

Ce rapport s'articule autour de cinq grandes sections : après la présentation des objectifs et de la planification du stage, les méthodes et approches théoriques (MIMO, ELM, MIMO-MH, réconciliation) seront détaillées. Ensuite, les données utilisées seront analysées pour en extraire les caractéristiques essentielles. Les résultats obtenus à partir des modèles seront ensuite présentés et comparés, avant de conclure par une synthèse des principaux enseignements et les perspectives futures ouvertes par ce travail.

1.2 Contexte Institutionnel du Stage

Pour mieux situer le contexte de ce stage au sein d'un écosystème de recherche et d'innovation, le paragraphe suivant présente brièvement les principales structures impliquées. Le centre Observation, Impacts, Énergie (OIE) et PERSEE (Procédés, Energies Renouvelables et Systèmes Energétiques), situés à Sophia-Antipolis, appartiennent au Département Énergétiques procédés - MINES PARIS - PSL¹. Les recherches menées au Département Énergétique et Procédés visent à proposer des solutions fiables, compétitives, respectueuses de l'environnement, et adaptées à une demande accrue de biens et de services.

1.2.1 Centre Observation, Impacts, Énergie (OIE)

Le centre Observation, Impacts, Énergie (OIE)² Mines Paris – PSL est un laboratoire de référence dans l'évaluation et la modélisation des ressources énergétiques renouvelables, des impacts environnementaux et des usages de l'énergie. Son approche originale repose sur l'analyse conjointe des dimensions temporelles et spatiales, grâce à l'exploitation de données in-situ, satellitaires et issues de modèles numériques de prévision météorologique et environnementale. À l'intersection des sciences de la donnée, de l'observation de la Terre et de la modélisation prospective (Figure 1), le centre développe des outils méthodologiques avancés, tels que les modèles Heliosat et les bibliothèques d'Analyse du Cycle de Vie (ACV en open source). Ses travaux sont valorisés par des partenariats industriels, académiques et institutionnels à l'échelle nationale et internationale. Fortement impliqué dans la formation, la production scientifique et la diffusion des connaissances, OIE contribue activement à la transition énergétique et à l'élaboration

1. <https://www.minesparis.psl.eu/recherche/17-centres-de-recherche-5-domaines-disciplinaires/energetiques-et-procedes/>

2. <http://www.oie.minesparis.psl.eu/>

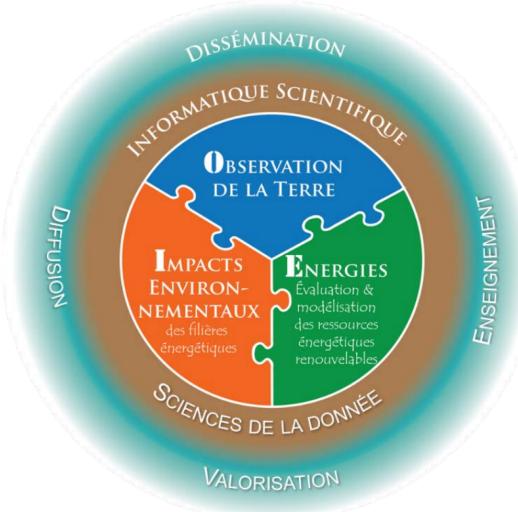


FIGURE 1 – Structuration de l’activité du centre OIE

de politiques publiques durables, tout en occupant un rôle central dans des initiatives européennes et mondiales liées à l’énergie, au climat et à la durabilité.

1.2.2 Centre PERSEE (Procédés, Energies Renouvelables et Systèmes Energétiques)

Le laboratoire PERSEE³ de Mines Paris – PSL, regroupe environ 60 chercheurs et ingénieurs. Il est structuré en trois axes de recherche (MATPRO, TeP et ERSEI) et couvre l’ensemble de la chaîne de l’énergie : de la recherche sur les matériaux et procédés (hydrogène, aérogels, plasma, stockage) à l’intégration des énergies renouvelables et des réseaux intelligents (prévision EnR, optimisation de microgrids, smart grids). Le laboratoire s’appuie sur une forte collaboration industrielle pour accompagner la transition énergétique et développer des solutions innovantes pour des systèmes énergétiques plus durables et résilients.

1.2.3 Chaire SciDoSol (SCIences des DOnnées appliquées à l’énergie SOLaire)

La chaire SciDoSol⁴, dédiée aux sciences des données appliquées au solaire, est soutenue par cinq mécènes : TotalEnergies⁵, Somfy⁶, RTE (Réseau de Transport d’Électricité)⁷, TSE (acteur du photovoltaïque et agrivoltaïque⁸) et CalibSun (solutions d’optimisation pour centrales photovoltaïques⁹). Ces partenaires accompagnent la recherche sur la caractérisation, la prévision et l’exploitation de la ressource solaire, avec des données issues de capteurs in-situ, satellites et modèles, dans l’objectif de renforcer l’intégration des EnR et d’accompagner la transition énergétique.

1.3 Un Bon Outil de Prévision Selon les Acteurs Nationaux

Afin de situer le cadre général de l’outil de prévision, nous allons examiner le point de vue détaillé de RTE et d’EDF, en matière de précision, de cohérence et de robustesse.

1.3.1 Point de Vue du Réseau de Transport d’Électricité (RTE)

Selon Augustin Touron, chargé d’Études R&D au Pôle Prévision Court-Terme et Équilibre Offre & Demande, l’outil de prévision idéal combine haute précision déterministe surtout dans les situations critiques, et mesures probabilistes pour évaluer les marges en intégrant les incertitudes. Il doit pouvoir produire des prévisions à différentes échelles, depuis la maille poste ou transformateur jusqu’à la maille nationale,

3. <https://www.persee.minesparis.psl.eu/>

4. <https://www.scidosol.fr/>

5. <https://www.totalenergies.fr/>

6. <https://www.somfy.fr/>

7. <https://www.rte-france.com/>

8. <https://www.tse.energy/>

9. <https://www.calibsun.com/fr>

avec une résolution temporelle d'au moins 15 minutes, voire plus fine pour capter des phénomènes comme les rafales. Il est également nécessaire de couvrir des horizons infra-journaliers, à J+1, et éventuellement hebdomadaires. La réconciliation inter-sources, l'intégration de facteurs non météo (pannes, écrêtements), la vision en quasi temps réel (notamment pour le photovoltaïque), et la capacité à agréger les prévisions de RTE, Gestionnaire de Réseau de Distribution (GRD), producteurs et autres acteurs sont essentiels. La rapidité de calcul, la montée en charge à des milliers d'installations, l'interprétabilité des résultats et un système structuré de monitoring et retour d'expérience (REX) pour détecter les dérives de performance sont également requis. Aujourd'hui, RTE s'appuie sur des modèles séparés pour la consommation, l'éolien et le photovoltaïque, auxquels s'ajoute un module sur les pertes. Les prévisions éoliennes et PV suivent une approche bottom-up : régressions variables météo au niveau des parcs, additionnées puis réconciliées, avec une résolution horaire bientôt portée à 15 minutes, et une correction à court terme pour l'éolien à partir des télémesures. L'horizon de prévision s'étend jusqu'à 72 h, avec production de quantiles à différentes mailles. La consommation est prévue via un modèle national jusqu'à J+2, renforcé par des modèles régionaux top-down pour les mailles postes. Des modèles directs nationaux pour l'éolien et le PV sont en préparation, ce qui pose des défis de réconciliation et de fusion entre approches. Augustin Touron souligne plusieurs priorités d'évolution : développer des modèles nationaux directs pour l'éolien et le photovoltaïque, intégrer des approches spatio-temporelles pour le très court terme, combiner des prévisions issues de différents fournisseurs météo, refondre le modèle éolien, renforcer les prévisions de consommation nette locale et généraliser l'outil de retour d'expérience.

1.3.2 Point de Vue de EDF

Selon Simon Poureau, Ingénieur Contrôle Commande - Aide à la prévision de production et référent données chez EDF Renouvelables, un bon outil de prévision multi-énergie serait un système capable d'anticiper précisément le futur, ce qui souligne le défi fondamental du domaine de la prévisions de la production d'électricité. Pour ce qui est des méthodes actuellement utilisées chez EDF, Simon Poureau n'a connaissance que de l'expertise qui se concentre principalement sur l'hydroélectricité où les approches restent disparates. Certains aménagements s'appuient sur des prévisions météorologiques pour anticiper les crues, notamment via le suivi des orages, mais ces méthodes ne sont pas très fiables. En pratique, cela nécessite une surveillance humaine complétée par des systèmes d'alarme capables de fournir des informations précises sur les conditions réelles du terrain. Par ailleurs, dans le domaine de l'équilibrage réseau (expérience qu'il a vécue en travaillant chez un agrégateur), le pilotage s'effectue principalement via la fréquence de RTE qui envoie des signaux spécifiques. Sur les améliorations souhaitables, bien qu'il ne connaisse pas spécifiquement les outils de RTE, plusieurs axes émergent. L'intégration massive de données apparaît comme prioritaire : croiser les paramètres météorologiques (température, ensoleillement, risque orageux), l'état des infrastructures (niveaux de barrage, tension réseau, surchauffes d'interconnexions) et les indicateurs de réseau permettrait d'anticiper des scénarios critiques comme les coupures. IA et ELM sont identifiés comme des leviers intéressants, c'est une piste déjà explorée en interne chez EDF via des datalakes produisant des tableaux de bord de pilotage.

1.3.3 Cohérence et Impact de la Thématique Explorée Durant ce Stage

Ce stage répond directement aux enjeux nationaux de la prévision énergétique. En combinant les approches MIMO, multi-horizon et ELM avec une réconciliation optimale, il contribue à renforcer la cohérence inter-sources, l'intégration de facteurs non météorologiques, la précision sur plusieurs horizons et la robustesse des outils de prévision. La mise en œuvre Python et la comparaison avec des benchmarks existants (tels que Matlab) visent à fournir une méthodologie transparente, reproductible et opérationnelle. Ce projet offre ainsi un impact tangible pour les acteurs comme RTE et EDF, en proposant une solution complète et modulaire, capable d'évoluer vers des systèmes massifs et intégrant les besoins exprimés par terrain.

2 Objectifs et Timeline

Les objectifs principaux de ce stage sont :

1. Réaliser une revue bibliographique des approches MIMO (Multi-Input Multi-Output), MIMO-MH (Multi-Input Multi-Output Multi-Horizon), SISO (Single-Input Single-Output), des algorithmes ELM (Extreme Learning Machine), ainsi que des techniques de réconciliation optimale des prévisions ;

2. Étudier et comparer les approches MIMO et SISO, en mettant en avant l'intérêt du MIMO pour la cohérence inter-sources, la réduction des erreurs globales et la meilleure adaptation aux interactions physiques entre les différentes sources d'énergie ;
3. Explorer l'intérêt d'une approche multi-horizon (MH), qui permet de prévoir simultanément à différents horizons temporels l'évolution de la production des différentes sources d'énergie, un aspect crucial pour les applications réelles telles que la gestion de réseaux et le dispatch ;
4. Développer en Python un modèle complet de prévision basé sur le couple MIMO-MH et l'ELM, combinant une approche théorique (formulation mathématique des modèles, méthodes de réconciliation) et une mise en œuvre pratique (implémentation efficace, optimisation des performances et reproductibilité) ;
5. Réaliser des simulations sur des données réelles pour valider la faisabilité et la pertinence des modèles proposés, en utilisant des métriques standards (RMSE, MAE, R^2 , etc.) afin de comparer les performances ;
6. Comparer les résultats obtenus avec le benchmark Matlab existant afin de valider l'implémentation Python, d'évaluer ses atouts (vitesse, précision, flexibilité) et de démontrer les avantages de cette solution ;
7. Proposer une méthodologie reproductible et open source, en fournissant les scripts et données nécessaires pour permettre à la communauté scientifique et industrielle d'adopter ces outils dans un contexte opérationnel.

Pour atteindre les objectifs fixés, les étapes suivantes ont été suivies pendant les 5 mois du stage :

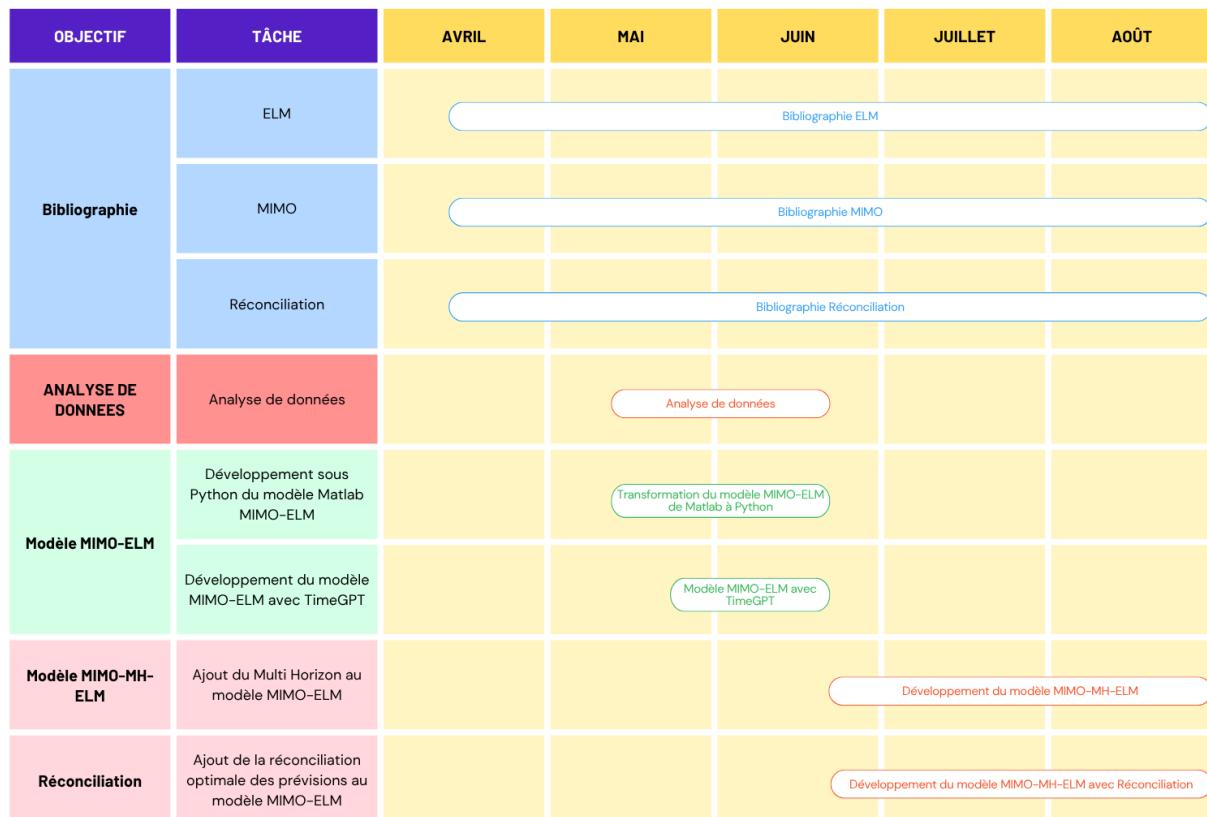


FIGURE 2 – Diagramme de Gantt

3 Méthodes et Théories

Avant de détailler les modèles et méthodes mis en œuvre dans ce travail, il est indispensable de dresser un état de l'art des principaux outils et approches mobilisés. Nous allons ainsi aborder les fondements théoriques et les applications des différentes techniques, en commençant par le modèle Extreme Learning Machine (ELM) qui joue un rôle central dans cette étude.

3.1 Extreme Learning Machine (ELM)

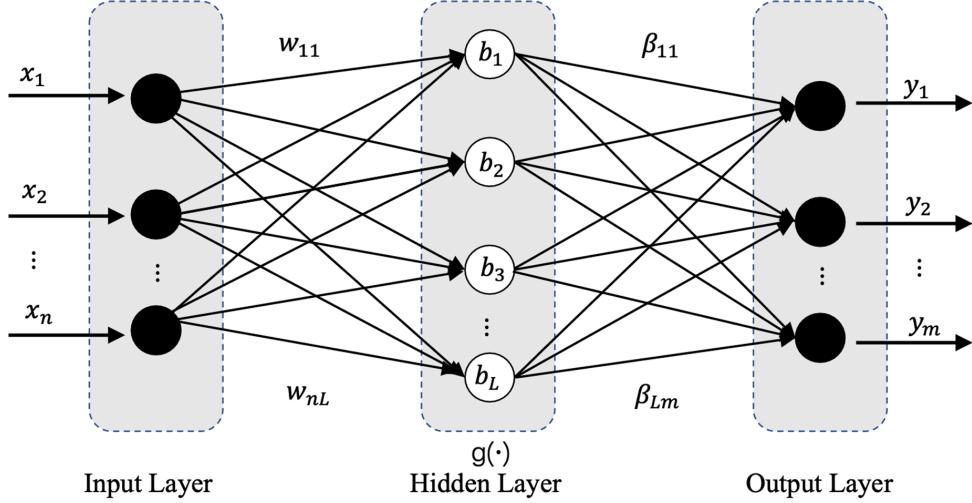


FIGURE 3 – Schéma de la structure du réseau de ELM

L'ELM est un réseau de neurones à une seule couche cachée, (voir Figure 3, [5, 6]), dont la spécificité réside dans le choix aléatoire et fixe des poids d'entrée. Seuls les poids de sortie sont déterminés par un calcul analytique explicite (closed-form en Anglais), ce qui permet un entraînement rapide [6–9].

On considère un ensemble de données d'apprentissage $D = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$, avec $\mathbf{x}_i \in \mathbb{R}^n$ et $\mathbf{y}_i \in \mathbb{R}^m$ (respectivement les entrées et les sorties).

La couche cachée comprend L neurones. Chaque neurone caché ℓ est défini par :

- un vecteur de poids d'entrée $\mathbf{w}_\ell \in \mathbb{R}^n$,
- un biais $b_\ell \in \mathbb{R}$,
- une fonction d'activation ReLU (Rectified Linear Unit, e.g. $g(\mathbf{w}_i, b_i, \mathbf{x}) = \max(0, \mathbf{w}_i^\top \mathbf{x} + b_i)$)
- un vecteur de poids de sortie $\boldsymbol{\beta}_\ell \in \mathbb{R}^m$ [9]

Contrairement aux réseaux classiques, les poids \mathbf{w}_ℓ et les biais b_ℓ sont fixés aléatoirement et ne sont pas ajustés durant l'apprentissage. Seule la matrice de sortie est estimée.

La sortie du ℓ ème neurone caché pour l'entrée \mathbf{x}_i est donnée par $h_\ell(\mathbf{x}_i) = g(\mathbf{w}_\ell^\top \mathbf{x}_i + b_\ell)$.

Ces sorties forment le vecteur $\mathbf{h}(\mathbf{x}_i) = [h_1(\mathbf{x}_i), \dots, h_L(\mathbf{x}_i)] \in \mathbb{R}^L$. La prédiction du modèle pour l'entrée \mathbf{x}_i s'exprime comme un produit vectoriel $\mathbf{f}(\mathbf{x}_i) = \mathbf{h}(\mathbf{x}_i)\boldsymbol{\beta}$, où la matrice $\boldsymbol{\beta} \in \mathbb{R}^{L \times m}$ regroupe les poids reliant la couche cachée à la couche de sortie [10]. Pour les N observations d'apprentissage, on forme la matrice des sorties de la couche cachée :

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}(\mathbf{x}_1) \\ \vdots \\ \mathbf{h}(\mathbf{x}_N) \end{bmatrix} = \begin{bmatrix} h_1(\mathbf{x}_1) & \cdots & h_L(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ h_1(\mathbf{x}_N) & \cdots & h_L(\mathbf{x}_N) \end{bmatrix} \in \mathbb{R}^{N \times L}. \quad (1)$$

La matrice des sorties cibles est quand à elle définie par ($^\top$ désigne la transposée du vecteur) :

$$\mathbf{T} = \begin{bmatrix} \mathbf{y}_1^\top \\ \vdots \\ \mathbf{y}_N^\top \end{bmatrix} = \begin{bmatrix} y_{11} & \cdots & y_{1m} \\ \vdots & \ddots & \vdots \\ y_{N1} & \cdots & y_{Nm} \end{bmatrix} \in \mathbb{R}^{N \times m}. \quad (2)$$

L'apprentissage du modèle [8] consiste à résoudre le système linéaire $\mathbf{H}\boldsymbol{\beta} = \mathbf{T}$ dont la solution minimale au sens des moindres carrés [11], utilisant la pseudo-inverse de Moore–Penrose (Cf. annexe A.1), est $\boldsymbol{\beta} = \mathbf{H}^\dagger \mathbf{T}$.

Pour améliorer la stabilité numérique et éviter les problèmes liés à un éventuel mauvais conditionnement de la matrice \mathbf{H} , on introduit une régularisation de type Ridge (ou Tikhonov). La solution des poids de sortie s'écrit alors :

$$\boldsymbol{\beta} = (\mathbf{H}^\top \mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{H}^\top \mathbf{T}, \quad (3)$$

où \mathbf{I} est la matrice identité et $\lambda > 0$ est un hyperparamètre de régularisation. Ce paramètre λ contrôle le compromis entre l'ajustement aux données et la complexité du modèle (taille des poids). Plus λ est

grand, plus la régularisation est forte (poids plus petits, risque de sous-apprentissage) ; à l'inverse, un λ trop petit réduit la régularisation et peut favoriser le sur-apprentissage.

En pratique, λ est choisi par validation croisée, en testant plusieurs valeurs pour identifier celle qui minimise l'erreur de prédiction sur un jeu de validation. Ainsi, la sortie finale du modèle pour une entrée \mathbf{x} s'exprime $\mathbf{f}(\mathbf{x}) = \mathbf{h}(\mathbf{x})\beta$.

Cette approche illustre la spécificité de l'ELM : seuls les poids de sortie β sont appris analytiquement, tandis que les poids d'entrée et les biais sont fixés aléatoirement. Cela confère au modèle une grande rapidité d'apprentissage, une bonne capacité de généralisation, et une simplicité d'implémentation.

3.2 Multi Input Multi Output (MIMO)

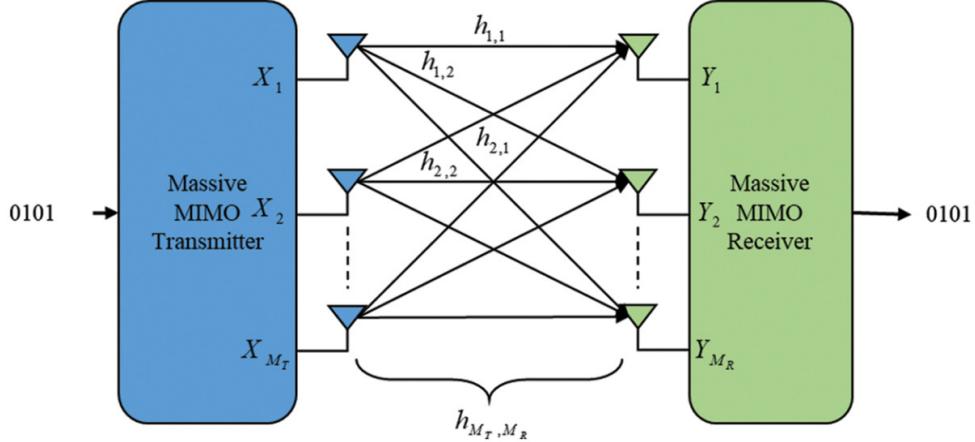


FIGURE 4 – Schéma de la structure du modèle MIMO

Le modèle MIMO (Multi-Input Multi-Output) constitue une généralisation des approches classiques SISO (Single-Input Single-Output). Alors que les modèles SISO traitent séparément chaque variable en entrée et en sortie, le MIMO prend en compte simultanément plusieurs séries temporelles, ce qui permet de capturer leurs interactions et de renforcer la cohérence globale des prévisions (Figure 5 [12])).

Mathématiquement, le modèle peut s'écrire sous la forme : $\mathbf{Y} = \mathbf{H}\mathbf{X}$ [13]

où :

- $\mathbf{X} \in \mathbb{R}^{M_T}$ représente le vecteur des entrées (ex. productions multi-sources),
- $\mathbf{Y} \in \mathbb{R}^{M_R}$ le vecteur des sorties (prévisions correspondantes),
- $\mathbf{H} \in \mathbb{R}^{M_R \times M_T}$ la matrice des coefficients reliant chaque entrée à chaque sortie.

Chaque élément h_{ij} de la matrice \mathbf{H} exprime l'influence de l'entrée X_j sur la sortie Y_i . Le modèle comporte donc $M_R \times M_T$ paramètres à estimer, ce qui permet de modéliser finement les corrélations croisées entre sources énergétiques.

Dans le cadre de la prévision énergétique multi-sources, le vecteur d'entrée \mathbf{X} regroupe les séries temporelles des différentes productions ou consommations (solaire, éolienne, hydraulique, thermique, biomasse, importations, etc.). Le vecteur de sortie \mathbf{Y} contient les prévisions correspondantes, ainsi que, le cas échéant, la consommation totale du réseau.

pour une entrée \mathbf{x} , la prédiction s'écrit : $\mathbf{f}(\mathbf{x}) = \mathbf{h}(\mathbf{x})\mathbf{W}_{\text{out}}$,

où $\mathbf{h}(\mathbf{x})$ désigne la représentation intermédiaire (par exemple les activations d'une couche cachée avec fonction ReLU) et \mathbf{W}_{out} la matrice de poids reliant cette représentation aux prédictions. Cette formulation permet de mutualiser les calculs entre les différentes sorties et d'optimiser la précision globale en exploitant les dépendances croisées.

Le modèle MIMO permet ainsi de capturer les corrélations entre sources et d'assurer la cohérence globale des prévisions. Cette approche est particulièrement pertinente lorsque les séries temporelles partagent des tendances communes (par exemple une dépendance aux conditions météorologiques) ou sont soumises à des contraintes physiques. Contrairement aux modèles indépendants, elle garantit une estimation conjointe de l'ensemble des sorties, renforçant la cohérence interne des prévisions.

3.3 Extension MIMO-MH : Multi-Horizon

Le modèle MIMO-MH (Multi-Input Multi-Output Multi-Horizon) est une généralisation directe du modèle MIMO, visant à intégrer simultanément les prévisions à différents horizons temporels. Dans un contexte énergétique, cela permet de capturer à la fois les interactions entre les différentes sources d'énergie et la dynamique temporelle de leur évolution future.

Le modèle MIMO-MH (Multi-Input Multi-Output Multi-Horizon) constitue une généralisation directe du modèle MIMO. Il vise à intégrer simultanément plusieurs horizons de prévision, ce qui permet de capturer à la fois les interactions entre les différentes sources d'énergie et la dynamique temporelle de leur évolution future.

Mathématiquement, le modèle MIMO-MH est formulé comme suit. Soit H le nombre d'horizons considérés (par exemple 1h, 3h, 6h, 12h, 24h). Pour chaque horizon $h = 1, \dots, H$, on définit :

- un vecteur d'entrée $\mathbf{X}^{(h)} \in \mathbb{R}^{M_T}$,
- un vecteur de sortie associé $\mathbf{Y}^{(h)} \in \mathbb{R}^{M_R}$

La relation de prévision à l'horizon h s'écrit :

$$\mathbf{Y}^{(h)} = \mathbf{H}^{(h)} \mathbf{X}^{(h)}, \quad (4)$$

où $\mathbf{H}^{(h)} \in \mathbb{R}^{M_R \times M_T}$ est la matrice de paramètre spécifique à l'horizon h .

L'extension MIMO-MH consiste à estimer simultanément l'ensemble des matrices $\{\mathbf{H}^{(h)}\}_{h=1}^H$ de manière cohérente.

Pour une vision unifiée, les H horizons sont regroupés en un vecteur d'entrée global $\mathbf{X}_{\text{MH}} \in \mathbb{R}^{H \times M_T}$ et en un vecteur de sortie global $\mathbf{Y}_{\text{MH}} \in \mathbb{R}^{H \times M_R}$:

$$\mathbf{X}_{\text{MH}} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \vdots \\ \mathbf{X}^{(H)} \end{pmatrix}, \quad \mathbf{Y}_{\text{MH}} = \begin{pmatrix} \mathbf{Y}^{(1)} \\ \vdots \\ \mathbf{Y}^{(H)} \end{pmatrix}. \quad (5)$$

Le modèle complet MIMO-MH s'écrit alors comme une relation bloc-diagonale :

$$\mathbf{Y}_{\text{MH}} = \begin{pmatrix} \mathbf{H}^{(1)} & & \\ & \ddots & \\ & & \mathbf{H}^{(H)} \end{pmatrix} \mathbf{X}_{\text{MH}}. \quad (6)$$

Cette formulation exploite la structure répétée des interactions pour chaque horizon tout en permettant un ajustement spécifique à chaque pas temporel. L'intérêt majeur du modèle MIMO-MH est de garantir la cohérence temporelle des prévisions tout en prenant en compte les interactions physiques et statistiques entre les sources d'énergie.

Au lieu de traiter chaque horizon indépendamment, l'agrégation des prévisions dans un cadre multi-horizon assure que les tendances temporelles (croissance, décroissance, cycles saisonniers) soient capturées et respectées à l'échelle du système complet.

En pratique, l'apprentissage des poids $\mathbf{H}^{(h)}$ ou de la matrice de sortie finale \mathbf{W}_{out} (dans le cas d'un réseau neuronal de type ELM) se fait conjointement pour l'ensemble des horizons. Cette approche mutualise les informations disponibles sur les différentes échelles temporelles, améliore la robustesse des prévisions et réduit les risques de sur-apprentissage liés à l'estimation séparée de chaque horizon.

Un encodage sinusoïdal est utilisé pour représenter les caractéristiques cycliques, selon [14] :

$$\sin_T(t) = \sin\left(2\pi \frac{t}{T}\right), \quad \cos_T(t) = \cos\left(2\pi \frac{t}{T}\right), \quad (7)$$

avec t l'heure (0–23) et $T = 24$. Cet encodage évite la discontinuité des variables catégorielles.

Pour les modèles MIMO et MIMO-MH développés au cours du stage (Figure 3), la paramétrisation est la suivante

- Entrées : 8 sources énergétiques observées sur 48 heures +2 composantes temporelles, soit 386 variables d'entrée ;
- Couche cachée : 1000 neurones ;
- Sorties : $8 \times 24 = 192$ sorties ;
- Horizon de prévision : 24 heure par pas horaire (par run).

Cette architecture permet de capter les interactions spatio-temporelles des sources et de la dynamique énergétique.

Nombre de paramètres :

- Entre l'entrée et la couche cachée : $386 \times 1000 = 386\,000$,
- Entre la couche cachée et la sortie : $1000 \times 8 = 8\,000$,
- Total : 394 000 paramètres.

Synthèse comparative :

- SISO : prédit une seule source pour un horizon donné,
- MIMO : prédit toutes les sources pour un horizon donné,
- MIMO-MH : prédit toutes les sources sur plusieurs horizons en même temps.



FIGURE 5 – Schéma des modèles SISO, MIMO, MIMO-MH

3.4 Réconciliation des Prévisions Hiérarchiques

La réconciliation vise à garantir la cohérence des prévisions à différents niveaux d'agrégation : sources individuelles, sous-systèmes et total global du réseau. Sans ce mécanisme, il est fréquent que la somme des prévisions par source diffère de la prévision du total, ce qui compromet la fiabilité opérationnelle.

La réconciliation dans le cadre MIMO-MH vise à garantir la cohérence des prévisions, à la fois entre les différentes sources d'énergie (cohérence spatiale) et entre les horizons temporels (cohérence temporelle) [15].

À chaque instant t , les prévisions non réconciliées issues d'un modèle MIMO ou MIMO-MH sont regroupées dans un vecteur $\hat{\mathbf{y}}_t^{\text{MH}}$, qui comprend l'ensemble des prévisions multi-sources et multi-horizons [16].

Soit :

- Un vecteur de prévisions non reconciliés $\hat{\mathbf{y}}_t^{\text{MH}}$ à chaque instant t issu d'un modèle MIMO ou MIMO-MH,
- Une matrice de contraintes hiérarchiques S , qui relie les niveaux désagrégés (sources individuelles) aux niveaux agrégés (totaux).

L'objectif est d'obtenir un vecteur de prévisions réconciliées \mathbf{Y} respectant les contraintes :

$$\hat{\mathbf{y}}_t^{\text{agg}} = \mathbf{S}\mathbf{b}_t, \quad (8)$$

où \mathbf{b}_t regroupe les prévisions au niveau le plus désagrégé (bottom level), et \mathbf{S} encode la structure d'agrégation du système énergétique.

$$S = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (9)$$

Propriétés de matrice d'agrégation S :

- Dimension : $S \in \mathbb{R}^{(8 \times 7)}$
- La première ligne contient la somme des 1 (contrainte d'agrégation pour la somme de toutes les sources sans la variable énergétique Total)
- On aura 8 lignes pour chaque variable énergétique somme de toutes les sources, Thermal, Hydro, Micro Hydro, Solar, Wind, BioEner (sans la variable énergétique Total)

Le modèle de prévision avant réconciliation est :

$$\hat{\mathbf{y}}_t^{\text{MH}} = \mathbf{S}\mathbf{b}_t + \boldsymbol{\varepsilon}_t, \quad (10)$$

où $\boldsymbol{\varepsilon}_t$ est l'erreur de réconciliation, supposée de moyenne nulle et de covariance diagonale \mathbf{W}_t .

La méthode des moindres carrés pondérés (WLS) fournit une solution analytique pour ajuster les prévisions \mathbf{b}_t de manière à minimiser la variance globale des erreurs, tout en imposant les contraintes structurelles de cohérence. La solution WLS s'écrit :

$$\mathbf{b}_t^{\text{WLS}} = (\mathbf{S}^\top \mathbf{W}_t^{-1} \mathbf{S})^{-1} \mathbf{S}^\top \mathbf{W}_t^{-1} \hat{\mathbf{y}}_t^{\text{MH}}, \quad (11)$$

où $\mathbf{W}_t = \text{diag}(\hat{\sigma}_1^2, \dots, \hat{\sigma}_n^2)$ est une matrice diagonale composée des variances historiques des erreurs de prévision [17, 18]. Les prévisions finales réconciliées sont obtenues par :

$$\tilde{\mathbf{y}}_t^{\text{MH}} = \mathbf{S}\mathbf{b}_t^{\text{WLS}}. \quad (12)$$

Cette approche garantit que les prévisions finales respectent simultanément les contraintes physiques du système énergétique et la cohérence temporelle entre les horizons de prévision [19, 20]. Elle offre un compromis efficace entre la précision des prévisions et la robustesse globale, tout en assurant une intégration cohérente des différentes sources et horizons du modèle MIMO-MH.

3.5 Méthodes Naïves et de Référence

Afin d'évaluer les performances des modèles avancés, deux approches de base, ne nécessitant aucun apprentissage, sont utilisées comme points de comparaison :

- Persistance horaire (Persistance h) : la prévision pour l'instant $t+h$ (h est l'horizon de prévision) est égale à la valeur observée à l'instant t $\hat{y}_{t+h} = y_t$.
- Persistance quotidienne (Persistance 24h) : la prévision pour l'heure t d'une journée donnée reprend la valeur observée à la même heure la veille $\hat{y}_t = y_{t-24}$.

Ces méthodes constituent des baselines minimales, permettant de mesurer les gains réels apportés par les modèles plus complexes.

En complément, deux méthodes avancées servent également de référence :

- TimeGPT¹⁰ [21] : basé sur une architecture Transformer pré-entraînée sur des milliers de séries temporelles. Il fournit des prévisions directement, sans apprentissage supplémentaire sur les données locales, en exploitant des représentations globales des dynamiques temporelles.
- Prophet AI : modèle statistique développé par Meta, adapté aux séries temporelles avec tendances et saisonnalités, offrant une approche robuste et facilement interprétable.

L'utilisation de ces méthodes de référence permet de positionner les performances du modèle MIMO-MH-ELM par rapport à des approches simples (persistance) et à des benchmarks avancés (TimeGPT, Prophet).

3.6 Métriques d'Evaluation

L'évaluation des performances du modèle MIMO-ELM repose sur plusieurs indicateurs standards :

- RMSE (Root Mean Square Error) : mesure l'erreur quadratique moyenne. Cette métrique est sensible aux valeurs extrêmes et pénalise fortement les grandes erreurs.
- MAE (Mean Absolute Error) : calcule l'erreur absolue moyenne. Plus robuste que le RMSE, il reflète la magnitude moyenne des écarts sans donner trop de poids aux valeurs aberrantes.
- MBE (Mean Bias Error) : détecte les biais systématiques :
 - $MBE > 0$: tendance à la sous-estimation,
 - $MBE < 0$: tendance à la surestimation.
- R^2 (coefficients de détermination) : indique la proportion de variance expliquée par le modèle.
 - $R^2 = 1$: prévision parfaite,
 - $R^2 = 0$: absence d'explicabilité (équivalent à la moyenne). Cet indicateur facilite également la comparaison entre méthodes.
- Gain par rapport à la persistance : mesure relative de performance. Par exemple, un gain de 0,45 signifie que l'ELM réduit l'erreur de 45% par rapport à la méthode de persistance. Un gain négatif indique au contraire que l'ELM fait moins bien que la persistance.

10. <https://github.com/Nixtla/nixtla>

Métriques	Formule	Normalisation
RMSE (Root Mean Square Error)	$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$	$n\text{RMSE} = \frac{\text{RMSE}}{\bar{y}}$
MAE (Mean Absolute Error)	$\text{MAE} = \frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $	$n\text{MAE} = \frac{\text{MAE}}{\bar{y}}$
MBE (Mean Bias Error)	$\text{MBE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)$	$n\text{MBE} = \frac{\text{MBE}}{\bar{y}}$
R^2 (coefficient de détermination)	$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$	
Gain par rapport à la persistance	$\frac{\text{Gain} = \frac{\text{RMSE}_{\text{persistance}} - \text{RMSE}_{\text{ELM}}}{\text{RMSE}_{\text{persistance}}}}$	

TABLE 1 – Métriques d’Évaluation

Ces métriques d’erreur assurent une évaluation rigoureuse et objective des performances des modèles.

4 Données Utilisées

Les performances des modèles de prévision dépendent directement de la qualité et de la représentativité des données. Pour capturer la dynamique des interactions entre sources, nous exploitons des séries temporelles horaires couvrant 2016–2022.

4.1 Nature des données

Les performances des modèles de prévision énergétique dépendent directement de la qualité et de la représentativité des données mobilisées. Pour analyser les interactions entre les différentes sources d’énergie, ce travail s’appuie sur des séries temporelles horaires détaillées, couvrant une période suffisamment longue pour refléter la variabilité réelle des productions.

Une série temporelle correspond à une suite d’observations indexées par le temps. Dans le cadre de ce rapport, elles représentent la production horaire d’électricité (en MWh) issue de différentes sources :

- Thermique,
- Hydraulique,
- Micro-hydraulique,
- Solaire photovoltaïque,
- Éolien,
- Bioénergies,
- Importations

Sont également inclus :

- Le coût moyen de production (€/MWh),
- La production totale (MWh).

Ces données sont fournies par EDF¹¹ pour la région Corse, ce qui en garantit la fiabilité.

4.2 Contexte énergétique corse

La Corse se distingue du reste de la France par son statut de Zone Non Interconnectée (ZNI). Le réseau est géré de manière centralisée par EDF SEI (Systèmes Énergétiques Insulaires), ce qui permet une meilleure adaptation aux contraintes locales mais entraîne des coûts de production plus élevés, du fait de la dépendance aux importations d’hydrocarbures. Ces surcoûts sont compensés par la Contribution au Service Public de l’Électricité (CSPE), permettant d’aligner les tarifs insulaires sur ceux de la France continentale (0,25 €/kWh en 2024).

11. <https://opendata-corse.edf.fr/pages/home0/>

Le réseau corse est marqué par plusieurs spécificités :

- Une demande variant de 130 à 450 MW, fortement corrélée aux conditions climatiques ($\approx 37\%$ de la consommation [22]),
- Une forte variabilité saisonnière liée au tourisme (≈ 3 millions de visiteurs/an),
- Une interconnexion limitée avec l'Italie via le câble sous-marin SACOI 2 (50 MW),
- Un mix énergétique composé en moyenne de 39,3 % thermique, 34,1 % EnR et 26,6 % importations [23].

4.3 Analyse des données

L'étude repose sur l'analyse de la puissance moyenne horaire (de 0h à 23h) par source d'énergie, calculée à différentes échelles temporelles :

- journalière,
- saisonnière,
- annuelle

Ces agrégations permettent de révéler les patterns temporels de production, essentiels pour la modélisation et la prévision.

4.4 Statistiques Descriptives

L'analyse du mix énergétique horaire moyen en Corse (2016–2022) (Tableau 2 , Figure 6) révèle des patterns temporels distincts selon les sources.

Le thermique a représenté 40,7% du mix (≈ 106 MWh en moyenne) Sa part a fluctué entre 36 % et 47 %, atteignant 47,5 % en 2022. Cette prédominance s'explique par la forte dépendance aux groupes électrogènes diesel, dont les émissions de CO_2 sont jusqu'à 15 fois supérieures à celles de la France continentale. Environ 70% de l'électricité consommée provient encore des combustibles fossiles ¹² [22].

Les Énergies renouvelables (EnR) contribuent à 34,1 % du mix ($\approx 80,4$ MWh/h). Leur répartition interne met en évidence des dynamiques contrastées :

- Hydraulique : principale EnR, avec une moyenne supérieure à 15 %. Sa contribution varie fortement selon les conditions climatiques : 19,9 % en 2020 (année humide) contre 12,3 % en 2022 (année de sécheresse).
- Solaire : en forte progression, passant de 7 % en 2016 à 11,7 % en 2022, portée par l'augmentation des capacités installées.
- Micro-hydraulique : relativement stable (1,7–3 %).
- Éolien et bioénergies : parts marginales (< 1 %), limitées par les contraintes d'installation et la petite taille des filières.

Les importations représentent en moyenne **28,5 %** du mix ($\approx 75,5$ MWh/h). Leur part recule légèrement (30,2 % en 2016 → 25,9 % en 2022), mais elles restent essentielles pour sécuriser l'approvisionnement, notamment la nuit (absence de solaire) ou en cas de faibles apports hydrauliques prolongés.

Comme le montrent le Tableau 7 et le tableau 3, la production horaire a globalement progressé entre 2016 et 2022. Toutefois, une baisse marquée en 2020 est observable, liée à la crise du COVID-19 et au confinement, qui ont entraîné une diminution de la consommation électrique.

En synthèse :

- Le thermique domine et tend à augmenter, reflétant la dépendance persistante aux hydrocarbures.
- Les EnR progressent mais restent très variables et dépendantes du climat.
- Les importations, bien que légèrement réduites, jouent un rôle crucial d'équilibrage.

12. <https://www.insee.fr/fr/statistiques/4481069>

Variable	Moy.	Écart-type	Max	Étendue	25%	50%	75%
Thermique (MWh)	106.14	46.57	311.02	311.02	64.96	100.46	137.20
Hydraulique (MWh)	47.08	49.61	196.58	199.15	4.39	29.11	79.82
Micro-hydraulique (MWh)	6.25	5.55	21.15	23.45	0.36	5.94	11.18
Solaire PV (MWh)	24.53	34.15	146.32	150.35	0.27	3.30	45.14
Éolien (MWh)	1.79	2.70	17.71	19.82	0.01	0.38	2.94
Bioénergies (MWh)	0.75	0.40	1.91	2.40	0.34	0.80	1.01
Importations (MWh)	74.26	36.38	154.56	203.72	51.28	75.06	97.16
Production totale (MWh)	260.81	62.17	510.94	510.94	216.22	256.36	304.01
Coût moy. prod. (€/MWh)	202.74	73.37	1290.95	1187.59	161.02	186.17	229.37

TABLE 2 – Résumé global de la production horaire moyenne (2016-2022)

Répartition de la production énergétique horaire moyenne globale par source d'énergie

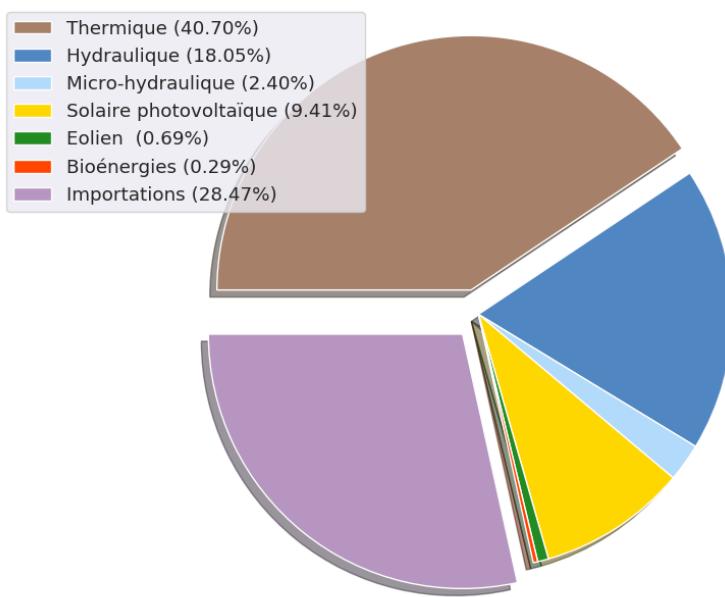


FIGURE 6 – Répartition de la production énergétique horaire moyenne globale par source d'énergie

Année	2016	2017	2018	2019	2020	2021	2022
Production totale (MWh)	249.81	260.21	260.40	265.17	251.47	270.49	268.14
Thermique (MWh)	98.24	116.18	93.44	110.88	90.53	106.34	127.40
Hydraulique (MWh)	47.67	39.11	65.11	41.23	49.94	53.50	32.94
Micro-hydraulique (MWh)	5.99	4.42	8.18	4.90	7.05	7.78	5.43
Solaire photovoltaïque (MWh)	17.62	20.76	20.91	24.85	27.04	29.10	31.42
Éolien (MWh)	3.75	2.49	1.51	1.04	1.28	1.36	1.10
Bioénergies (MWh)	1.04	0.93	0.97	0.82	0.71	0.45	0.32
Importations (MWh)	75.48	76.28	70.26	81.43	74.89	71.95	69.51

TABLE 3 – Résumé par année de la production horaire moyenne

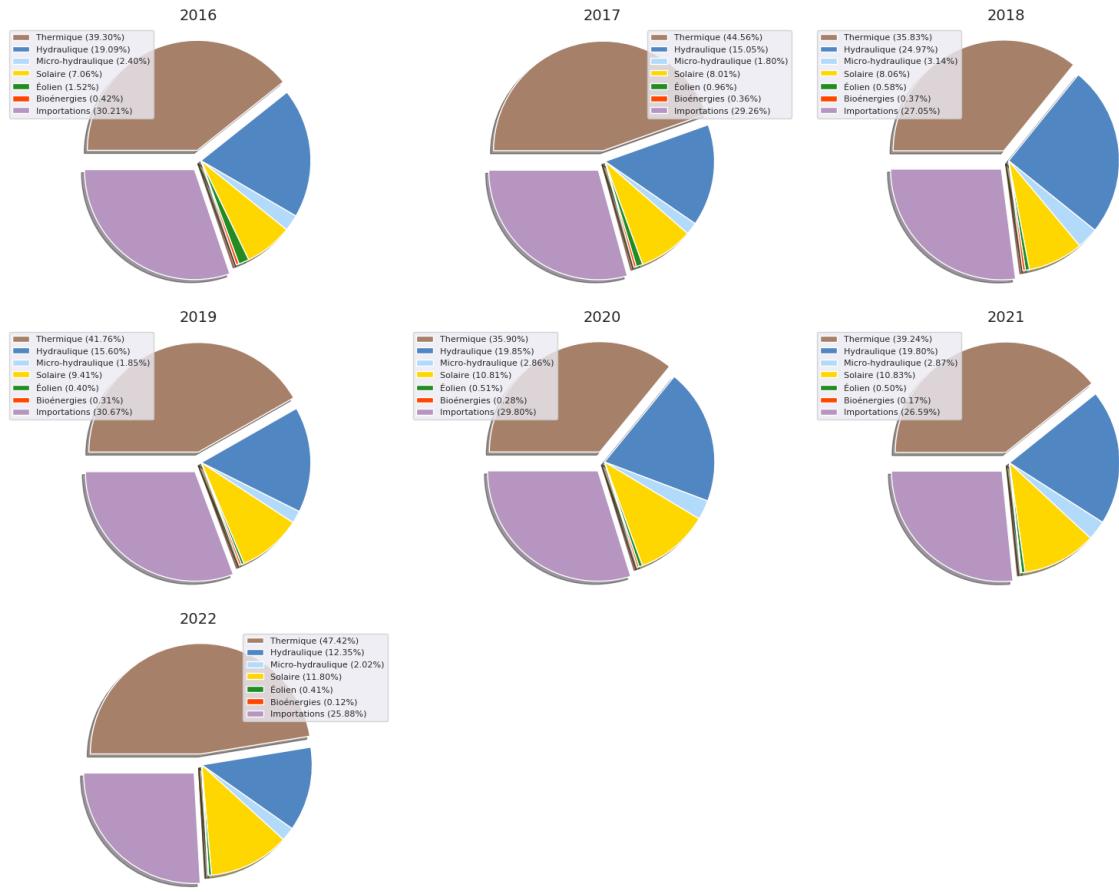


FIGURE 7 – Production énergétique horaire moyenne par source d'énergie (2016-2022)

4.5 Saisonnalités et Tendances

Les graphiques 8, 9 et le tableau 4 font apparaître des différences marquées dans les patterns de production horaire (0h à 23h) selon l'hiver, le printemps, l'été et l'automne

En hiver, la production totale moyenne horaire est de 315.38 MWh et est la plus élevée des 4 saisons. La production hydraulique bénéficie de l'abondance des pluies et du remplissage des lacs, ce qui assure une contribution relativement élevée au mix électrique (26.6% qui représente 84 MWh). D'un autre côté, le solaire reste très faible à cette période de l'année (4.85%), en raison d'un ensoleillement limité (angle solaire bas , jours courts où le pic de production solaire est entre 6 et 16h). Donc cette configuration impose une dépendance importante aux importations, qui couvrent encore environ 31% de la demande, soit 97.8 MW, notamment pour compenser les pointes nocturnes, lorsque la demande en énergie est au plus haut et que les ENR sont insuffisants. Les pics de production thermique (33.39%) se concentrent d'ailleurs entre 18h et 21h, période charnière marquée par les besoins en chauffage résidentiel, tandis que l'hydraulique reste dépendant de la variabilité des précipitations.

Au printemps, la production totale moyenne horaire s'élève à 244.4 MWh. L'ensoleillement s'intensifie, ce qui augmente la part du solaire ($\approx 11\%$), tandis que les pluies printanières et la fonte des neiges maintiennent une bonne disponibilité de l'hydraulique ($\approx 26\%$). Cette combinaison permet de réduire la dépendance aux importations, qui chutent autour de 25%, soit 61.28 MWh, rendant la saison plus propice à une couverture accrue par les ENR.

En été, La demande explose, portée par l'afflux touristique. En moyenne, la production totale horaire s'élève à 256.3 MWh. L'hydraulique, à l'inverse, s'effondre en ayant les réservoirs au plus bas, dû à la sécheresse estivale et aux faibles apports naturels, ce qui réduit sa contribution à moins de 6% du mix. Le solaire atteint son maximum, dépassant les 13% (pic de production entre 6h et 18h), Mais c'est insuffisant pour répondre à l'énorme demande. De ce fait, on a recours massivement au thermique (jusqu'à 53% de la production en moyenne horaire et atteint son pic de production entre 16h et 21h à 170 MWh contre 135.5 MWh globalement) et aux importations (environ 27%) pour maintenir l'équilibre du réseau. Le solaire, malgré sa progression, ne parvient pas à compenser l'effondrement hydraulique ni à couvrir les

périodes de forte sollicitation.

À l'automne, la production totale moyenne horaire atteint 227.75 MWh. les réserves hydrauliques amorcent leur reconstitution (11.5%), tandis que l'ensoleillement décline progressivement. C'est dans ce contexte intermédiaire que le thermique (46.7%) et les importations (30.35%) reprennent leur rôle central pour assurer la stabilité du réseau, même si la production solaire reste encore significative (9.26%).

Saison	Automne	Hiver	Printemps	Été
Production totale (MWh)	227.75	315.38	244.44	256.31
Thermique (MWh)	106.31	105.31	77.32	135.59
Hydraulique (MWh)	26.34	83.96	64.88	13.57
Micro-hydraulique (MWh)	2.76	9.89	10.66	1.72
Solaire PV (MWh)	21.09	15.23	27.33	34.18
Éolien (MWh)	1.42	2.34	2.13	1.28
Bioénergies (MWh)	0.69	0.71	0.82	0.76
Importations (MWh)	69.11	97.84	61.28	69.19

TABLE 4 – Résumé par saison de la production horaire moyenne

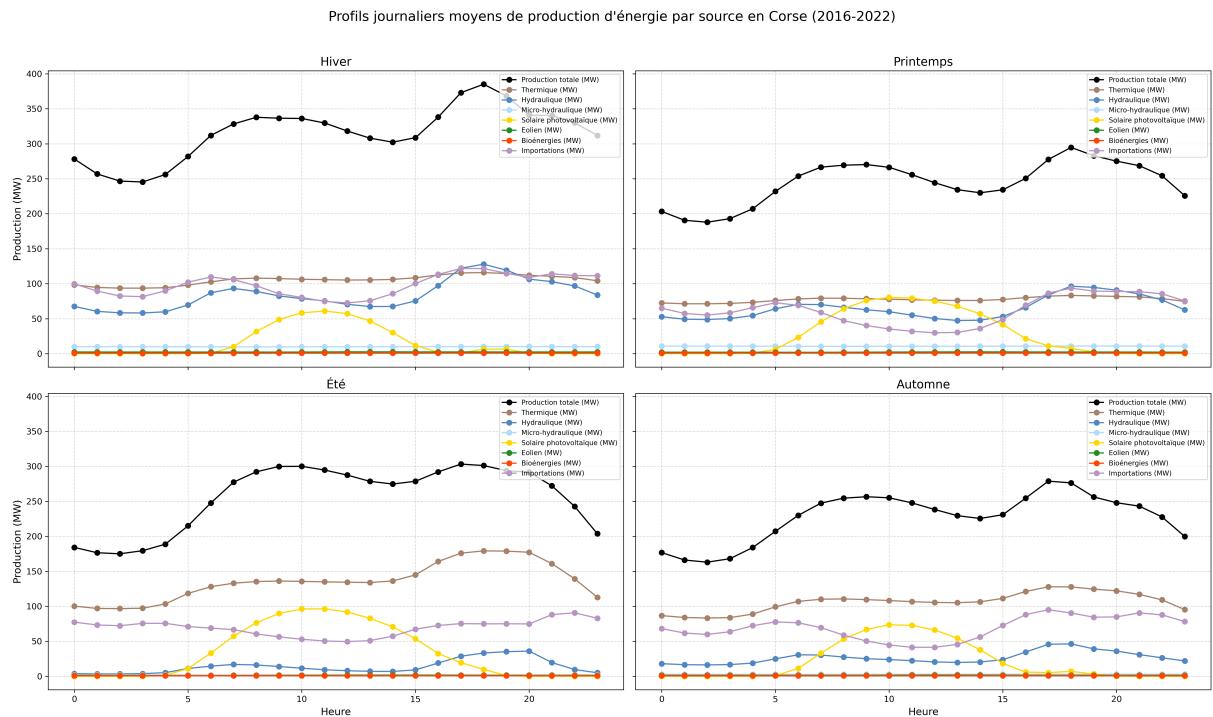


FIGURE 8 – Profils journaliers moyens de production d'énergie par source en Corse (2016-2022)

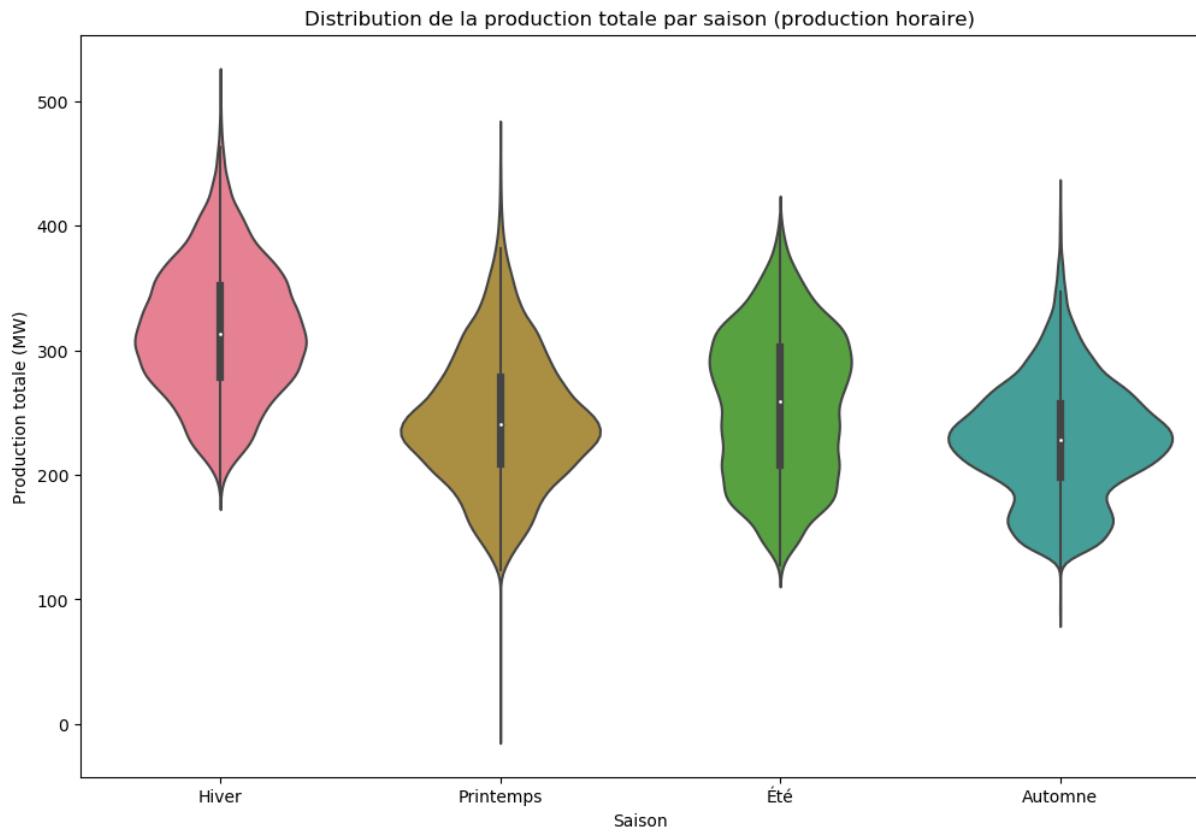


FIGURE 9 – Violin plot de la production totale (MWh) par saison de 2016 à 2022

4.6 Dépendances Statistiques et variables explicatives

La Figure 10 met en évidence les interdépendances du système énergétique corse à travers la corrélation de Spearman entre les différentes sources de production. Cette mesure statistique évalue la force et le sens d'une relation monotone (croissante ou décroissante) entre deux variables.

- Bleu : corrélation négative,
- Rouge : corrélation positive,
- Blanc/neutre : absence de corrélation (≈ 0).

4.6.1 Principaux résultats

Trois dynamiques structurantes apparaissent clairement :

1. Synergie hydraulique – micro-hydraulique Corrélation positive élevée (+0,764) :
 - Lorsque la production hydraulique augmente, la micro-hydraulique suit la même tendance.
 - Ces deux sources dépendent des mêmes conditions hydriques et météorologiques.
2. Substitution thermique – hydraulique
 - Corrélations négatives fortes (0,534 avec l'hydraulique ; 0,638 avec la micro-hydraulique).
 - En période sèche, la baisse de l'hydraulique est compensée par une augmentation du thermique pour maintenir l'équilibre du réseau.
3. Opposition solaire – importations
 - Corrélation négative (0,39) : les importations via le câble SACOI 2 diminuent lorsque la production solaire est optimale.

On observe également une faible corrélation positive entre thermique et importations (+0,18), suggérant un usage conjoint lors des pics de demande. Les autres corrélations restent proches de zéro et donc négligeables.

Ces résultats illustrent la logique opérationnelle du gestionnaire de réseau insulaire (EDF SEI) :

- substituer le thermique à l'hydraulique en période de sécheresse,

- exploiter la complémentarité hydraulique–micro-hydraulique,
- réduire les importations lorsque les EnR locales (notamment le solaire) sont abondantes.

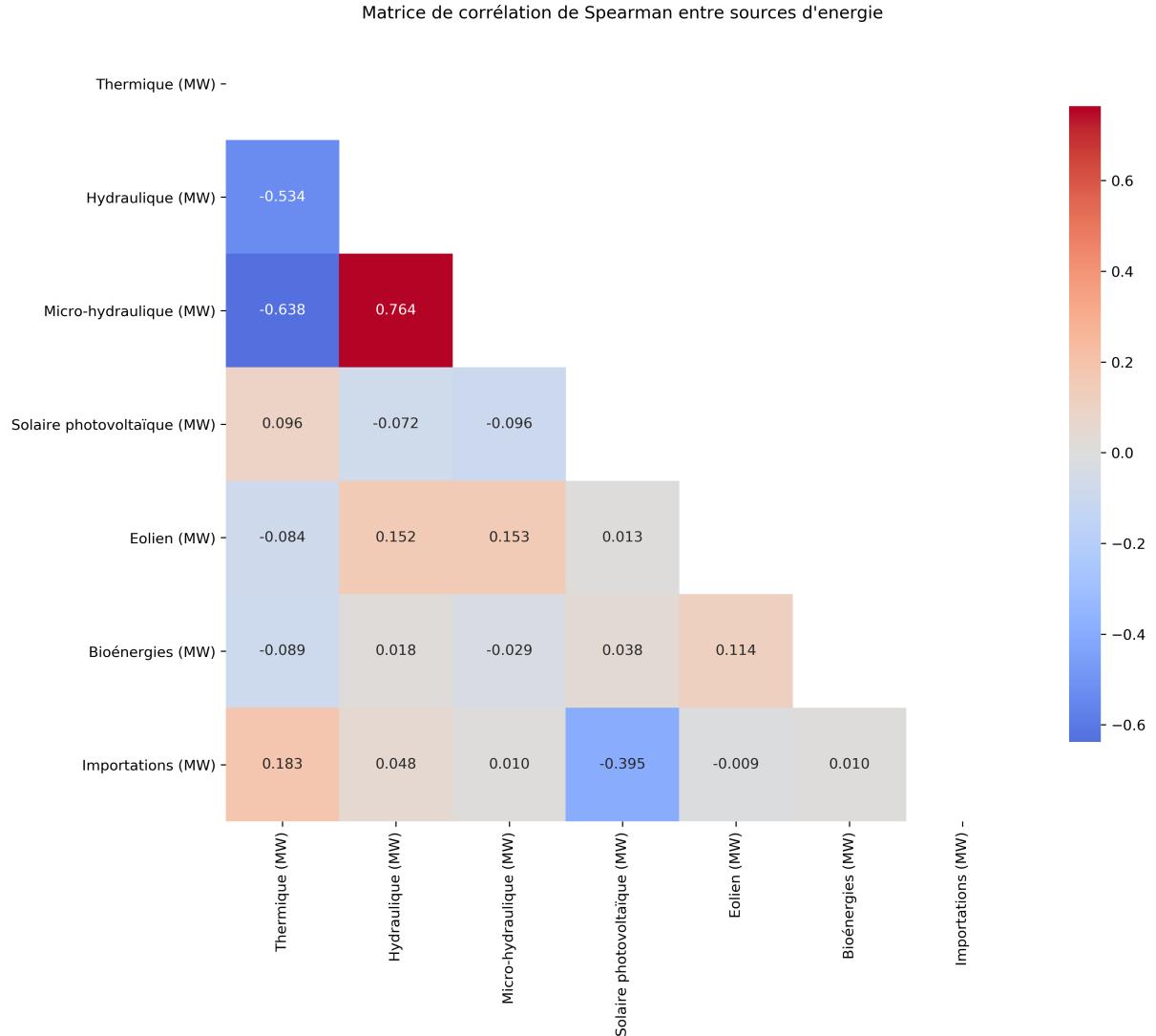


FIGURE 10 – Heatmap de la corrélation de Spearman des sources d'énergies

4.6.2 Analyse de significativité

- Un test t de Student [24] a été appliqué pour comparer la significativité saisonnière des productions :
- Sur 42 comparaisons, 41 sont significatives à $p < 0,05$ (97,6 %) et 40 à $p < 0,01$ (95,2 %).
 - Cela confirme une variabilité saisonnière robuste pour la quasi-totalité des sources.
 - Une seule comparaison apparaît marginale : thermique hiver vs automne, significative à $p \approx 0,037$. La différence de moyenne reste faible (105,3 MWh vs 106,3 MWh).

4.7 Résumé et perspectives

L'analyse met en évidence que l'été concentre les principales difficultés structurelles du système énergétique corse. La période estivale se caractérise par :

- une forte demande, accentuée par l'afflux touristique et l'usage croissant de la climatisation,
 - des apports hydrauliques très faibles liés à la sécheresse estivale,
 - une production solaire en progression, mais encore insuffisante pour couvrir l'ensemble des besoins.
- Ce déséquilibre impose un recours accru au thermique et aux importations pour maintenir la stabilité du réseau. À l'inverse, le printemps et l'automne offrent une meilleure disponibilité des énergies renouvelables, mais la dépendance au thermique reste dominante tout au long de l'année.

Ces constats soulignent plusieurs défis énergétiques majeurs pour la Corse :

1. Réduire la dépendance aux énergies fossiles en renforçant la part des énergies renouvelables (notamment le solaire et l'hydraulique).
2. Développer des solutions de stockage (batteries, STEP) pour compenser la variabilité des productions intermittentes.
3. Sécuriser l'approvisionnement grâce à une nouvelle interconnexion avec le continent (projet SACOI 3).
4. Accroître la flexibilité du réseau, via la gestion active de la demande, l'agrégation de petits producteurs et l'intégration optimisée des ressources locales.

Donc, la transition énergétique en Corse repose sur un équilibre entre :

- production locale,
- capacité de stockage,
- interconnexion avec le continent.

Cet équilibre est essentiel pour garantir une alimentation stable, durable et décarbonée, tout en réduisant les émissions de gaz à effet de serre.

5 Résultats et Discussion

5.1 Corrélation entre variables

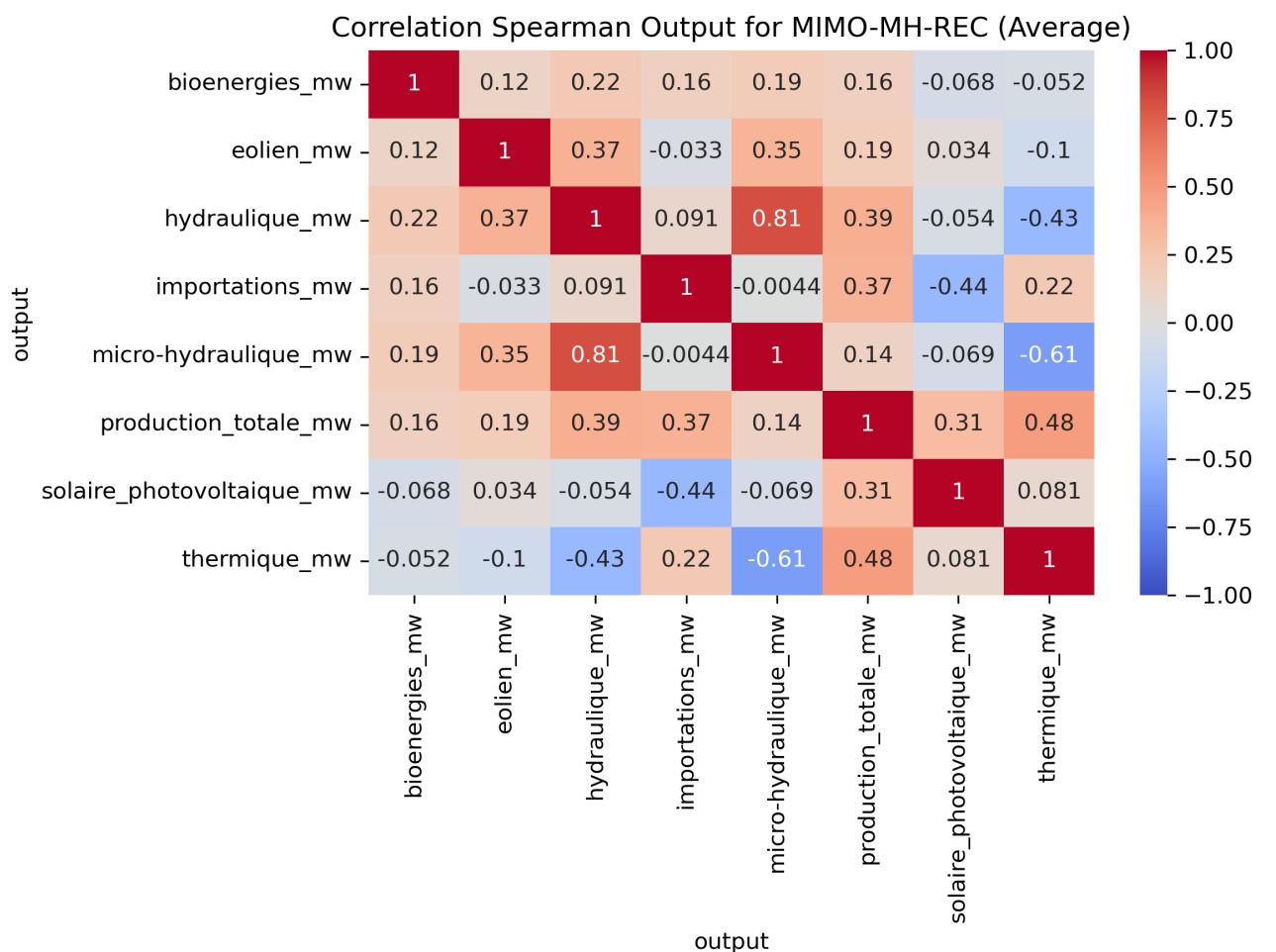


FIGURE 11 – Correlation de spearman

La matrice de corrélation de Spearman pour les outputs du modèle MIMO-MH-REC (Figure 11) révèle des corrélations significatives entre les variables énergétiques, telles que des associations positives

fortes entre la production hydraulique et les importations (0.91), ou entre la production totale et le solaire photovoltaïque (0.81), ainsi que des corrélations négatives notables comme entre les bioénergies et le solaire (-0.68). Cette mesure non paramétrique, utile pour détecter des relations monotones sans assumer une distribution normale des données, souligne les interdépendances multivariées inhérentes au système énergétique, justifiant l'approche MIMO-MH-REC qui intègre ces liens pour une modélisation plus holistique.

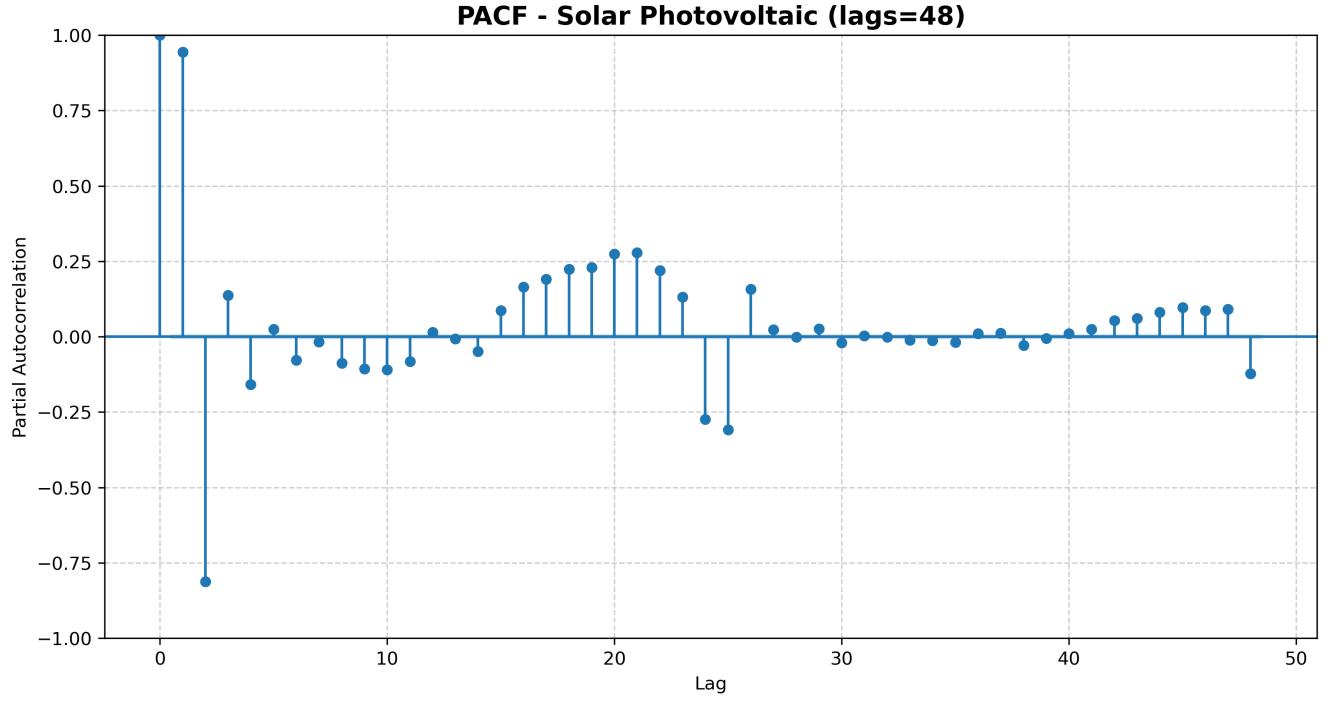


FIGURE 12 – $\text{pacf}_{\text{solar}}_{\text{photovoltaic}}$

Par ailleurs, la fonction d'autocorrélation partielle (PACF) pour la production solaire photovoltaïque (Figure 12), qui quantifie les corrélations directes à différents lags en contrôlant les effets intermédiaires, met en évidence une forte continuité d'une heure à l'autre (pics élevés aux lags courts) et un rythme quotidien clair (oscillations cycliques autour de 24 lags), reflétant les patterns diurnes influencés par l'ensoleillement. Le modèle MIMO-MH-REC capture efficacement ces dépendances temporelles et multivariées, comme en attestent ses performances en prévision pour les sources volatiles comme le solaire.

5.2 Résultat à court terme (horizon 1 h)

La comparaison des résultats obtenus avec le benchmark Matlab (tableau 5 et figure 13) a permis de valider la justesse de l'implémentation du modèle sous Python. Les valeurs des principales métriques (nRMSE, R^2 , Gain) se révèlent très proches de celles de la version initiale, confirmant ainsi la robustesse de la conversion.

À l'horizon d'une heure, les performances sont particulièrement satisfaisantes : le nRMSE atteint une valeur de 0,10, tandis que le R^2 se situe à un niveau proche de 1, traduisant une excellente capacité explicative. L'allure des courbes prédictives confirme également la cohérence méthodologique.

L'un des objectifs majeurs de cette migration vers Python était de rendre le modèle accessible à une communauté scientifique plus large, dans une logique d'ouverture et de reproductibilité. À ce titre, l'ensemble du code, des données et des résultats a été mis à disposition en open source sur ce dépôt Git : https://github.com/YJheelan/Program_SISO_Prophet_Time_gpt_MIMO_MH_Reconciliation.

TABLE 5 – Comparaison des performances du modèle MIMO entre Matlab / Python à l'horizon 1h

Source	nRMSE	Gain	nMAE	nMBE	R ²	Écart max (pp)
Total_MW	0.022 / 0.0217	0.668 / 0.67	0.014 / 0.014	-0.000 / -0.000	0.991 / 0.896	+0.005 (Gain)
Thermal_MW	0.051 / 0.051	0.346 / 0.347	0.035 / 0.034	-0.003 / -0.003	0.980 / 0.980	+0.009 (Gain)
Hydro_MW	0.223 / 0.223	0.338 / 0.339	0.141 / 0.141	0.012 / 0.012	0.968 / 0.968	+0.002 (nMBE)
Solar_MW	0.179 / 0.179	0.619 / 0.620	0.109 / 0.109	0.010 / 0.010	0.982 / 0.982	-0.000 (nMBE)
Wind_MW	0.422 / 0.422	0.024 / 0.023	0.226 / 0.226	0.003 / 0.003	0.918 / 0.918	-0.000 (Gain)
BioEner_MW	0.122 / 0.122	-0.068 / -0.074	0.062 / 0.063	0.032 / 0.032	0.946 / 0.945	+0.009 (Gain)
Import_MW	0.137 / 0.137	0.269 / 0.270	0.098 / 0.097	0.004 / 0.004	0.911 / 0.911	+0.002 (Gain)

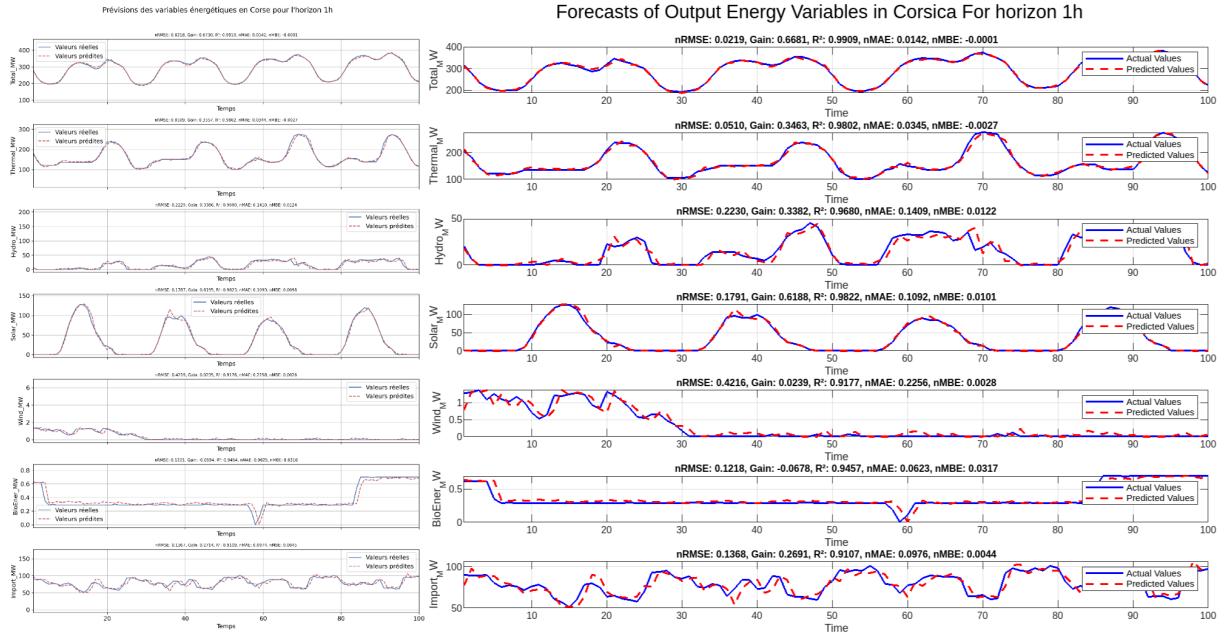


FIGURE 13 – Prévision énergétique multi-sources via une approche MIMO sur python (à gauche) et Matlab (à droite)

La figure 14 présente les comparaisons temporelles entre les prédictions du modèle MIMO, les valeurs réelles, la persistance horizon et la persistance quotidienne pour chaque variable énergétique. Cela permet de mettre en évidence la capacité du modèle à suivre les tendances générales des séries temporelles, bien que certaines fluctuations rapides ne soient pas parfaitement capturées.

Comparaison des modèles pour l'horizon 1h

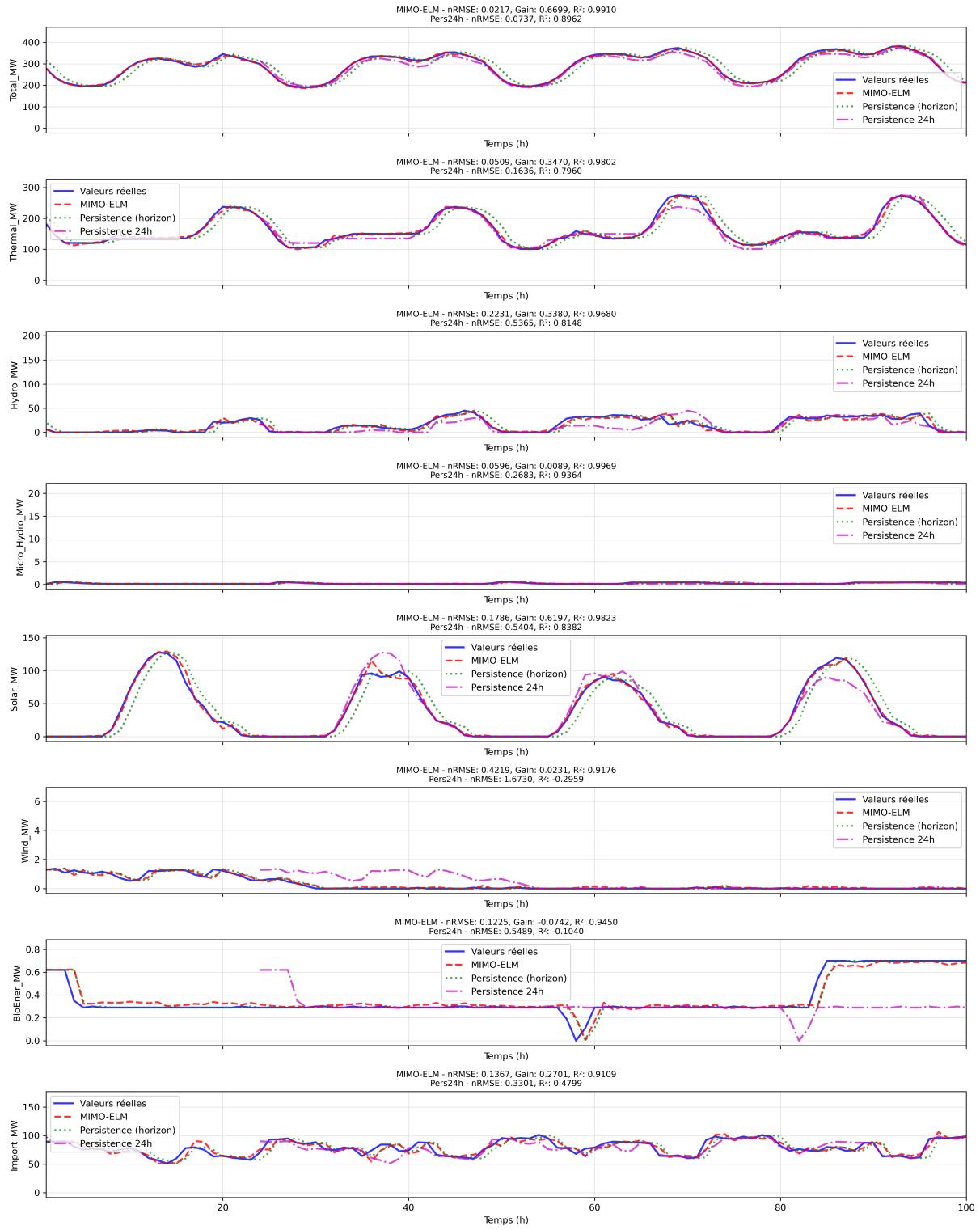


FIGURE 14 – Prévisions des variables énergétiques en Corse pour l'horizon 1h sur Python via MIMO et comparaison avec la persistance horizon et la persistance quotidienne

5.3 Réconciliation

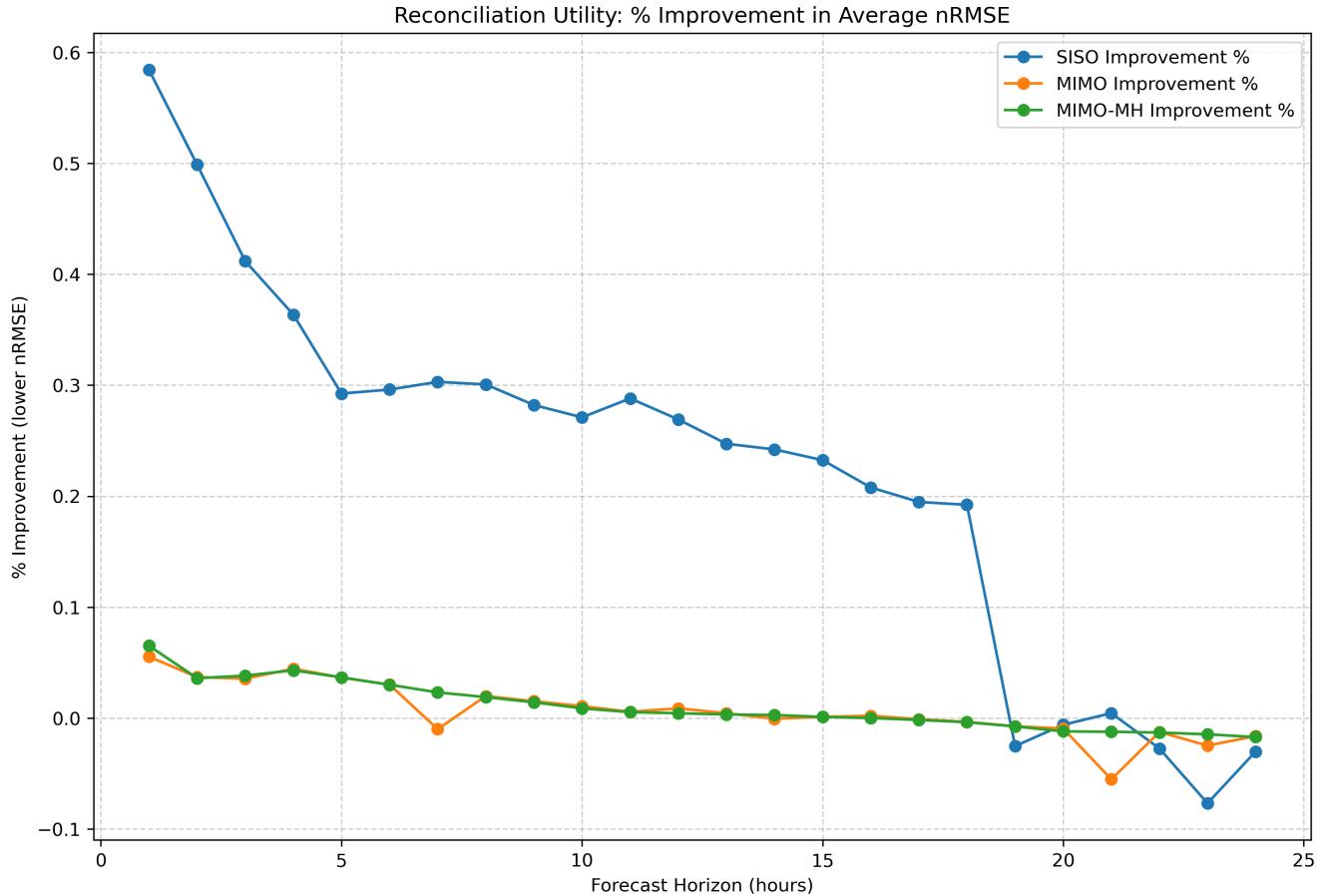


FIGURE 15 – Amélioration apportée par la réconciliation

La figure 15 illustre l’impact de l’application d’une technique de réconciliation hiérarchique sur la performance des modèles. Cette méthode vise à garantir la cohérence des prévisions, en s’assurant que la somme des productions individuelles (solaire, éolien, etc.) corresponde bien à la prévision de la production totale à chaque horizon. Pour le modèle SISO, la réconciliation apporte une amélioration spectaculaire de la précision. La réduction de l’erreur (nRMSE) atteint près de 60% pour le premier horizon et reste supérieure à 20% jusqu’à 18 heures. Cela démontre que les prévisions de base générées indépendamment par le modèle SISO sont fortement incohérentes, et que le processus de réconciliation corrige efficacement ces erreurs structurelles. À l’inverse, pour les modèles MIMO et MIMO-MH, l’amélioration est marginale, voire légèrement négative sur les horizons longs. L’amélioration ne dépasse jamais 5% et devient nulle après 8 heures. Ceci suggère que ces stratégies, en prédisant l’ensemble des variables et des horizons simultanément, produisent déjà des prévisions de base qui sont intrinsèquement cohérentes. La réconciliation n’a donc que peu d’erreurs à corriger et son application n’apporte pas de gain significatif.

La réconciliation est un outil essentiel pour améliorer la fiabilité des modèles simples qui traitent chaque série de manière indépendante (comme le SISO), mais son utilité est très limitée pour des architectures multi-sorties (MIMO) déjà structurellement cohérentes.

5.4 Résultats

Les Figures 16 à 23 présentent une visualisation comparative des performances des cinq modèles (MIMO-MH-REC, MIMO-REC, Prophet, SISO-REC et TimeGPT) pour chaque source d’énergie (bio-énergies, éolienne, hydraulique, importations, micro-hydraulique, production totale, solaire photovoltaïque et thermique), évaluées via les métriques nRMSE (normalized Root Mean Square Error), nMAE (normalized Mean Absolute Error), nMBE (normalized Mean Bias Error) et R^2 (coefficient de détermination) sur des horizons de prévision de 1 à 24 heures.

La Figure 16 offre un résumé global des métriques normalisées agrégées. Globalement, les graphiques révèlent une dégradation progressive des performances avec l'augmentation de l'horizon pour toutes les métriques d'erreur (nRMSE et nMAE augmentent de manière monotone, tandis que nMBE oscille autour de zéro avec des biais occasionnels positifs ou négatifs), et R^2 diminue de valeurs proches de 1 à court terme vers 0.8-0.9 à long terme), reflétant l'accumulation d'incertitudes temporelles inhérentes aux séries énergétiques. SISO-REC (rouge) se distingue par des courbes relativement lisses sans pics extrêmes notables dans la plupart des panels, indiquant une stabilité améliorée par rapport à des configurations antérieures, bien qu'il reste légèrement au-dessus des modèles MIMO pour les erreurs. MIMO-MH-REC (bleu) et MIMO-REC (orange) montrent des courbes superposées et souvent inférieures aux autres pour les erreurs (nRMSE et nMAE) sur les sources stables comme l'hydraulique, les importations et la micro-hydraulique, surpassant Prophet (vert) de 5 - 10% en termes relatifs à long terme. Pour les sources volatiles (éolienne et solaire), les MIMO maintiennent des courbes basses et croissantes linéairement. Prophet excelle pour les agrégats stables (production totale et thermique), avec des courbes plus lisses et des nMBE proches de zéro, minimisant les biais systématiques. Dans les panels R^2 , les modèles MIMO atteignent des valeurs supérieures (proches de 0.95-0.98 à court terme) pour les bioénergies et l'éolien, indiquant une meilleure explication de la variance, alors que Prophet domine pour le solaire avec une décroissance plus graduelle.

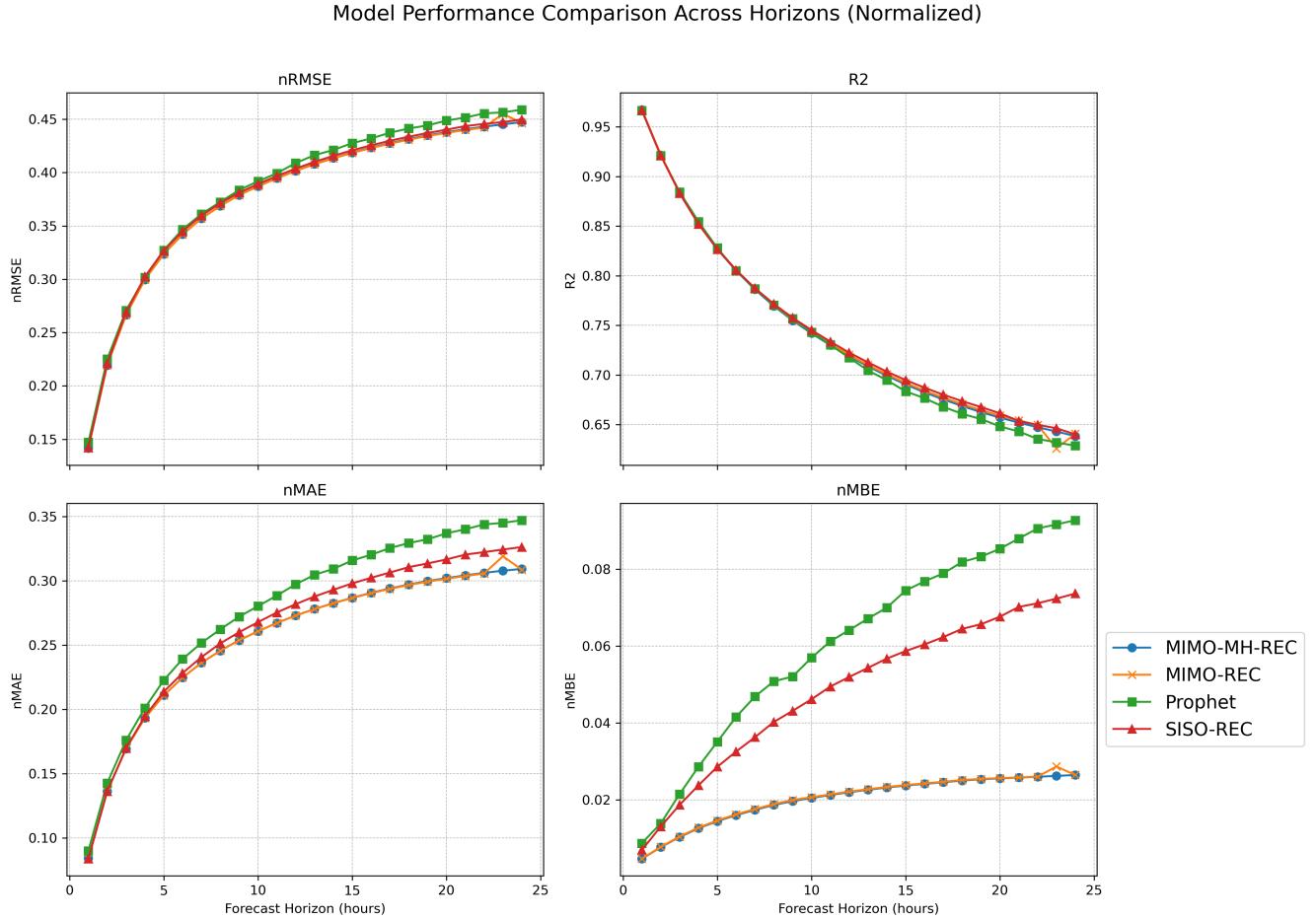


FIGURE 16 – Benchmarks des métriques normalisées (moyenne des variables énergétiques)

Une analyse détaillée par métrique suit, basée sur les tendances visuelles des courbes et des comparaisons relatives.

5.4.1 nRMSE

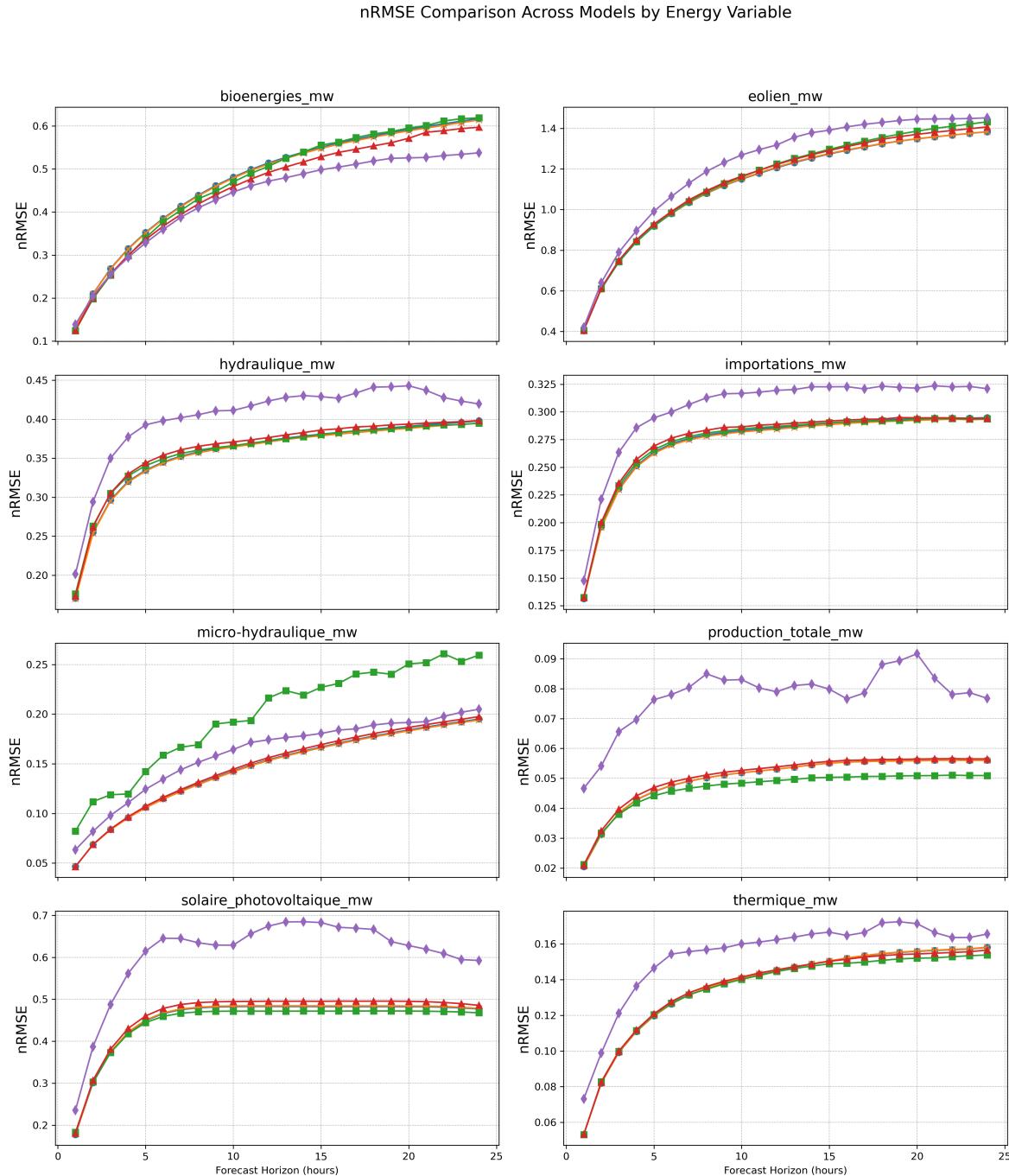


FIGURE 17 – Benchmarks nRMSE

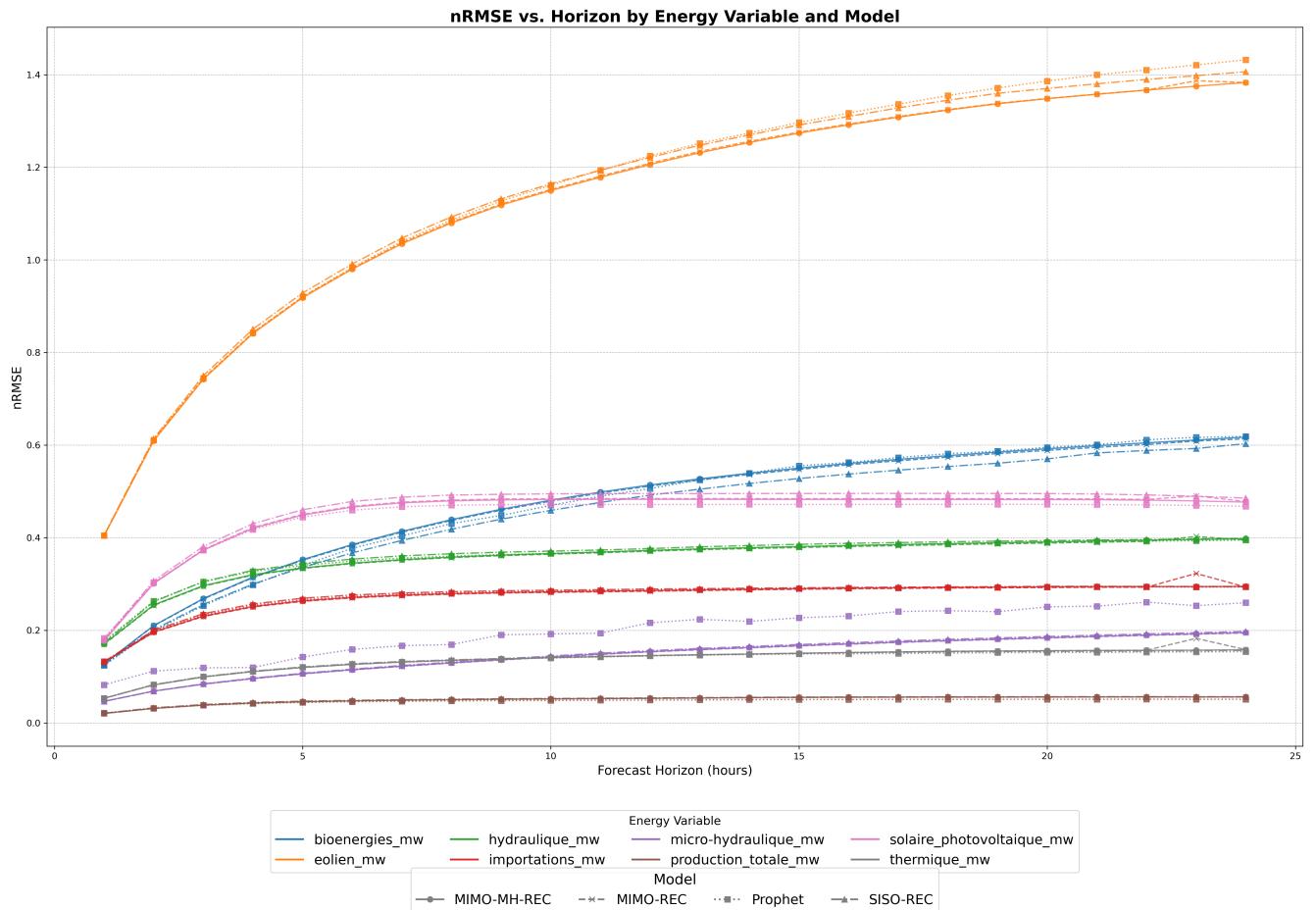


FIGURE 18 – Benchmarks par variable (nRMSE)

Les Figures 17 et 18 illustrent l'évolution du nRMSE (normalized Root Mean Square Error) pour les modèles MIMO-MH-REC, MIMO-REC, Prophet, SISO-REC et TimeGPT. Cette métrique mesure l'erreur relative normalisée, où des valeurs plus faibles indiquent une meilleure précision de prévision, évaluée pour des horizons de prévision allant de 1 à 24 heures. Globalement, les courbes confirment une augmentation progressive et plutôt linéaire du nRMSE avec l'horizon pour tous les modèles et sources, reflétant la difficulté croissante des prévisions à long terme due à l'accumulation d'incertitudes.

Les tendances visuelles mettent en évidence :

- Une production totale hautement prédictible (erreurs entre 2,2 % et 5,5 %), effet direct de l'agrégation des différentes sources.
- Une production éolienne difficile à anticiper, avec des erreurs atteignant 138 % aux horizons longs, reflet de sa forte variabilité météorologique.
- Des performances stables pour le thermique et les importations, avec des nRMSE inférieurs à 35 %.
- Un production solaire sous les 50% d'erreurs pour les modèles MIMO et Prophet.

La dynamique temporelle confirme une évolution progressive des performances : excellentes à court terme (1–6 h), acceptables à moyen terme (7–15 h), et plus critiques au-delà de 16 h.

Les modèles MIMO-MH-REC (bleu) et MIMO-REC (orange) dominent souvent les panels pour l'hydraulique et la micro-hydraulique avec des courbes les plus basses, indiquant une supériorité ou une équivalence par rapport à Prophet (vert) et TimeGPT (violet), qui tend à surperformer à court terme pour les bioénergies mais à diverger vers le haut pour l'éolien et le solaire. Prophet offre la courbe la plus stable pour la production totale, restant sous 6% d'erreur à 24h, et SISO-REC (rouge) suit de près les MIMO sans écarts majeurs, suggérant une bonne généralisation multivariée. TimeGPT, bien que compétitif initialement, présente des valeurs globalement plus élevées pour les sources volatiles, soulignant une moindre robustesse à long terme. Il est à noter que le modèle TimeGPT-1-long-horizon est utilisé ici avec le paramètre finetune steps = 0. Une augmentation de ce paramètre améliorerait encore ses performances, en particulier pour les sources volatiles telles que l'éolien et le solaire.

Pour une analyse détaillée par source, incluant les moyennes du nRMSE sur les 24 horizons et les comparaisons horizon par horizon, les tableaux 7 à 13 et 14 en annexe B fournissent les valeurs numériques exhaustives.

5.4.2 nMAE

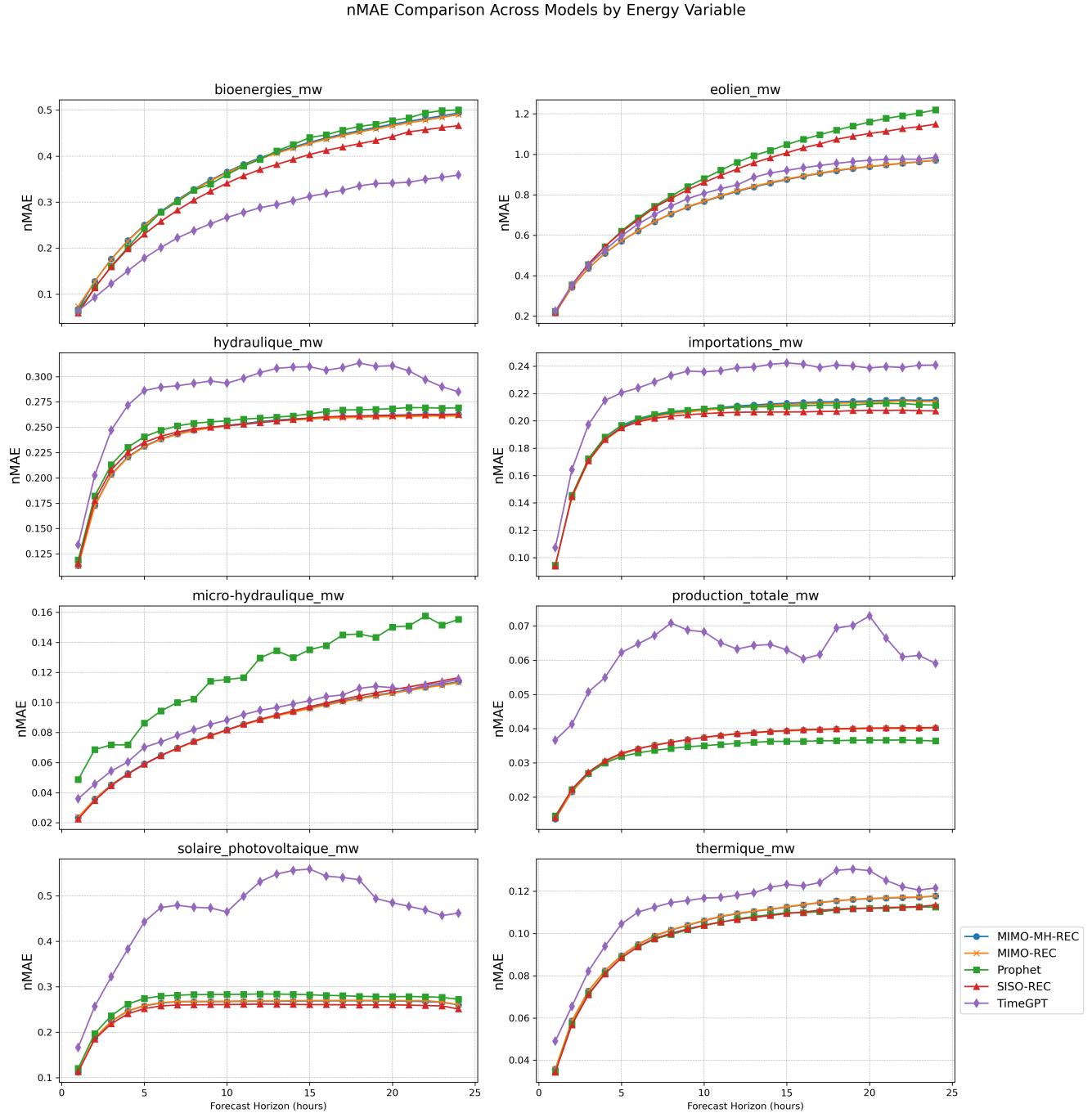


FIGURE 19 – Benchmarks nMAE

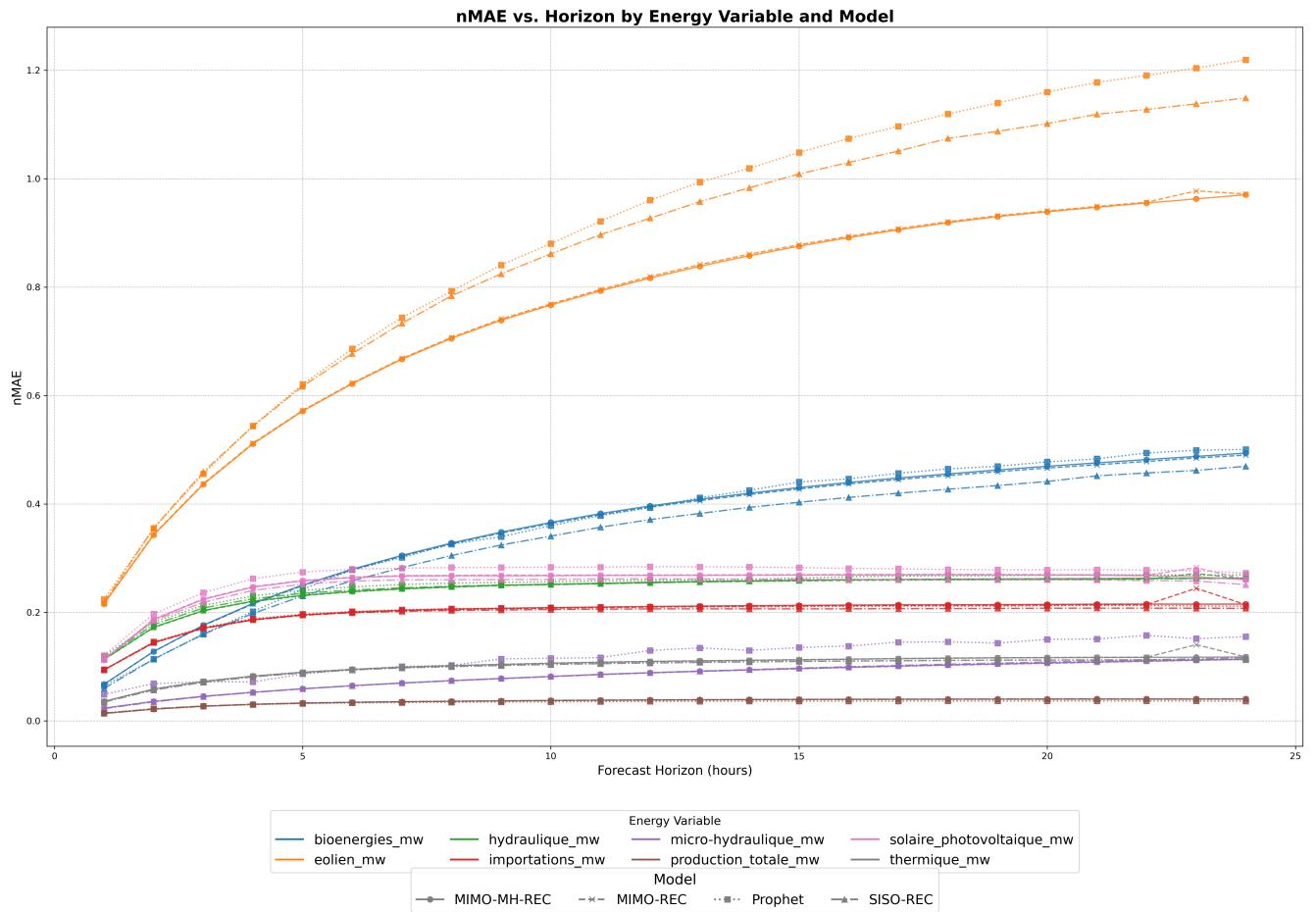


FIGURE 20 – Benchmarks par variable (nMAE)

Concernant le nMAE (Figure 19 et 20), les patterns sont similaires au nRMSE mais avec des valeurs absolues plus basses ($\sim 0.3 - 0.4$ pour éolien à 24h), soulignant une erreur absolue relative moindre. Prophet émerge comme le meilleur pour le solaire et la production totale (courbes vertes les plus plates), tandis que les modèles MIMO sont excellents pour les importations et le thermique (courbes bleu/orange confondues sous les autres). TimeGPT montre une performance dégradée pour les volatiles, avec une courbe violette s'élevant rapidement après 5h pour l'éolien et le solaire.

5.4.3 nMBE

nMBE Comparison Across Models by Energy Variable

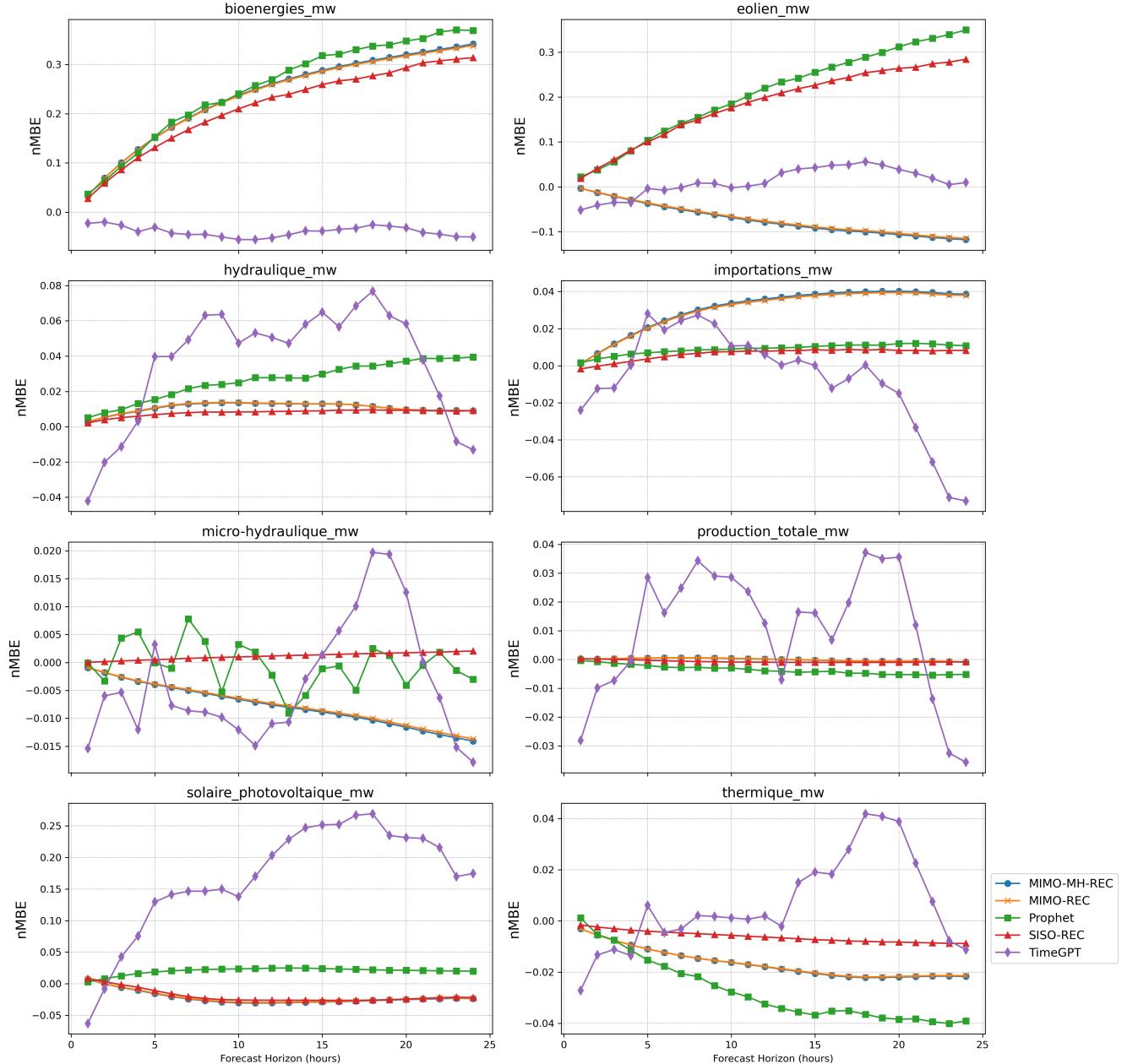


FIGURE 21 – Benchmarks nMBE

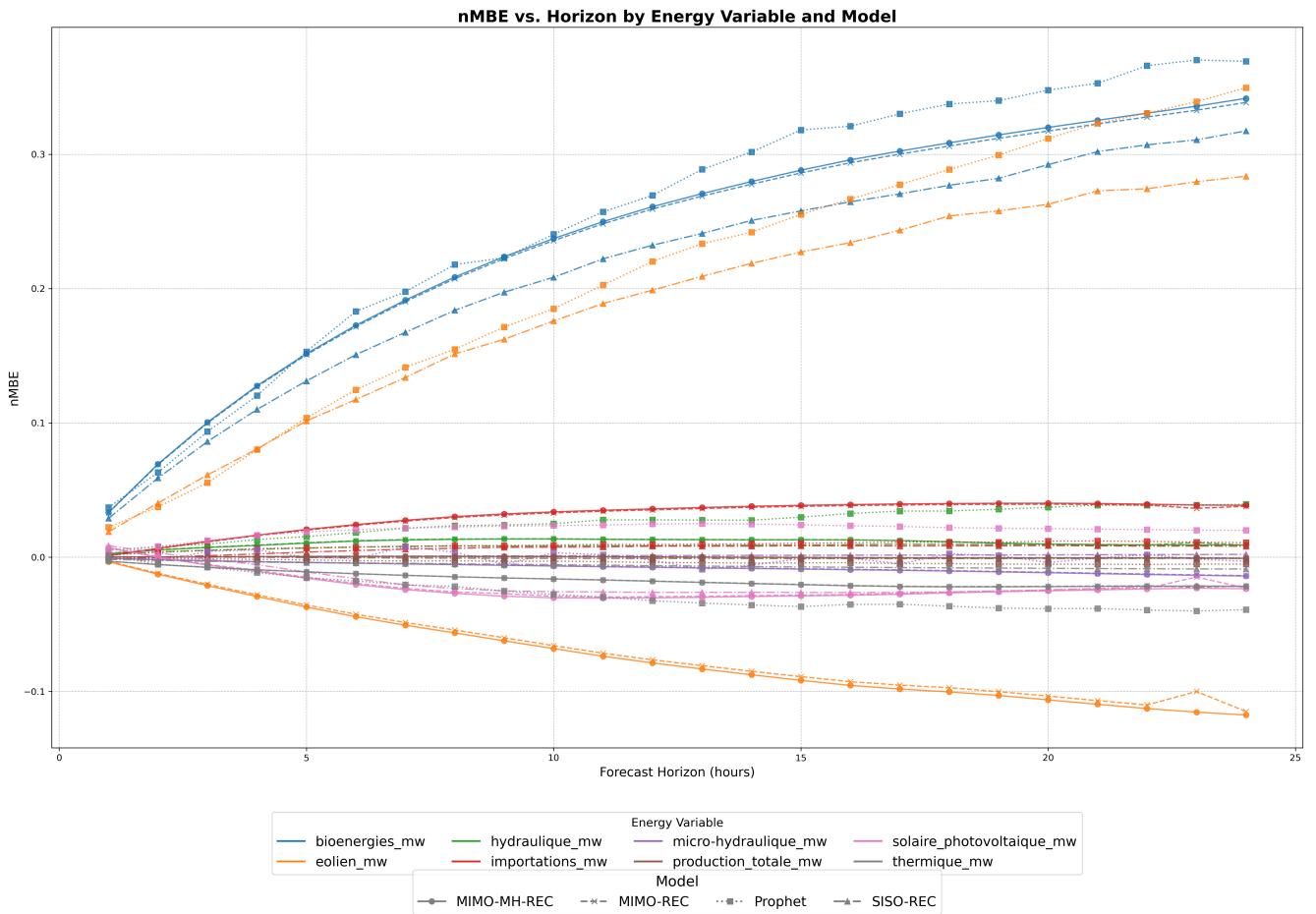


FIGURE 22 – Benchmarks par variable (nMBE)

Les Figures 21 et 22 révèle des biais faibles globalement (< 0.1), avec des oscillations autour de zéro : positifs pour Prophet sur le solaire (jusqu'à 0.2 à mi-horizon, reflétant une surestimation diurne), et négatifs pour TimeGPT sur l'hydraulique (vers -0.05 à long terme). Les MIMO minimisent les biais pour la plupart des sources, avec des lignes bleu/orange proches de l'axe horizontal.

5.4.4 R^2

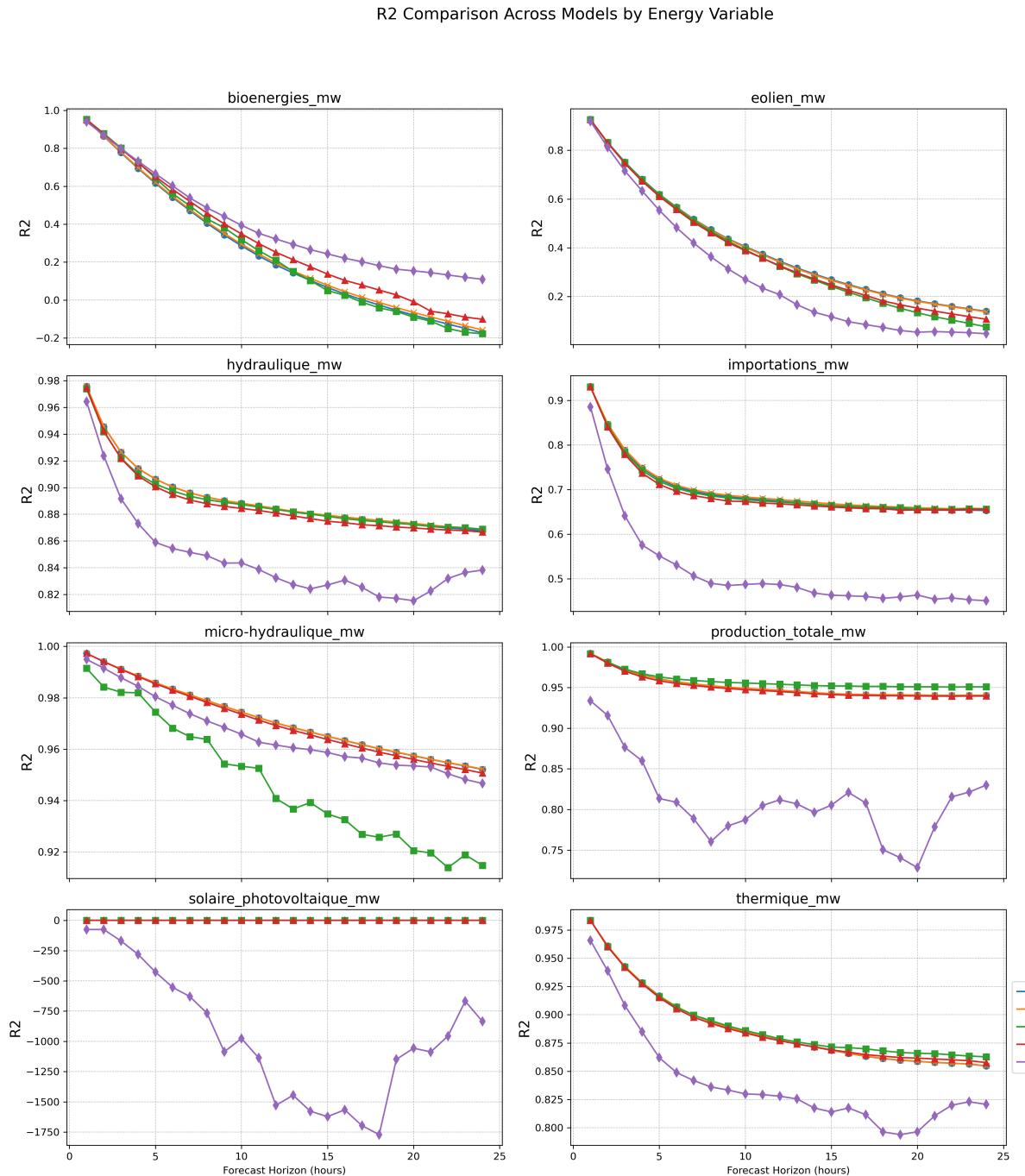


FIGURE 23 – Benchmarks R^2

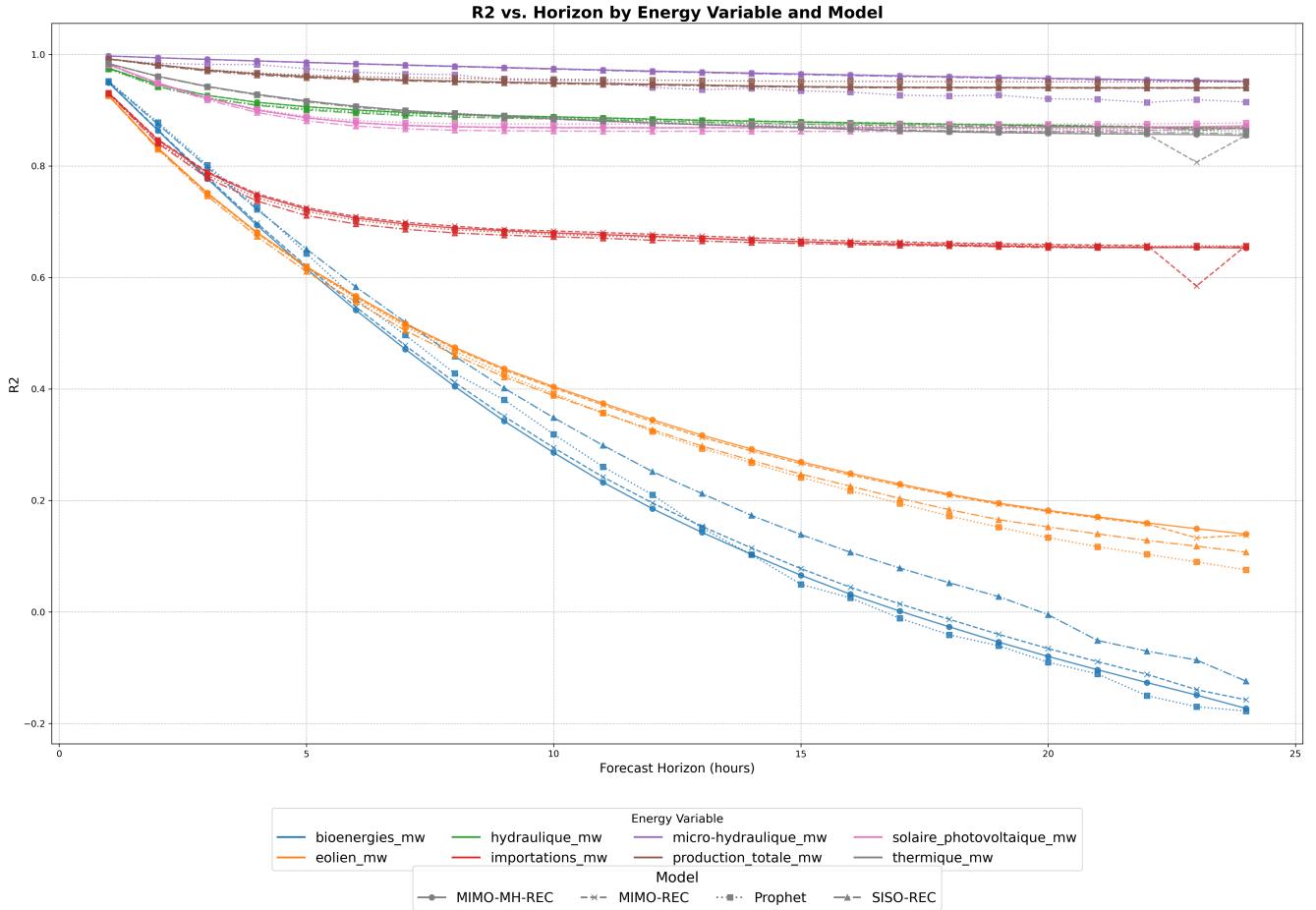


FIGURE 24 – Benchmarks par variable (r^2)

Enfin, pour le R^2 (Figure 23 et 24), les valeurs initiales élevées (0.95+) diminuent vers 0.7 - 0.9, avec les modèles MIMO atteignant les pics les plus hauts pour l'éolien et les bioénergies (courbes orange/bleu descendant le moins vite), indiquant une meilleure capture de la variance non linéaire. Prophet maintient des R^2 solides pour les agrégats (proche de 0.98 à 24h pour production totale), tandis que TimeGPT chute plus bas pour le solaire (~ 0.6).

5.5 Résumé des résultats

Les figures 16 à 23 et les tableaux 7 à 14 présentent les performances des cinq modèles (MIMO-MH-REC, MIMO-REC, Prophet, SISO-REC et TimeGPT) pour chacune des huit sources d'énergie. Globalement, les erreurs augmentent avec l'horizon, en raison de l'accumulation progressive d'incertitudes. L'évaluation comparative de ces cinq modèles, appliquée aux huit sources d'énergie et aux horizons de 1 à 24 h, révèle une dégradation monotone des performances : les valeurs de nRMSE et nMAE croissent de 0,02–0,4 (à court terme) à 0,05–1,4 (à long terme), tandis que le R^2 décroît de plus de 0,95 à 0,7–0,9. Cette tendance s'explique par l'accumulation d'incertitudes temporelles.

5.5.1 Comparaison par source

- Bioénergies : TimeGPT est le plus performant (nRMSE moyen = 0,432), grâce à sa capacité à modéliser des tendances stables. Prophet domine légèrement à court terme.
- Éolienne : MIMO-MH-REC et MIMO-REC sont les plus robustes (nRMSE $\approx 1,13$). Prophet reste compétitif à court terme, mais TimeGPT se dégrade rapidement.
- Hydraulique : MIMO-REC domine (0,355), suivi de près par MIMO-MH et Prophet ; TimeGPT reste en retrait.
- Importations : MIMO-REC offre les meilleures performances (0,271), légèrement devant Prophet et MIMO-MH.

- Micro-hydraulique : MIMO-REC et MIMO-MH dominent largement (0,145), Prophet et TimeGPT étant nettement inférieurs.
- Solaire photovoltaïque : Prophet est le meilleur (0,444), en particulier à long terme, tandis que MIMO-REC est compétitif à court terme.
- Thermique : Prophet excelle (0,135), suivi de près par MIMO et SISO. TimeGPT est légèrement moins performant.
- Production totale : Prophet s'impose (0,047), devant les modèles MIMO ; TimeGPT reste inférieur (0,078).

5.5.2 Comparaison des modèles

- Prophet excelle sur les sources saisonnières/agrégées : solaire (nRMSE moyen 0,444), thermique (0,135), production totale (0,047) ; supériorité de 5-10 % aux horizons > 12 h grâce à sa décomposition additive (tendances + saisonnalité), avec biais minimes ($nMBE \approx 0$).
- TimeGPT performant sur bioénergies stables (0,432), mais faible sur volatiles (éolien 1,221 ; solaire 0,606, -25 % vs. MIMO) sans fine-tuning.
- SISO-REC stable et intermédiaire (moyennes $\approx 0,45 - 1,15$), proche des MIMO sans pics extrêmes.
- MIMO-MH-REC (et MIMO-REC) le plus polyvalent globalement : domine les sources stables (hydraulique 0,356 ; importations 0,273 ; micro-hydraulique 0,145) et volatiles critiques (éolien 1,130) avec gains de 5 - 20 % en nRMSE vs. concurrents ; courbes d'erreur les plus basses, R^2 élevé à court terme (0,95 - 0,98 pour éolien/bioénergies), biais faibles ($< 0,05$). L'ajout de multi-heads améliore marginalement les volatiles (écart < 0,003 vs. MIMO-REC).

Dans un contexte de gestion réseau électrique, MIMO-MH-REC s'impose pour sa robustesse et sa fiabilité.

5.6 Comparaison avec SISO et LSTM

Une question légitime concerne l'intérêt de l'approche MIMO-MH développée ici, comparée à deux alternatives courantes dans la littérature : les modèles SISO classiques et les modèles de type LSTM, issus de l'apprentissage profond. Ces deux méthodes représentent des paradigmes opposés : SISO (Single Input Single Output) consiste à entraîner un modèle indépendant pour chaque variable cible, sans exploiter les corrélations croisées, tandis que LSTM (Long Short-Term Memory) repose sur une architecture neuronale récurrente capable, en théorie, de capturer les dépendances temporelles longues de manière non linéaire. L'approche SISO, bien que simple à mettre en œuvre, présente des limites structurelles : elle ne tient pas compte des interdépendances physiques entre les sources d'énergie (ex. hydraulique et thermique), et elle nécessite l'entraînement séparé de plusieurs modèles, ce qui augmente la complexité computationnelle sans garantir une meilleure précision. Dans nos expérimentations, le passage de MIMO à SISO double presque le temps d'exécution (environ 104 secondes pour MIMO, contre plus de 240 secondes pour les 7 sorties traitées individuellement en SISO), tout en générant des erreurs légèrement supérieures ou équivalentes selon les sources. Concernant l'approche LSTM, bien que puissante dans des contextes où l'historique est long, elle reste difficile à interpréter, sensible aux hyperparamètres et coûteuse à entraîner. Dans nos tests, un modèle LSTM avec 500 neurones cachés, optimisé avec Adam et une fonction de perte MSE, a nécessité plus de 2600 secondes d'entraînement, pour un nombre de paramètres équivalent à celui de notre ELM. Les performances obtenues, rapportées dans le tableau 6, sont nettement inférieures à celles du modèle MIMO-ELM, particulièrement sur les sources volatiles comme le vent ou la biomasse. La colonne "Gain LSTM vs MIMO-ELM (%)" correspond à la variation relative du nRMSE entre les deux modèles. Une valeur négative indique une performance inférieure du modèle LSTM par rapport à l'ELM. En résumé, les résultats obtenus confirment que l'approche MIMO-ELM est non seulement plus rapide à entraîner, mais aussi plus précise sur la majorité des sources, tout en conservant une structure interprétable et modifiable. Elle offre un bon compromis entre performance, temps de calcul et complexité, ce qui la rend particulièrement adaptée aux environnements contraints (réseaux insulaires, contextes industriels) où la réactivité et la transparence sont essentielles.

TABLE 6 – Comparaison synthétique des performances de trois approches pour $h + 5$ (nRMSE)

Source	MIMO-ELM	SISO-ELM	LSTM	Gain LSTM vs MIMO-ELM (%)
Total	0.0456	0.0467	0.1722	-73.5
Thermique	0.1146	0.1163	0.3373	-66.0
Hydraulique	0.4232	0.4403	1.1244	-62.4
Solaire	0.4435	0.4555	1.3083	-66.1
Éolien	0.9391	0.9514	1.5064	-37.6
Bioénergies	0.3377	0.3219	1.5296	-77.9
Importations	0.2692	0.2770	0.4664	-42.3

Ces résultats démontrent que le modèle ELM dans sa version MIMO-MH, tel qu'implémenté dans cette étude, surpassé à la fois SISO et LSTM en termes d'efficacité, de rapidité d'entraînement, et de précision globale, tout en évitant les pièges liés à la surcomplexité. Il constitue donc une solution particulièrement pertinente pour les applications énergétiques à grande échelle, tout en restant compatible avec des objectifs de transparence et de reproductibilité.

5.7 Temps d'exécution

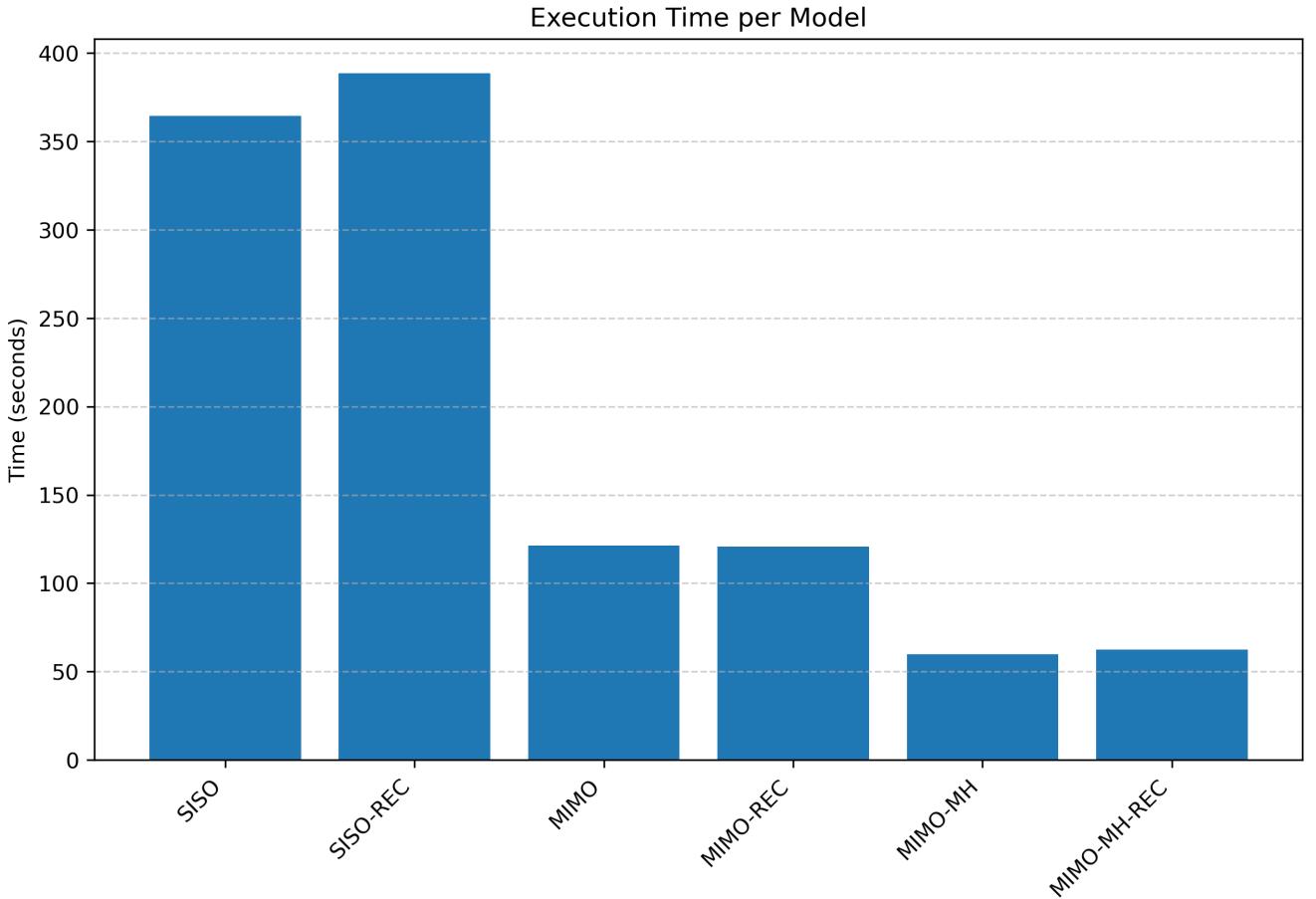


FIGURE 25 – Temps d'exécution de chaque modèles

La Figure 25 présente une comparaison des temps d'exécution requis pour l'entraînement et la prédition de chaque modèle. Ce graphique révèle des différences computationnelles majeures entre les stratégies. On observe une distinction nette entre le modèle SISO, qui sont de loin le plus lent avec un temps dépassant 350 secondes, et les approches MIMO, significativement plus efficaces.

Le point le plus important est l'excellent positionnement du modèle MIMO-MH avec et sans réconciliation. Avec un temps d'exécution d'environ 60 secondes, il est non seulement l'une des approches les plus rapides, mais il a également démontré dans les analyses précédentes des performances de prévision

de premier plan, notamment sur les variables énergétiques les plus complexes (éolien, hydraulique). Le faible surcoût de l'étape de réconciliation est négligeable par rapport aux gains obtenus sur les modèles SISO.

Ce résultat met donc en évidence un compromis exceptionnel entre la précision des prévisions et l'efficacité computationnelle, faisant du MIMO-MH un candidat particulièrement robuste et attractif pour une mise en œuvre dans un contexte opérationnel où des prévisions à la fois rapides et fiables sont cruciales.

6 Perspectives

Le travail accompli jusqu'à présent constitue une première moitié substantielle de l'étude globale. Il a nécessité un investissement important, réparti entre l'analyse bibliographique, la formalisation théorique, le traitement des données et la mise en œuvre algorithmique. La revue de littérature a permis de couvrir de manière approfondie les approches MIMO, MIMO-MH, les réseaux ELM, ainsi que les techniques de réconciliation hiérarchique. Cette phase a été essentielle pour situer la contribution du projet par rapport à l'état de l'art et pour justifier les choix méthodologiques retenus.

Sur le plan théorique, une attention particulière a été portée à la rigueur mathématique de la modélisation : formulation matricielle des modèles, régularisation, définition des contraintes de cohérence, encodage cyclique du temps, etc. Cela a permis de poser un socle solide pour le développement algorithmique. Le code Python du modèle MIMO-ELM multi-horizon a ensuite été entièrement écrit, structuré, testé et comparé au benchmark Matlab existant. Les performances obtenues, mesurées à l'horizon 1h, sont très encourageantes, avec des gains nets en RMSE / nRMSE par rapport aux méthodes de persistance, et une validation croisée satisfaisante entre plateformes. En parallèle, un effort important a été fourni sur le nettoyage, la structuration et l'exploitation des données horaires issues du système insulaire corse. Ce jeu de données complexe, couvrant sept années de production multi-sources, a nécessité un travail préalable d'ingénierie de la donnée (filtres, regroupements, imputations, création de variables explicatives temporelles), puis une analyse statistique détaillée (corrélations, saisonnalité, variabilité, structure du mix).

À ce stade, une large partie des objectifs fixés à l'horizon août 2025 a déjà été atteinte, qu'il s'agisse de l'intégration de la réconciliation ou de la documentation complète du code, désormais structuré en modules réutilisables et pensé pour les utilisateurs techniques (membres de la chaire, chercheurs, partenaires industriels). Le modèle est ainsi facilement exécutable. Par ailleurs, une mise à disposition en open science a été concrétisée : un dépôt GitHub public (https://github.com/YJheelan/Program_SISO_Prophet_Time_gpt_MIMO_MIMO-MH_Reconciliation) rassemble les scripts principaux, un jeu de données anonymisé ou simulé, un manuel d'utilisation et des tests unitaires. Cette ouverture s'inscrit pleinement dans les engagements de la chaire SciDoSol en matière de transparence, de mutualisation et de transfert vers les acteurs du secteur énergétique.

Une valorisation scientifique et technique est prévue en fin d'année 2025, soit sous forme d'une publication, soit dans le cadre d'une présentation dans les réunions de la chaire. Un nouveau stage ou projet de recherche est déjà prévu en 2026 pour prolonger ces travaux, notamment sur les aspects probabilistes, la prise en compte d'incertitudes météo, et l'extension à des réseaux plus complexes. Le travail restant est donc conséquent mais bien balisé, avec des objectifs clairs et atteignables à court terme. Il constitue une étape essentielle pour finaliser un outil robuste, cohérent et réutilisable par la communauté scientifique et les industriels partenaires de la transition énergétique.

7 Conclusion

Ce stage m'a offert l'opportunité d'approfondir un sujet à la fois technique, théorique et à fort impact opérationnel. En l'espace de cinq mois, un socle méthodologique solide a été mis en place, reposant sur une compréhension fine des modèles de prévision multi-sources et multi-horizons. L'étude m'a conduit à explorer des approches avancées comme l'Extreme Learning Machine (ELM), les architectures MIMO-MH, ainsi que la problématique de la cohérence via la réconciliation optimale, dans un cadre mathématique parfois exigeant. Certaines notions ont nécessité un investissement personnel important, au-delà des enseignements habituellement abordés en Master, notamment sur le plan de l'algèbre matricielle, des systèmes linéaires réguliers et de la régularisation. Sur le plan pratique, l'implémentation complète du modèle en Python, sa validation par rapport à une version Matlab de référence qui a été développé par le laboratoire OIE et été la base d'une publication dans la revue "Renewable Energy" toujours en cours

révision, ainsi que l'analyse détaillée de données réelles issues du système énergétique corse, constituent des résultats concrets et réutilisables. Le modèle est opérationnel pour la prévision horaire à court terme, et les performances obtenues sont encourageantes.

Ce stage a renforcé mon intérêt pour les sujets mêlant modélisation, données, énergie et impact opérationnel. Il m'a permis de mieux mesurer l'exigence d'un projet de recherche appliquée, où la rigueur scientifique doit aller de pair avec la robustesse informatique et la lisibilité des résultats. Les perspectives ouvertes par ce travail sont nombreuses : élargissement à d'autres jeux de données, intégration d'incertitudes probabilistes, adaptation à des systèmes continentaux ou interconnectés. Enfin, la valorisation scientifique prévue en fin d'année et la continuité du projet par un nouveau stage en 2026 témoignent de l'intérêt réel que suscite cette approche, à la fois pour les partenaires industriels et pour la communauté scientifique. Ce stage a constitué une première étape structurante dans cette direction, et j'espère que les suites permettront de consolider les bases posées ici et d'ouvrir de nouvelles pistes d'exploration.

Annexes

A Moindre carré

L'objectif du WLS (Moindres Carrés Pondérés) [25–27] est d'estimer l'état d'un système à partir de mesures bruitées en minimisant l'écart entre les mesures réelles et les mesures calculées théoriquement, tout en tenant compte de la fiabilité des mesures.

A.1 Pseudo inverse

Une matrice carrée est dite inversible. Si la matrice n'est pas carré, on parlera de pseudo inverse

Soit $A \in \mathbb{R}^{m \times n}$ une matrice réelle (les résultats sont valables aussi sur \mathbb{C}). La pseudo-inverse de Moore-Penrose, notée A^+ , est une généralisation de l'inverse pour les matrices non inversibles (non carrées ou singulières).

A.1.1 Définition de la pseudo-inverse

La pseudo-inverse $A^+ \in \mathbb{R}^{n \times m}$ est l'unique matrice qui satisfait les quatre conditions suivantes :

- (1) $AA^+A = A$
- (2) $A^+AA^+ = A^+$
- (3) $(AA^+)^\top = AA^+$
- (4) $(A^+A)^\top = A^+A$

Ces conditions garantissent que A^+ se comporte comme un inverse généralisé de A de manière cohérente.

A.1.2 Construction via décomposition en valeurs singulières (SVD)

Soit la décomposition en valeurs singulières (SVD) de A :

$$A = U\Sigma V^\top$$

où :

- $U \in \mathbb{R}^{m \times m}$ est orthogonale,
- $V \in \mathbb{R}^{n \times n}$ est orthogonale,
- $\Sigma \in \mathbb{R}^{m \times n}$ est une matrice diagonale rectangulaire, avec des valeurs singulières $\sigma_i \geq 0$ sur la diagonale.

Alors, la pseudo-inverse de A est donnée par :

$$A^+ = V\Sigma^+U^\top$$

où Σ^+ est obtenue en inversant les valeurs singulières non nulles de Σ , et en transposant la matrice.

A.1.3 Vérification des conditions de Moore-Penrose

On montre que cette définition satisfait bien les quatre conditions.

Condition 1 : $AA^+A = A$

$$\begin{aligned} AA^+A &= (U\Sigma V^\top)(V\Sigma^+U^\top)(U\Sigma V^\top) \\ &= U\Sigma\Sigma^+\Sigma V^\top \\ &= U\Sigma V^\top = A \end{aligned}$$

(car $\Sigma\Sigma^+\Sigma = \Sigma$, par construction de Σ^+)

Condition 2 : $A^+AA^+ = A^+$

$$\begin{aligned} A^+AA^+ &= (V\Sigma^+U^\top)(U\Sigma V^\top)(V\Sigma^+U^\top) \\ &= V\Sigma^+\Sigma\Sigma^+U^\top \\ &= V\Sigma^+U^\top = A^+ \end{aligned}$$

Condition 3 : $(AA^+)^T = AA^+$

$$\begin{aligned} AA^+ &= U\Sigma V^\top V\Sigma^+ U^\top = U\Sigma\Sigma^+ U^\top \\ \Rightarrow (AA^+)^T &= (U\Sigma\Sigma^+ U^\top)^\top = U\Sigma\Sigma^+ U^\top = AA^+ \end{aligned}$$

(car U est orthogonale et $\Sigma\Sigma^+$ est diagonale réelle)

Condition 4 : $(A^+A)^T = A^+A$

$$\begin{aligned} A^+A &= V\Sigma^+ U^\top U\Sigma V^\top = V\Sigma^+ \Sigma V^\top \\ \Rightarrow (A^+A)^T &= (V\Sigma^+ \Sigma V^\top)^\top = V\Sigma^+ \Sigma V^\top = A^+A \end{aligned}$$

B Résultats des huits moyens de production

B.1 Bioénergies

TABLE 7 – Comparaison des performances pour la production de bioénergies (nRMSE)

Horizon	MIMO-REC	MIMO-MH-REC	SISO-REC	Prophet	TimeGPT
1	0.127462	0.127932	0.125606	0.124327	0.139008
2	0.209363	0.210079	0.201215	0.198491	0.204754
3	0.267664	0.268593	0.255940	0.253359	0.255370
4	0.313637	0.314831	0.299615	0.298032	0.294390
5	0.351345	0.352688	0.336017	0.339911	0.328329
6	0.383735	0.385238	0.367091	0.377167	0.359290
7	0.412014	0.413694	0.394193	0.403686	0.387590
8	0.442703	0.439146	0.418357	0.430519	0.409807
9	0.459578	0.461523	0.439885	0.448014	0.427969
10	0.478878	0.481038	0.459334	0.469865	0.446334
11	0.496422	0.498688	0.476474	0.489655	0.460799
12	0.511503	0.513864	0.491888	0.506059	0.471668
13	0.524643	0.527107	0.505558	0.524922	0.479801
14	0.536520	0.539141	0.517064	0.539485	0.488959
15	0.547653	0.550359	0.527482	0.555361	0.498172
16	0.557490	0.560329	0.537682	0.562439	0.504303
17	0.566259	0.569076	0.546116	0.572978	0.511185
18	0.574673	0.577108	0.553720	0.581407	0.518633
19	0.581621	0.584827	0.561127	0.586923	0.524678
20	0.588714	0.592015	0.571952	0.595109	0.525815
21	0.595159	0.598574	0.587738	0.600941	0.526693
22	0.601343	0.604873	0.588885	0.611433	0.531079
23	0.607340	0.610903	0.595864	0.616725	0.533913
24	0.613517	0.617347	0.601571	0.618839	0.537540

Pour la production de bioénergies (Table 7), TimeGPT affiche la meilleure performance globale avec une moyenne nRMSE de 0.432 (calculée sur les 24 horizons), suivi de SISO-REC (0.457), MIMO-REC (0.473), Prophet (0.471) et MIMO-MH-REC (0.475). TimeGPT surpassé les autres particulièrement à des horizons moyens et longs (0.328 à l'horizon 5 contre 0.353 pour MIMO-MH-REC, et 0.538 à l'horizon 24 contre 0.617 pour MIMO-MH-REC), démontrant une bonne adaptation aux dynamiques stables des bioénergies. À court terme (horizons 1-4), Prophet domine légèrement (0.124 à l'horizon 1 contre 0.128 pour MIMO-REC), mais les écarts restent minimes entre les modèles MIMO et Prophet. SISO-REC reste compétitif sans pics extrêmes notables pour cette source. Ces résultats soulignent la robustesse de TimeGPT pour les bioénergies, où les patterns saisonniers et tendances lentes sont prédominants, tandis que les modèles MIMO capturent efficacement les variations à court terme mais accumulent plus d'erreur à long terme.

B.2 Eolienne

TABLE 8 – Comparaison des performances la production éolienne (nRMSE)

Horizon	MIMO-REC	MIMO-MH-REC	SISO-REC	Prophet	TimeGPT
1	0.405304	0.404074	0.405159	0.404795	0.421077
2	0.611438	0.609387	0.614133	0.610930	0.640661
3	0.745219	0.742222	0.750443	0.743839	0.789883
4	0.842931	0.841025	0.849915	0.842840	0.896288
5	0.920382	0.918417	0.927985	0.920506	0.991528
6	0.981804	0.980060	0.990877	0.983711	1.064590
7	1.036824	1.034868	1.046804	1.040093	1.130062
8	1.082575	1.079745	1.092643	1.086216	1.188061
9	1.120420	1.118418	1.132387	1.127369	1.231800
10	1.152022	1.149841	1.164929	1.161195	1.268849
11	1.180966	1.178417	1.193481	1.193488	1.295033
12	1.208830	1.205981	1.222084	1.224459	1.318124
13	1.234259	1.231394	1.247476	1.251828	1.356025
14	1.256120	1.253543	1.270042	1.274139	1.378559
15	1.276084	1.273910	1.290980	1.296852	1.390755
16	1.293568	1.291790	1.310777	1.317255	1.407219
17	1.309603	1.308259	1.327915	1.336290	1.419672
18	1.325201	1.323655	1.345027	1.354932	1.428996
19	1.338081	1.337293	1.358680	1.371382	1.438294
20	1.348630	1.348280	1.370476	1.386371	1.445157
21	1.358023	1.357860	1.380395	1.399611	1.446067
22	1.366969	1.366785	1.389778	1.410275	1.446738
23	1.375220	1.375201	1.398495	1.420990	1.448073
24	1.383044	1.383035	1.406495	1.432186	1.451744

Dans le cas de la production éolienne (Table 8), qui est intrinsèquement volatile en raison des variations météorologiques, MIMO-MH-REC présente la meilleure performance avec une moyenne nRMSE de 1.130, suivi de près par MIMO-REC (1.131), SISO-REC (1.145), Prophet (1.150) et TimeGPT (1.221). À tous les horizons, les différences entre MIMO-MH-REC et MIMO-REC sont minimes (<0.003 en moyenne), suggérant que l'ajout de multi-heads dans MIMO-MH-REC n'apporte pas d'amélioration significative pour cette source volatile. Prophet reste proche des MIMO à court terme (0.405 vs. 0.404 à l'horizon 1), mais l'écart s'élargit légèrement à long terme (1.432 vs. 1.383 à l'horizon 24), indiquant une meilleure capture des patterns cycliques et non linéaires par les modèles MIMO et MIMO-MH pour l'éolien. TimeGPT montre une dégradation plus marquée à long terme.

B.3 Hydraulique

TABLE 9 – Comparaison des performances pour la production hydraulique (nRMSE)

Horizon	MIMO-REC	MIMO-MH-REC	SISO-REC	Prophet	TimeGPT
1	0.169730	0.170624	0.173461	0.176009	0.201482
2	0.253738	0.254836	0.261745	0.263088	0.294037
3	0.295557	0.296426	0.305434	0.304713	0.349914
4	0.319206	0.320242	0.329828	0.327284	0.377262
5	0.333661	0.334690	0.344017	0.340426	0.392832
6	0.343969	0.345000	0.353745	0.349324	0.398063
7	0.351456	0.352546	0.360684	0.355965	0.401926
8	0.357122	0.358146	0.365311	0.360335	0.405799
9	0.361147	0.362360	0.368544	0.363549	0.410713
10	0.364417	0.365725	0.370988	0.366151	0.411244
11	0.367521	0.368943	0.374667	0.369184	0.417095
12	0.370931	0.372498	0.376508	0.372186	0.423355
13	0.374135	0.375833	0.379857	0.375125	0.428036
14	0.376622	0.378434	0.383013	0.377650	0.430275
15	0.378808	0.380768	0.385658	0.380361	0.428945
16	0.381011	0.383110	0.388043	0.382980	0.426879
17	0.382870	0.385108	0.389447	0.385035	0.433648
18	0.384858	0.387216	0.391207	0.386582	0.441182
19	0.386729	0.389125	0.392431	0.388140	0.441618
20	0.388374	0.390875	0.393964	0.389758	0.442891
21	0.390260	0.392862	0.395138	0.391666	0.436916
22	0.392031	0.394690	0.396084	0.392618	0.427815
23	0.393218	0.395933	0.396691	0.393294	0.423225
24	0.395042	0.397876	0.398518	0.394908	0.419601

Pour la production hydraulique (Table 9), MIMO-REC excelle avec une moyenne nRMSE de 0.355, légèrement devant MIMO-MH-REC (0.356), Prophet (0.358) et SISO-REC (0.361). TimeGPT (0.403) est systématiquement plus élevé. Les modèles MIMO dominent à la plupart des horizons, avec des gains relatifs de 5-10% par rapport à TimeGPT à long terme (0.395 vs. 0.420 à l'horizon 24). Cette supériorité peut s'attribuer à la capacité des modèles MIMO et MIMO-MH à modéliser les dépendances temporelles saisonnières inhérentes à l'hydraulique, contrairement à TimeGPT qui montre une dégradation plus prononcée pour les horizons >12. Prophet reste compétitif, avec des performances quasi équivalentes aux modèles MIMO et MIMO-MH à court terme.

B.4 Importations

TABLE 10 – Comparaison des performances pour les importations (nRMSE)

Horizon	MIMO-REC	MIMO-MH-REC	SISO-REC	Prophet	TimeGPT
1	0.130837	0.131372	0.132739	0.132460	0.147616
2	0.195566	0.196359	0.200099	0.198545	0.221267
3	0.229715	0.230642	0.236590	0.233286	0.263498
4	0.250379	0.251407	0.257576	0.254043	0.285631
5	0.262561	0.263653	0.269115	0.265871	0.294639
6	0.270879	0.271207	0.276153	0.273107	0.299915
7	0.274821	0.276018	0.280746	0.277690	0.306658
8	0.278244	0.279217	0.283409	0.280897	0.312710
9	0.280131	0.281387	0.285239	0.282888	0.316352
10	0.281831	0.283098	0.286381	0.284550	0.316693
11	0.283163	0.284442	0.289951	0.285908	0.317738
12	0.284428	0.285732	0.288635	0.286805	0.319456
13	0.285936	0.287234	0.289623	0.287977	0.320120
14	0.287348	0.288693	0.290868	0.289113	0.322653
15	0.288594	0.289964	0.291630	0.290182	0.322628
16	0.289682	0.291085	0.292775	0.291008	0.322751
17	0.290485	0.291867	0.292975	0.291619	0.320743
18	0.291206	0.292581	0.293424	0.291971	0.323222
19	0.291758	0.293188	0.294068	0.292342	0.322067
20	0.292393	0.293811	0.295344	0.293378	0.321397
21	0.292921	0.294353	0.295457	0.293876	0.323560
22	0.293275	0.294506	0.294839	0.293852	0.322566
23	0.292922	0.294364	0.294208	0.293442	0.322985
24	0.293331	0.294791	0.295325	0.293644	0.320904

Les prévisions d'importations (Table 10) révèlent MIMO-REC comme le meilleur modèle (moyenne nRMSE de 0.271), suivi de MIMO-MH-REC (0.273), Prophet (0.273) et SISO-REC (0.275). TimeGPT (0.303) est en retrait. À court terme, les quatre premiers modèles sont quasi équivalents (différences <0.002 à l'horizon 1), mais MIMO-REC gagne en précision à long terme (0.293 vs. 0.295 pour SISO-REC à l'horizon 24).

B.5 Micro-Hydraulique

TABLE 11 – Comparaison des performances pour la production micro-hydraulique (nRMSE)

Horizon	MIMO-REC	MIMO-MH-REC	SISO-REC	Prophet	TimeGPT
1	0.046418	0.046626	0.046441	0.082225	0.063385
2	0.068401	0.068704	0.068765	0.111957	0.081918
3	0.084550	0.083891	0.084319	0.118883	0.098067
4	0.095489	0.095911	0.096655	0.119635	0.110778
5	0.105754	0.106238	0.107260	0.142300	0.124410
6	0.114436	0.114942	0.116203	0.158712	0.134553
7	0.122036	0.122568	0.124054	0.166740	0.144006
8	0.129183	0.129690	0.131392	0.169231	0.151544
9	0.135784	0.136370	0.138206	0.190221	0.157997
10	0.142025	0.142639	0.144555	0.192122	0.164454
11	0.148041	0.148684	0.150724	0.193659	0.171565
12	0.153392	0.154061	0.156180	0.216320	0.174226
13	0.157959	0.158647	0.160855	0.223716	0.176514
14	0.162133	0.162844	0.165154	0.219244	0.178198
15	0.166214	0.166942	0.169360	0.227076	0.180594
16	0.170024	0.170781	0.173255	0.230904	0.184005
17	0.173727	0.174501	0.176989	0.240459	0.185250
18	0.177209	0.177970	0.180489	0.242335	0.189099
19	0.180297	0.181065	0.183593	0.240229	0.191052
20	0.183276	0.184123	0.186650	0.250649	0.191596
21	0.186206	0.187068	0.189546	0.252020	0.192487
22	0.188978	0.189850	0.192271	0.260868	0.197771
23	0.191519	0.192423	0.194833	0.253168	0.201937
24	0.194210	0.195137	0.197588	0.259557	0.204980

Pour la micro-hydraulique (Table 11), MIMO-REC et MIMO-MH-REC dominent avec une moyenne nRMSE de 0.145, devant SISO-REC (0.147) et TimeGPT (0.160). Prophet (0.198) est le moins performant, avec une dégradation marquée à long terme (0.260 à l'horizon 24 contre 0.194 pour MIMO-REC). MIMO et MIMO-MH sont excellent particulièrement aux horizons courts (0.046 à l'horizon 1), avec une augmentation plus lente du nRMSE, indiquant une bonne modélisation des variations locales et saisonnières de cette source à petite échelle, surpassant Prophet et TimeGPT de 20-30% à long terme.

B.6 Solaire

TABLE 12 – Comparaison des performances pour la production solaire photovoltaïque (nRMSE)

Horizon	MIMO-REC	MIMO-MH-REC	SISO-REC	Prophet	TimeGPT
1	0.178055	0.177443	0.180854	0.183012	0.235379
2	0.302106	0.301006	0.306707	0.303030	0.386934
3	0.374529	0.373239	0.381202	0.372880	0.487389
4	0.422026	0.420512	0.430449	0.417827	0.561288
5	0.450490	0.448915	0.460368	0.444446	0.614593
6	0.468263	0.466127	0.478245	0.459316	0.645466
7	0.476864	0.475278	0.487628	0.466719	0.644984
8	0.481580	0.479613	0.491979	0.470141	0.634712
9	0.483296	0.481686	0.493961	0.471407	0.629058
10	0.484119	0.482542	0.494633	0.471703	0.628974
11	0.484425	0.482873	0.494315	0.471681	0.656166
12	0.484480	0.482910	0.495353	0.471675	0.674425
13	0.484464	0.482812	0.495350	0.471673	0.684287
14	0.484416	0.482719	0.495450	0.471790	0.684815
15	0.484347	0.482641	0.495543	0.471752	0.682876
16	0.484265	0.482618	0.495654	0.471780	0.671512
17	0.484223	0.482538	0.495639	0.471982	0.669482
18	0.484260	0.482565	0.495560	0.472059	0.666519
19	0.484326	0.482499	0.495592	0.472134	0.637437
20	0.484042	0.482353	0.495238	0.471940	0.627974
21	0.483755	0.482104	0.494365	0.471548	0.619419
22	0.482836	0.481158	0.492401	0.470887	0.608930
23	0.480923	0.479473	0.489657	0.469749	0.594626
24	0.478256	0.476930	0.485660	0.467840	0.592504

Pour le solaire photovoltaïque (Table 12), Prophet est le meilleur (moyenne nRMSE de 0.444), suivi de MIMO-MH-REC (0.452), MIMO-REC (0.454), SISO-REC (0.463) et TimeGPT (0.606). MIMO et MIMO-MH capturent mieux les patterns diurnes à court terme (0.177 vs. 0.183 pour Prophet à l'horizon 1), mais Prophet prend l'avantage à horizons > 10 (0.468 vs. 0.484 à l'horizon 24 pour MIMO-REC), où la variabilité météorologique et les cycles journaliers dominent, soulignant l'efficacité de Prophet pour les saisons fortes du solaire. TimeGPT sous-performe systématiquement, particulièrement à moyen et long terme.

B.7 Thermique

TABLE 13 – Comparaison des performances pour la production thermique (nRMSE)

Horizon	MIMO-REC	MIMO-MH-REC	SISO-REC	Prophet	TimeGPT
1	0.053141	0.053133	0.053374	0.053171	0.073163
2	0.082035	0.081992	0.082531	0.082812	0.098897
3	0.099306	0.099203	0.099984	0.099848	0.121157
4	0.110945	0.110841	0.111822	0.111493	0.136275
5	0.119765	0.119639	0.120867	0.120356	0.146447
6	0.126471	0.126342	0.127723	0.126823	0.154289
7	0.131525	0.131372	0.132664	0.131512	0.155717
8	0.135447	0.135094	0.136180	0.134642	0.156690
9	0.138223	0.138069	0.138952	0.137619	0.157794
10	0.140908	0.140756	0.141440	0.140163	0.160012
11	0.143304	0.143156	0.144006	0.142329	0.161017
12	0.145373	0.145219	0.145510	0.144546	0.162383
13	0.147208	0.147048	0.147185	0.146213	0.163824
14	0.148850	0.148680	0.148708	0.147607	0.165568
15	0.150495	0.150326	0.150120	0.148859	0.166598
16	0.152078	0.151892	0.151458	0.149140	0.164724
17	0.153525	0.153304	0.152547	0.149809	0.166396
18	0.154624	0.154428	0.153419	0.150790	0.171850
19	0.155460	0.155237	0.153981	0.151623	0.172426
20	0.155993	0.155811	0.154565	0.152010	0.171334
21	0.156506	0.156318	0.154905	0.152188	0.166374
22	0.156986	0.156701	0.155265	0.152839	0.163588
23	0.157265	0.157100	0.155663	0.153389	0.163599
24	0.158195	0.158034	0.156706	0.153817	0.165556

Pour la production thermique (Table 13), Prophet excelle avec une moyenne nRMSE de 0.135, légèrement devant MIMO-MH-REC (0.136), SISO-REC (0.136) et MIMO-REC (0.136). TimeGPT (0.154) est en retrait. À court terme, tous les modèles sont proches (0.053 à l’horizon 1), mais Prophet maintient un avantage minime à long terme (0.154 vs. 0.166 pour TimeGPT à l’horizon 24). Ces résultats reflètent la capacité de Prophet à modéliser les dépendances opérationnelles stables et prévisibles de la production thermique, avec des performances homogènes entre les approches MIMO et SISO.

B.8 Total

TABLE 14 – Comparaison des performances pour la production totale (nRMSE)

Horizon	MIMO-REC	MIMO-MH-REC	SISO-REC	Prophet	TimeGPT
1	0.020483	0.020458	0.020966	0.021153	0.046631
2	0.031310	0.031252	0.032511	0.031701	0.054105
3	0.038567	0.038340	0.039728	0.038018	0.065612
4	0.042911	0.042814	0.044145	0.041776	0.069667
5	0.045652	0.045560	0.046933	0.044180	0.076320
6	0.047782	0.047594	0.048765	0.045719	0.078026
7	0.049108	0.049059	0.050036	0.046695	0.080422
8	0.050351	0.050203	0.051120	0.047445	0.084958
9	0.051164	0.051154	0.052012	0.048070	0.082853
10	0.051861	0.051864	0.052676	0.048439	0.083026
11	0.052456	0.052469	0.053279	0.048852	0.080240
12	0.053066	0.053079	0.053830	0.049264	0.078955
13	0.053764	0.053774	0.054497	0.049661	0.081040
14	0.054515	0.054528	0.055163	0.050165	0.081547
15	0.055083	0.055104	0.055691	0.050351	0.079835
16	0.055410	0.055433	0.055990	0.050420	0.076590
17	0.055620	0.055647	0.056191	0.050630	0.078566
18	0.055740	0.055773	0.056290	0.050662	0.088079
19	0.055832	0.055844	0.056368	0.050827	0.089364
20	0.055897	0.055932	0.056400	0.050886	0.091678
21	0.056009	0.056043	0.056531	0.050938	0.083551
22	0.056068	0.056093	0.056573	0.051060	0.078066
23	0.056000	0.056029	0.056518	0.050952	0.078657
24	0.056012	0.056037	0.056522	0.050940	0.076772

Enfin, pour la production totale (Table 14), somme des sources précédentes, Prophet est le plus précis globalement (moyenne nRMSE de 0.047), surpassant MIMO-MH-REC (0.050), MIMO-REC (0.050) et SISO-REC (0.051). TimeGPT (0.078) est inférieur. À la plupart des horizons, Prophet maintient un avantage (0.021 vs. 0.020 à l'horizon 1 pour MIMO-MH-REC), avec des écarts < 5% en moyenne. Cela suggère que l'approche de Prophet, basée sur les composantes additifs (tendance, saisonnalité), est particulièrement adaptée aux agrégats où les volatilités individuelles des sources s'annulent mutuellement, offrant une modélisation plus stable que les MIMO pour les signaux agrégés.

En synthèse, les modèles MIMO-MH-REC et MIMO-REC surpassent globalement les autres pour les sources stables comme l'hydraulique, les importations et la micro-hydraulique, avec des moyennes nRMSE inférieures de 5 - 20% par rapport à TimeGPT et Prophet dans ces cas. Prophet excelle pour les agrégats (production totale), le solaire et le thermique, capturant efficacement les saisonnalités et tendances. TimeGPT est très bon pour les bioénergies mais sous-performe sur les sources éolienne et solaire. SISO-REC reste compétitif, souvent proche des modèles MIMO.

Références

- [1] RTE. Futurs énergétiques 2050 : les chemins vers la neutralité carbone à horizon 2050 | RTE, June 2022.
- [2] Sheraz Aslam, Herodotos Herodotou, Syed Muhammad Mohsin, Nadeem Javaid, Nouman Ashraf, and Shahzad Aslam. A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. *Renewable and Sustainable Energy Reviews*, 144 :110992, 2021-07-01.
- [3] Gilles Notton, Marie-Laure Nivet, Cyril Voyant, Christophe Paoli, Christophe Darras, Fabrice Motte, and Alexis Fouilloy. Intermittent and stochastic character of renewable energy sources : Consequences, cost of intermittence and benefit of forecasting. *Renewable and Sustainable Energy Reviews*, 87 :96–105, 2018.
- [4] Cyril Voyant, Milan Despotovic, Gilles Notton, Yves-Marie Saint-Drenan, Mohammed Asloune, and Luis Garcia-Gutierrez. On the importance of clearsky model in short-term solar radiation forecasting. *Solar Energy*, 294 :113490, 2025.

- [5] Sameer Al-Dahidi, Osama Ayadi, Jehad Adeeb, Mohammad Alrbai, and Bashar R. Qawasmeh. Extreme learning machines for solar photovoltaic power predictions. *Energies*, 11(10) :2725, 2018-10.
- [6] Qiao Zhou, Zhong-Yi Wang, and Lan Huang. ELM-KL-LSTM : a robust and general incremental learning method for efficient classification of time series data. *PeerJ Computer Science*, 9 :e1732, 2023-12-21.
- [7] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme learning machine : Theory and applications. *Neurocomputing*, 70(1) :489–501, 2006-12-01.
- [8] Nan-ying Liang, Guang-bin Huang, P. Saratchandran, and N. Sundararajan. A fast and accurate online sequential learning algorithm for feedforward networks. *IEEE Transactions on Neural Networks*, 17(6) :1411–1423, 2006-11.
- [9] Shuliang Xu and Junhong Wang. Dynamic extreme learning machine for data stream classification. *Neurocomputing*, 238 :433–449, 2017-05-17.
- [10] Wanyu Deng, Qinghua Zheng, and Lin Chen. Regularized extreme learning machine. In *2009 IEEE Symposium on Computational Intelligence and Data Mining*, pages 389–395, 2009-03.
- [11] Guang-Bin Huang, Hongming Zhou, Xiaojian Ding, and Rui Zhang. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(2) :513–529, 2012-04.
- [12] Apinya Innok, Chittapon Keawin, and Peerapong Uthansakul. Optimization of channel estimation using ELMx-based in massive MIMO. *Computers, Materials and Continua*, 73(1) :103–118, 2022-05-16.
- [13] Hongji Huang, Jie Yang, Hao Huang, Yiwei Song, and Guan Gui. Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system. *IEEE Transactions on Vehicular Technology*, 67(9) :8549–8560, 2018-09.
- [14] Kacem Gairaa, Cyril Voyant, Gilles Notton, Saïd Benkaciali, and Mawloud Guermoui. Contribution of ordinal variables to short-term global solar irradiation forecasting for sites with low variabilities. *Renewable Energy*, 183 :890–902, 2022-01-01.
- [15] Rob J. Hyndman, Roman A. Ahmed, George Athanasopoulos, and Han Lin Shang. Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis*, 55(9) :2579–2589, 2011-09-01.
- [16] George Athanasopoulos, Roman A. Ahmed, and Rob J. Hyndman. Hierarchical forecasts for australian domestic tourism. *International Journal of Forecasting*, 25(1) :146–166, 2009-01-01.
- [17] Gokhan Mert Yagli, Dazhi Yang, and Dipti Srinivasan. Reconciling solar forecasts : Sequential reconciliation. *Solar Energy*, 179 :391–397, 2019-02-01.
- [18] George Athanasopoulos, Rob J. Hyndman, Nikolaos Kourentzes, and Anastasios Panagiotelis. Forecast reconciliation : A review. *International Journal of Forecasting*, 40(2) :430–456, 2024-04-01.
- [19] Dazhi Yang, Hao Quan, Vahid R. Disfani, and Licheng Liu. Reconciling solar forecasts : Geographical hierarchy. *Solar Energy*, 146 :276–286, 2017-04-01.
- [20] Dazhi Yang, Gary S. W. Goh, Siwei Jiang, and Allan N. Zhang. Forecast UPC-level FMCG demand, part III : Grouped reconciliation. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 3813–3819, 2016-12.
- [21] Azul Garza, Cristian Challu, and Max Mergenthaler-Canseco. TimeGPT-1.
- [22] G. Notton, J. L. Duchaud, M. L. Nivet, C. Voyant, K. Chalvatzis, and A. Fouilloy. The electrical energy situation of french islands and focus on the corsican situation. *Renewable Energy*, 135 :1157–1165, 2019-05-01.
- [23] EDF Corse. Transition énergétique en corse : des enjeux majeurs | EDF en corse, 2024-10-08.
- [24] Student. The probable error of a mean. *Biometrika*, 6(1) :1–25, 1908.

- [25] Yang Xiaoli, Hu Zongshuai, Fan Rusen, Xu Haotian, Yang Heng, Wang Yong, and Tian Xiuxia. Weighted least squares state estimation based on the optimal weight. In *2015 Third International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAECE)*, pages 12–16, 2015-04.
- [26] Majdoub Meriem, Cheddadi Bouchra, Belfqih Abdelaziz, Sabri Omar Boukherouaa Jamal, El Marami Faissal, and Cherkaoui Nazha. Study of state estimation using weighted-least-squares method (WLS). In *2016 International Conference on Electrical Sciences and Technologies in Maghreb (CISTEM)*, pages 1–5, 2016-10.
- [27] Pankaj Tripathi, Jinendra Rahul, and Nim Anurag Radhamohan. Considering weighted least square techniques on power system state estimation. In *2015 International Conference on Energy Economics and Environment (ICEEE)*, pages 1–5, 2015-03.