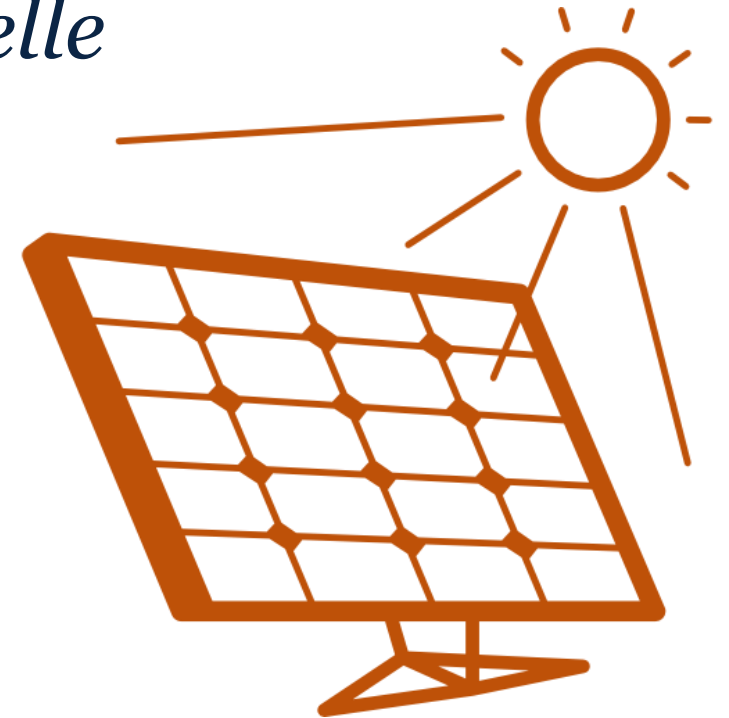


Prédiction Photovoltaïque

Multi-Horizon par Intelligence Artificielle



Inasse LAHMAR

Sous la supervision du **Pr. Yacine Yaddaden**

Département de mathématiques, informatique et génie _ UQAR
Département Green Tech Institute _ Université Mohammed VI Polytechnique _ UM6P

Plan de la présentation

01 Contexte et problématique

Intermittence des énergies renouvelables et enjeux des réseaux

02 Objectifs du projet

Simuler, prédire et comparer

03 Simulation MATLAB/Simulink

Centrale photovoltaïque de 1 MW connectée au réseau

04 Base de données générée

65 241 observations horaires

05 Pipeline et modèles d'intelligence artificielle

*3 modèles d'apprentissage automatique (Machine Learning **ML**) + 3 modèles d'apprentissage profond (Deep Learning **DL**)*

06 Résultats multi-horizons

t+1h, t+3h, t+6h et t+11h

07 Conclusion et Synthèse

Problématique

Pourquoi prédire la production photovoltaïque ?

! Le défi de l'intermittence

L'intégration massive des énergies renouvelables pose un défi majeur aux réseaux électriques : leur **intermittence naturelle**.



La production solaire dépend fortement des conditions météorologiques (irradiance, nuages, température, etc.), ce qui rend difficile l'équilibre instantané entre offre et demande sur le réseau.

Une prévision précise devient donc **essentielle** pour assurer la stabilité des **réseaux électriques intelligents** (smart grids) et optimiser leur planification.



Déséquilibre offre/demande

Instabilités de fréquence et de tension



Gestion complexe du réseau

Réserves tournantes coûteuses



Intégration d'énergie renouvelable limitée

Frein au déploiement à grande échelle



Planification difficile

Peu de visibilité sur la production future

Objectifs

01



Simuler

Une centrale solaire PV 1 MW à Marrakech

Modéliser sous MATLAB/Simulink une centrale solaire photovoltaïque de 1 MW connectée au réseau ONEE (30 kV / 50 Hz) afin de générer une base de données de production horaire réaliste.



02



Prédire et comparer

Six modèles d'intelligence artificielle AI

Développer et comparer six modèles d'intelligence artificielle, trois d'apprentissage automatique **ML** et trois d'apprentissage profond **DL**, pour prédire la puissance AC produite.



03



Explorer

Données de simulation vs Données de terrain

Explorer le potentiel des données simulées comme alternative aux données de terrain, coûteuses et souvent incomplètes, pour l'entraînement des modèles prédictifs.



Approche méthodologique

Un pipeline en cinq étapes



Pourquoi Marrakech ?

Marrakech bénéficie d'environ 3 000 heures d'ensoleillement par an et d'une irradiance globale horizontale comprise entre 2 100 et 2 200 kWh/m²/an → Un site de référence pour le solaire photovoltaïque au Maroc.

Simulation MATLAB/Simulink

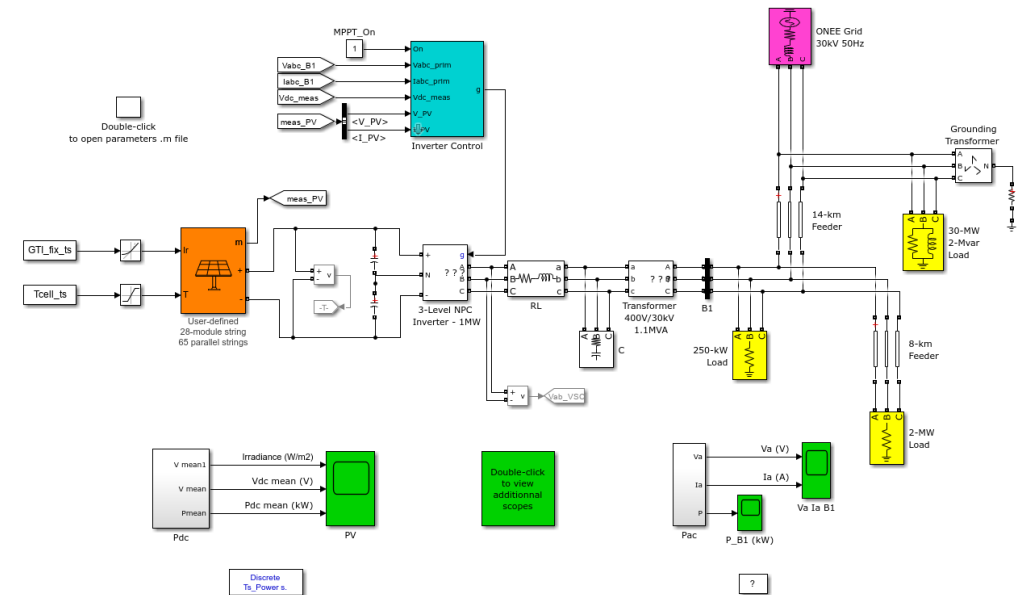
Une centrale solaire PV de 1 MW



Architecture de la centrale

- **Modules PV** LONGi LR5-72HPH-550M (550 Wc)
- **Configuration** 28 en série × 65 strings en parallèle
- **Puissance totale** 1 MW crête
- **Modèle électrique** Une diode, calibré STC + NOCT
- **Erreur Pmax** Inférieure à 5 %
- **Onduleur** NPC 3 niveaux avec contrôle MPPT
- **Injection réseau** ONEE — 30 kV / 50 Hz

Schéma Simulink



Le modèle à une diode est calibré aux conditions STC et NOCT, avec une erreur inférieure à 5 % sur la puissance maximale. L'onduleur NPC 3 niveaux et la stratégie MPPT assurent une injection optimale dans le réseau électrique.

- STC: Standard Test Conditions.
- NOCT: Nominal Operating Cell Temperature.

Base de données de production

14 variables d'entrée, une cible : la puissance AC produite



65 241

Observations horaires



14

Variables d'entrée



16 ans

Période couverte (2010–2026)



9h–19h

Plage horaire de production



Variables d'entrée

- **Rayonnement** GHI, DNI, DHI
- **Température** Température ambiante, humidité relative
- **Atmosphère** Pression, couverture nuageuse, indice poussière
- **Vent** Vitesse et direction du vent
- **Temporelles** Mois, jour, heure, variables cycliques



Prétraitement & séparation

- **Filtrage** Heures de production seulement (9h–19h)
- **Nettoyage** Suppression des valeurs aberrantes et manquantes
- **Normalisation** StandardScaler (fitté sur le train uniquement)
- **Split** Chronologique — 80 % / 10 % / 10 %

Variable cible : Puissance AC produite (P_ac en MW)

Six modèles d'intelligence artificielle

Trois modèles d'apprentissage automatique *ML*, trois d'apprentissage profonde *DL*



Apprentissage automatique ML

● Random Forest

Agrégation d'arbres décorrélés (bagging). Robuste au bruit et facile à interpréter.

● XGBoost

Gradient boosting optimisé. Référence sur données tabulaires.

● LightGBM

Boosting par feuilles (leaf-wise). Très rapide et performant sur gros volumes.



Apprentissage profonde DL

● LSTM

Long Short-Term Memory. Capture les dépendances longues grâce à ses portes (input, forget, output).

● BiLSTM

LSTM bidirectionnelle. Parcourt la séquence dans les deux sens pour un contexte plus riche.

● GRU

Gated Recurrent Unit. Architecture plus légère que LSTM, souvent aussi performante.

Évaluation multi-horizon

Horizons de prédiction évalués



t + 1 h

Très court terme

Ajustement en temps quasi réel



t + 3 h

Court terme

Planification intrajournalière



t + 6 h

Moyen terme

Gestion demi-journée



t + 11 h

Long terme

Planification du jour suivant

Métriques de performance

RMSE

Root Mean Square Error

$$\sqrt{(1/n \cdot \sum (\hat{y}_i - y_i)^2)}$$

Erreur quadratique moyenne en MW. Pénalise fortement les grosses erreurs.
Plus faible = meilleur.

R²

Coefficient de détermination

$$1 - \sum_i (y_i - \hat{y}_i)^2 / \sum_i (y_i - \bar{y})^2$$

Proportion de variance expliquée par le modèle. Compris entre 0 et 1 ;
1 = ajustement parfait.

Résultats — tableau comparatif

RMSE (MW) et R^2 pour les six modèles sur les quatre horizons

Modèle	t+1h RMSE	t+1h R^2	t+3h RMSE	t+3h R^2	t+6h RMSE	t+6h R^2	t+11h RMSE	t+11h R^2
Random Forest	0.0362	0.9547	0.0527	0.9146	0.0411	0.9270	0.0111	0.6131
XGBoost	0.0346	0.9586	0.0530	0.9137	0.0410	0.9275	0.0088	0.7580
LightGBM	0.0344	0.9590	0.0527	0.9142	0.0411	0.9269	0.0092	0.7327
LSTM	0.0365	0.9551	0.0528	0.9059	0.0649	0.8576	0.0712	0.8286
BiLSTM	0.0361	0.9559	0.0542	0.9009	0.0658	0.8537	0.0701	0.8338
GRU	0.0354	0.9576	0.0529	0.9054	0.0639	0.8620	0.0697	0.8358

HORIZONS COURTS

LightGBM

$R^2 = 0.9590$ à t+1h

HORIZONS MOYENS

XGBoost

$R^2 = 0.9275$ à t+6h

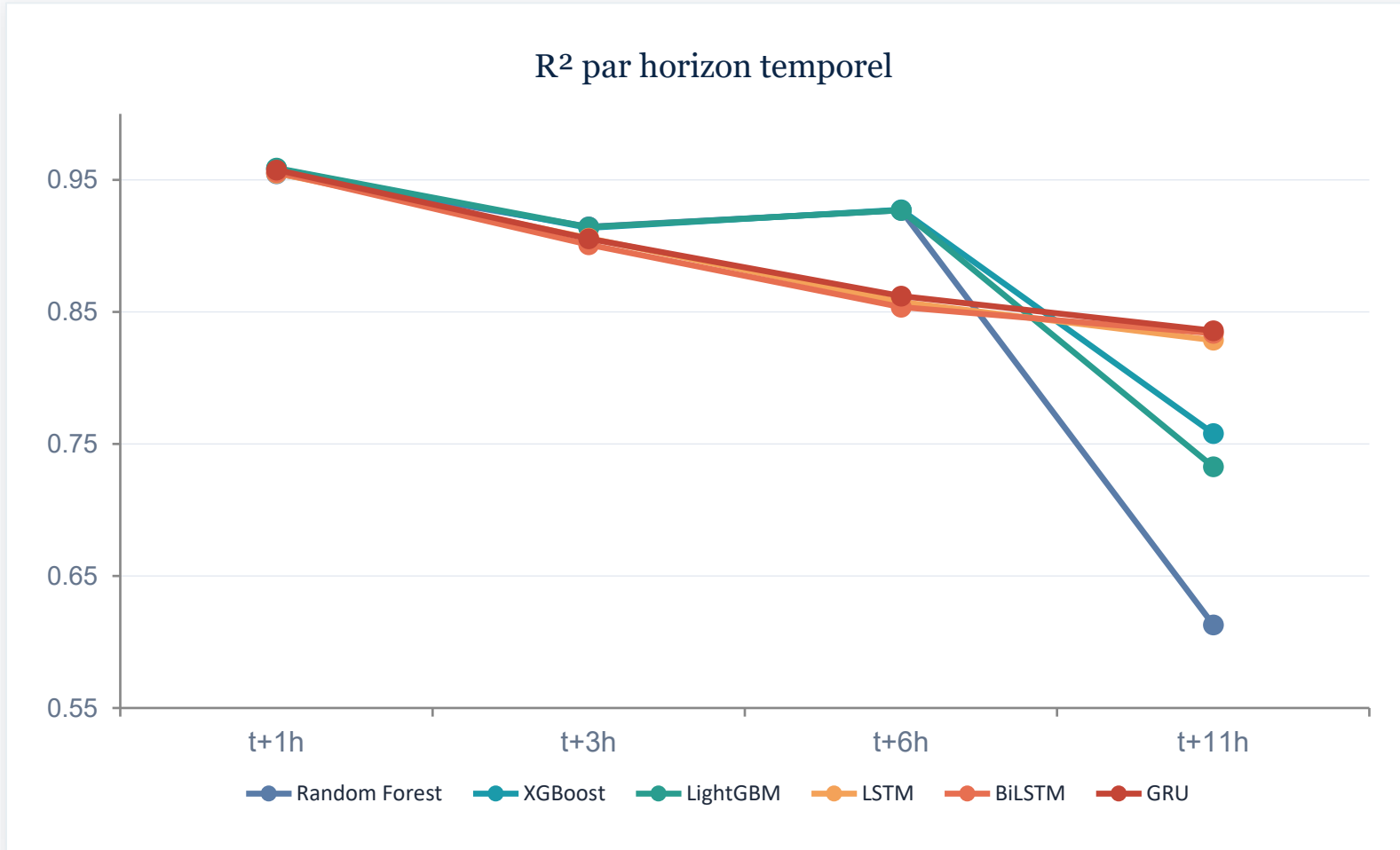
HORIZONS LONGS

GRU

$R^2 = 0.8358$ à t+11h

Évolution du R^2 par horizon

Deux dynamiques distinctes : la dégradation des modèles *ML* vs la résilience des modèles *DL*



LECTURE

Deux régimes

1 Chute brutale du ML à t+11h

Random Forest chute à $R^2=0.61$: les modèles à arbres peinent sur les horizons très longs.

2 Résilience du DL

LSTM, BiLSTM et GRU gardent $R^2 > 0.82$ à t+11h grâce à leur mémoire temporelle.

3 Complémentarité

ML pour le court terme, DL pour le long terme : l'hybridation est prometteuse.

Analyse des résultats



Pourquoi le ML domine à court terme ?

Sur les horizons t+1h à t+6h, la production suit de près les variables météo instantanées (GHI, température). Les arbres de décision capturent très bien ces relations non-linéaires et peuvent s'appuyer sur des valeurs récentes quasi identiques à la cible.



Pourquoi le ML s'effondre à t+11h ?

Random Forest, XGBoost et LightGBM n'ont aucune mémoire temporelle explicite. À t+11h, l'information instantanée est trop éloignée de la cible et le modèle n'a pas de mécanisme pour suivre la dynamique journalière.



La force des réseaux récurrents

LSTM, BiLSTM et GRU disposent de cellules mémoire permettant de capturer des dépendances longues sur 33 pas (3 jours). Cette mémoire leur permet de maintenir $R^2 > 0.82$ même à t+11h, là où le ML chute.



GRU : le meilleur compromis long-terme

Avec un $R^2=0.8358$ à t+11h, GRU s'impose sur les horizons longs. Son architecture plus légère que LSTM lui permet de mieux généraliser et de réduire les risques de surapprentissage sur ce jeu de données.

Conclusion

6

modèles

Comparés sur une même base

Trois algorithmes d'apprentissage automatique **ML** (Random Forest, XGBoost, LightGBM) et trois d'apprentissage profond **DL** (LSTM, BiLSTM, GRU) évalués de façon équitable sur quatre horizons de prédiction.

0.959

R^2 max à $t+1h$

LightGBM

Le meilleur score global est obtenu par LightGBM sur l'horizon le plus court. Sur les horizons courts et moyens, l'apprentissage automatique ML reste la voie la plus efficace.

0.836

R^2 à $t+11h$

GRU

À $t+11h$, GRU s'impose nettement grâce à sa capacité à capturer les dépendances temporelles. L'apprentissage profond devient incontournable dès que l'horizon dépasse quelques heures.

La complémentarité entre apprentissage automatique et apprentissage profond est la clé d'une prévision photovoltaïque robuste sur tous les horizons.

MERCI

de votre attention

Inasse LAHMAR et Pr. Yacine Yaddaden