

PROJET D'ANALYSE SOUS R

# **ANALYSE STATISTIQUE DE LA PRODUCTION MENSUELLE D'ÉLECTRICITÉ EN TUNISIE (2005–2024)**

Projet réalisé par Ines Elouaer, dans le cadre du cours assuré par le Professeur  
Abdallah Khmaies, au titre de l'année universitaire 2024–2025.

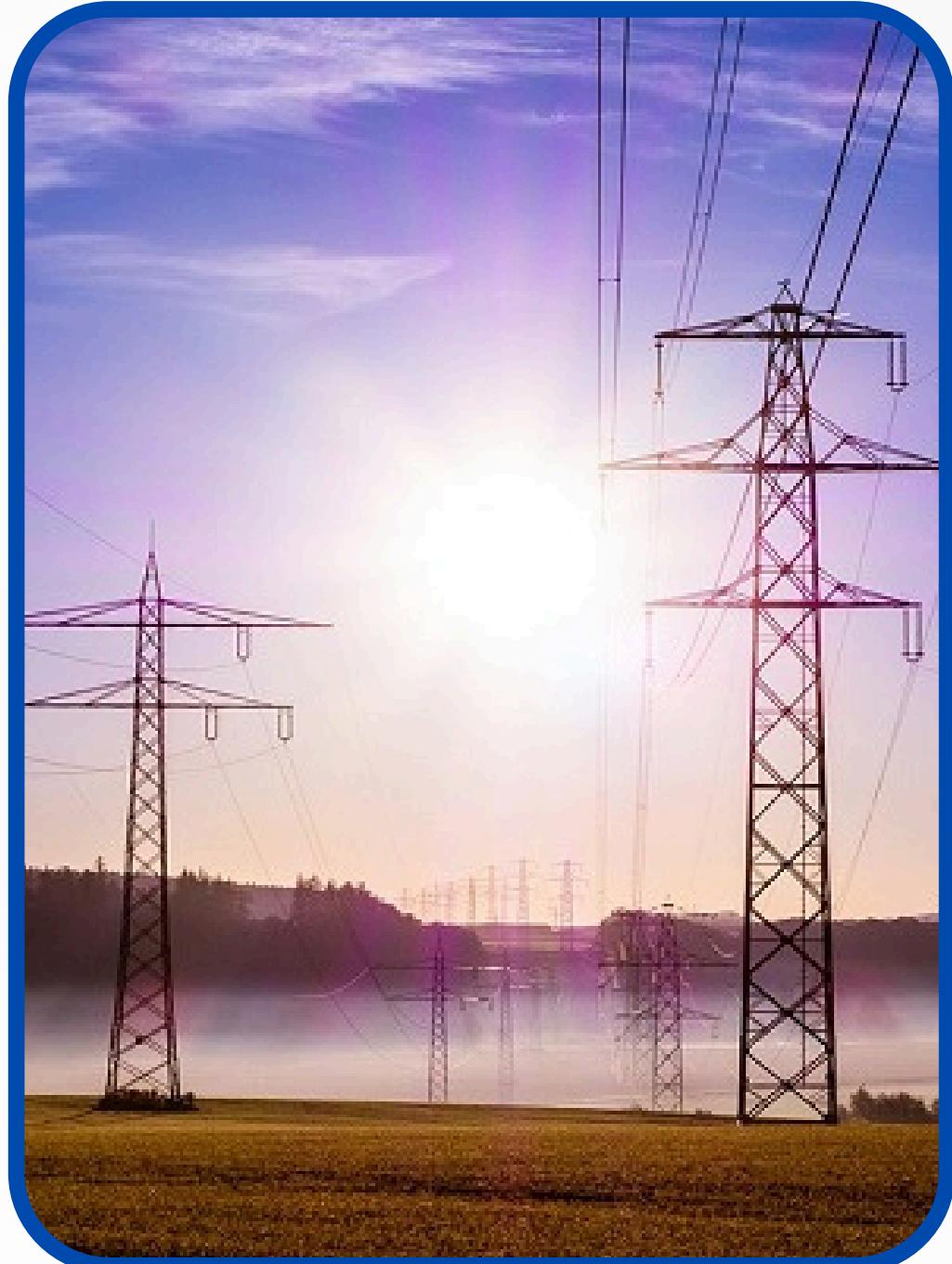
# CONTEXTE & MOTIVATION

## CONTEXTE

- L'électricité est un secteur stratégique pour l'économie.
- Des données publiques mensuelles permettent d'analyser son évolution.
- La production repose sur plusieurs acteurs (STEG, IPP, solaire).

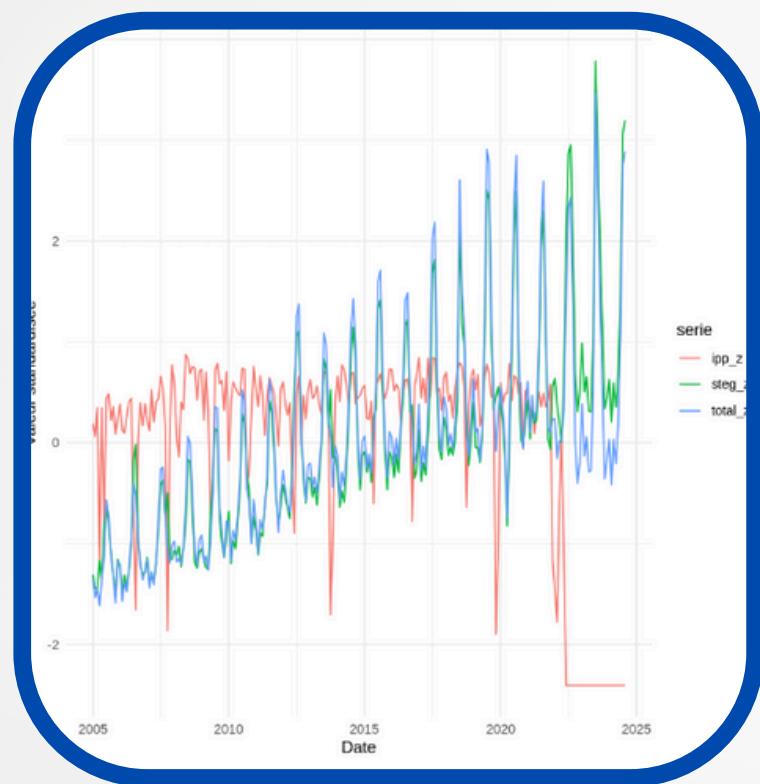
## MOTIVATION

- Étudier les tendances et la saisonnalité.
- Comparer les contributions des producteurs.



# OBJECTIFS DU PROJET

Ce projet vise à analyser l'évolution de la production d'électricité en Tunisie sur la période étudiée, en comparant les différents types de producteurs. Il cherche à mettre en évidence les tendances générales, les effets de saisonnalité et les différences de niveaux de production entre la STEG et les producteurs indépendants. L'analyse repose sur des méthodes statistiques exploratoires et des modèles simples et interprétables, dans une démarche de compréhension des données.



DONNÉES PUBLIQUES (TUNISIE)

# DESCRIPTION DES DONNÉES

236

20 ANS

5

Observations mensuelles  
(2005 – 2024)

Période d'étude

Variables principales :  
STEG,  
IPP,  
solaire,  
auto-producteurs ,  
production totale (en GWh)

# DÉMARCHE DU PROJET

- Inspection des données
- Nettoyage & préparation
- Gestion des dates & NA
- Transformation
- Analyse descriptive & graphiques
- Tests statistiques
- Modélisation & diagnostic

- 1
- 2
- 3
- 4
- 5
- 6
- 7

Le projet suit une démarche classique d'analyse de données : nettoyage, transformation en format tidy, analyse descriptive , tests statistiques et modélisation. Cette approche garantit des résultats cohérents et interprétables.

# NETTOYAGE DES DONNÉES

Les valeurs manquantes observées dans les variables solaire et auto-producteurs correspondent à des périodes où ces sources n'étaient pas encore présentes.

Ces NA ont donc été conservés volontairement, car ils ont une signification temporelle et ne traduisent pas une erreur de mesure.

1.  
**RENOMMAGE  
DES  
VARIABLES**

noms courts et explicites

2.  
**CONVERSION  
DES DATES**

formats mixtes → format Date

3.  
**VÉRIFICATION  
DES  
DOUBLONS**

aucun doublon détecté

4.  
**ANALYSE DES  
VALEURS  
MANQUANTES**

identification par variable

# TRANSFORMATION DES DONNÉES

Les données initiales étaient organisées en format large, avec une colonne par producteur. Afin de faciliter les visualisations, les comparaisons et les analyses statistiques sous R, les données ont été transformées en format long (tidy data). Cette structure est plus cohérente, standardisée et adaptée aux bonnes pratiques de l'analyse de données.

Format initial (Wide)

Date	STEG	IPP	Solaire	Sauto-producteur
Jan 2024	1,120	920	14	-
Feb 2024	975	789	14	-
Mar 2024	1,182	841	26	30
Apr 2024	1,131	900	28	35

Format transformé (Long / Tidy)

Date	Producteur	Production
Jan 2024	STEG	1,120 GWh
Feb 2024	IPP	920 GWh
Feb 2024	Solaire	975 GWh
Mar 2024	Solaire	789 GWh
Mar 2024	Solaire	14 GWh
Apr 2024	Sauto-producteur	30 GWh

# ANALYSE DESCRIPTIVE ET EXPLORATOIRE DES DONNÉES

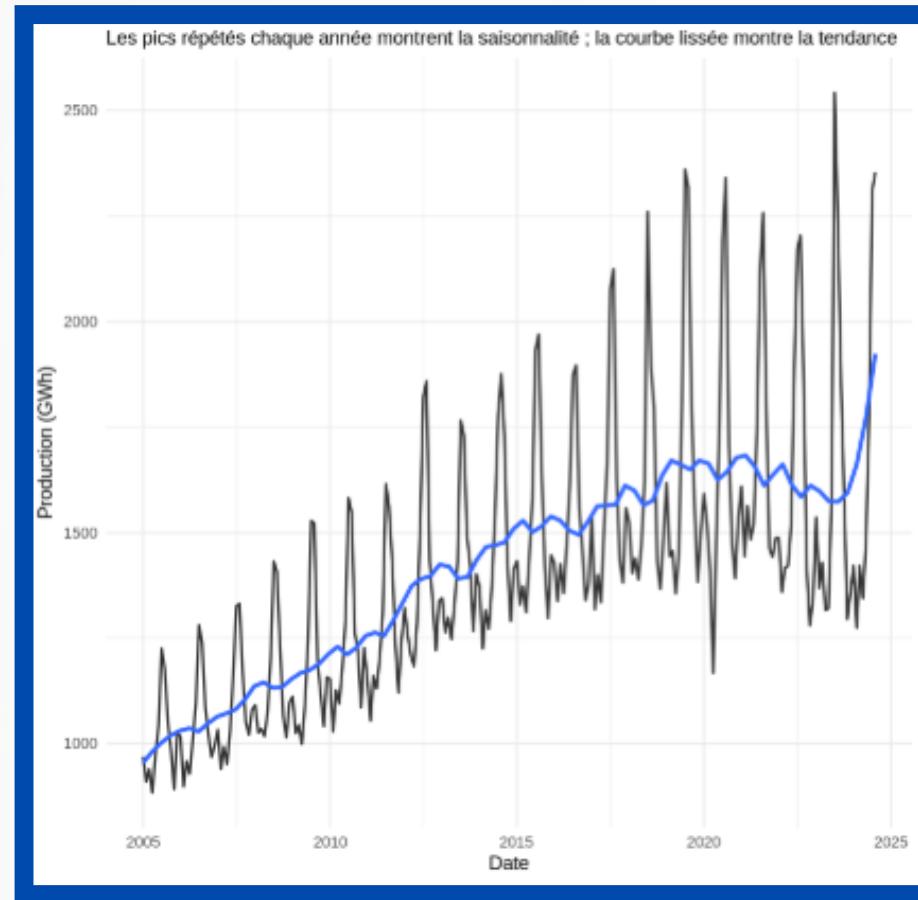


Figure 1 – Évolution de la production totale d'électricité

On observe une croissance progressive de la production, surtout marquée après 2010, avec des pics estivaux récurrents.

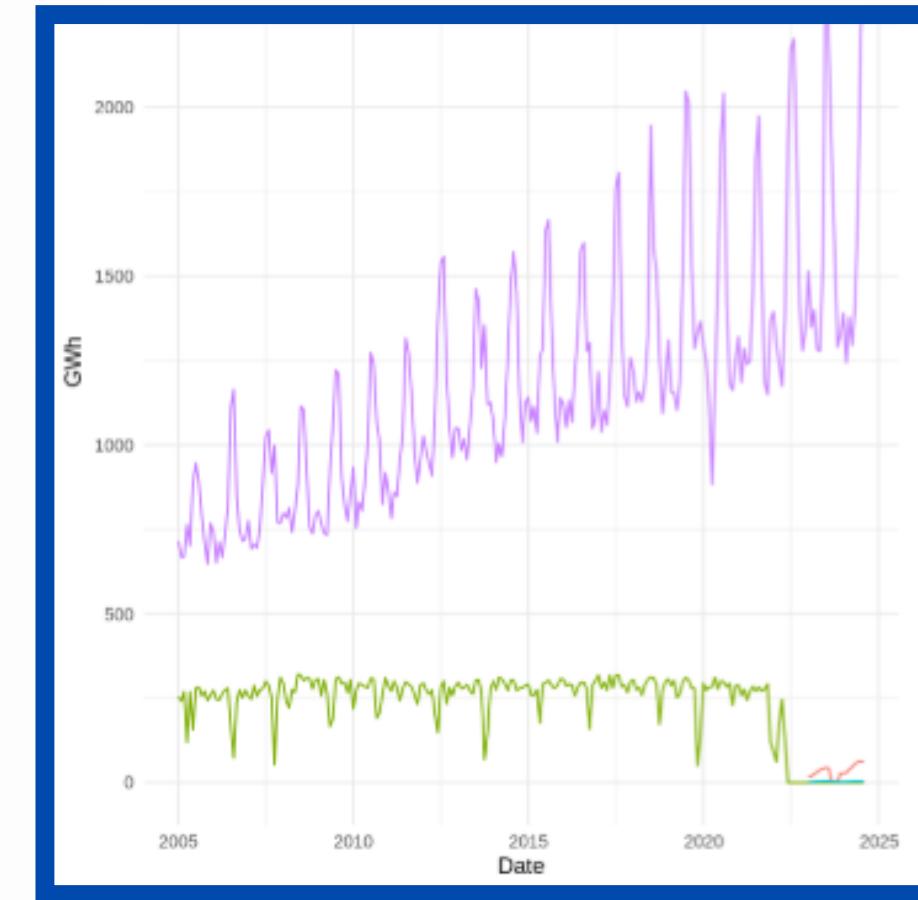


Figure 2 – Production mensuelle par type de producteur

Cette figure montre clairement la différence d'échelle entre les producteurs et l'émergence récente du solaire.

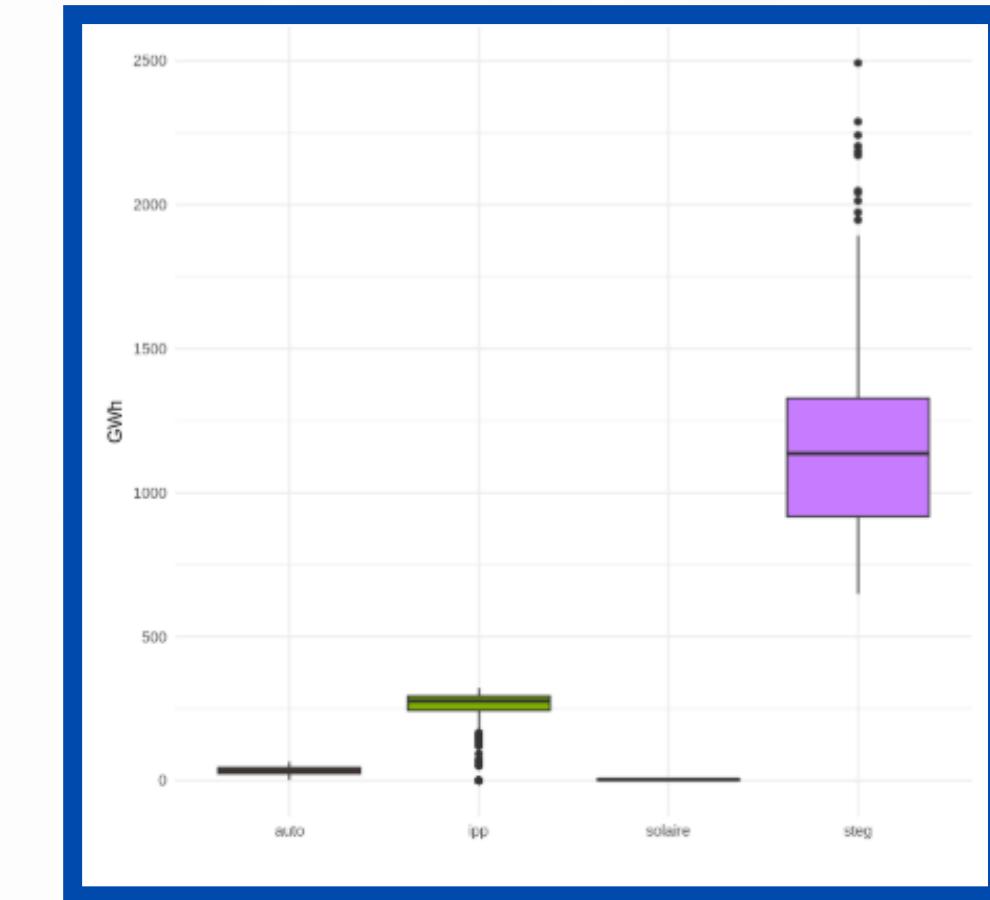


Figure 3 – Distribution de la production par producteur (Boxplot)

Les outliers ne sont pas des erreurs mais correspondent à des périodes de forte demande.

# ANALYSE STATISTIQUE COMPARAISON STEG VS IPP

Question statistique :

- La production mensuelle de la STEG est-elle significativement différente de celle des IPP ?

Méthodes utilisées :

- Un test de normalité (Shapiro-Wilk) a été réalisé, suivi d'un test t apparié afin de comparer les productions mensuelles des deux producteurs.

Résultats principaux :

- Différence hautement significative
- p-value < 0,001
- STEG produit en moyenne  $\approx 936$  GWh de plus que les IPP

→ Conclusion : la STEG domine statistiquement la production.

# MODÉLISATION & DIAGNOSTIC DU MODÈLE

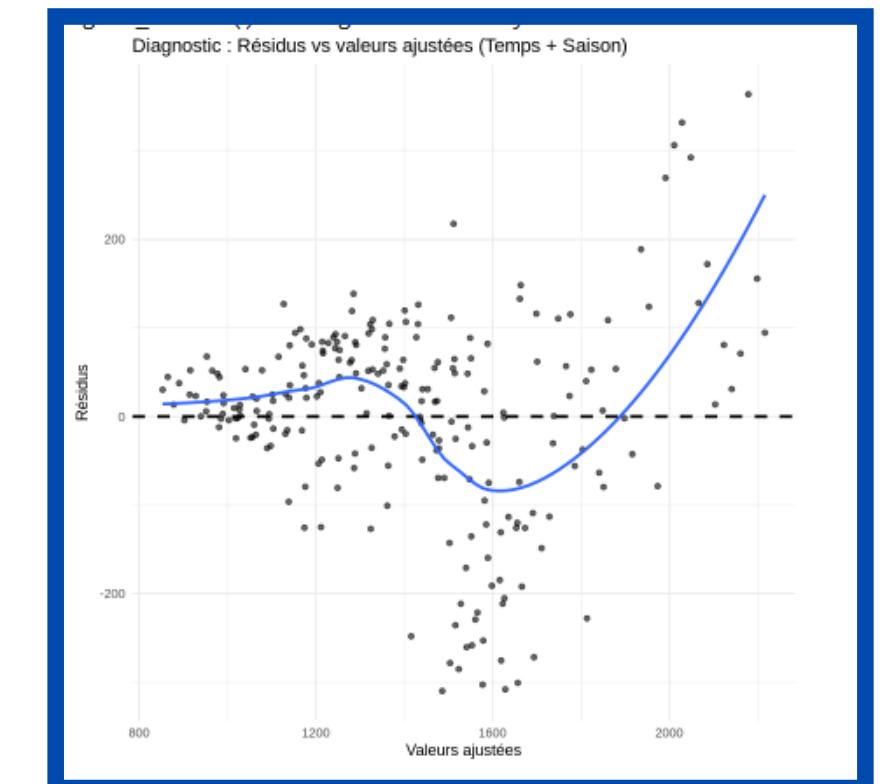
## modélisation :

Plusieurs modèles de régression ont été testés pour analyser l'évolution de la production totale d'électricité. Le modèle combinant la tendance temporelle et la saisonnalité a été retenu, car il offre le meilleur compromis entre performance statistique et interprétation simple.

modele	R2	R2_ajuste
<chr>	<dbl>	<dbl>
Temps	0.4431654	0.4407858
Temps + Saison	0.8792726	0.8727761

L'analyse des résidus montre qu'ils sont globalement centrés autour de zéro. Malgré une légère non-linéarité, le modèle reste acceptable pour une analyse exploratoire.

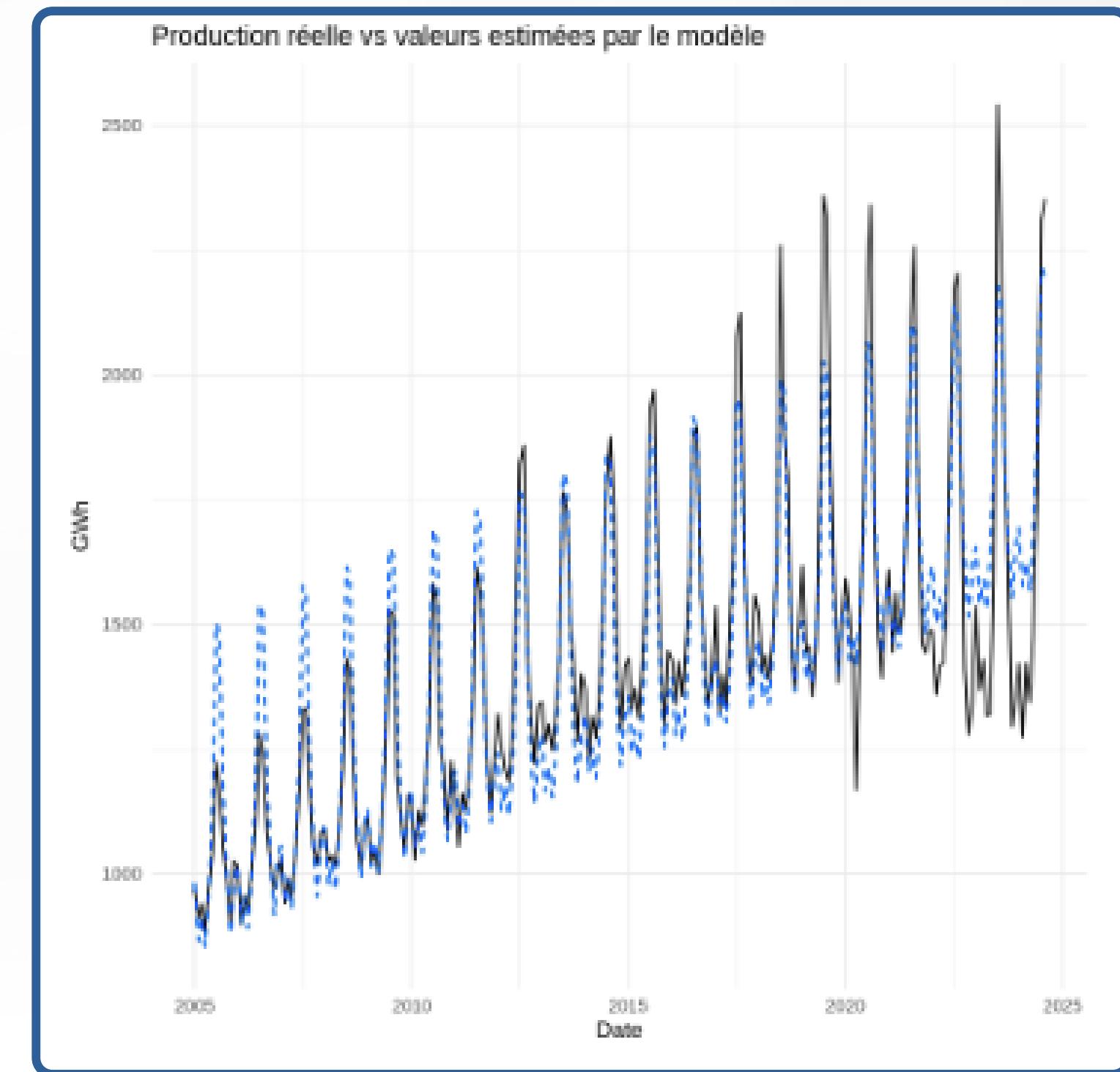
La comparaison des modèles montre une amélioration significative du pouvoir explicatif lorsque la saisonnalité est intégrée.



# RÉEL VS VALEURS ESTIMÉES

## Observation

- Le modèle capte bien la tendance et la saisonnalité
- Certains pics extrêmes restent difficiles à expliquer



# OUR COMPANY VISION

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore

1

2

3

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore

# CONCLUSION

L'analyse montre une augmentation globale de la production d'électricité en Tunisie, avec une forte saisonnalité et une prédominance de la STEG. La production solaire apparaît récemment et reste encore limitée mais en progression. Les modèles statistiques confirment ces tendances et offrent une bonne base pour des analyses futures plus avancées.

## LIMITES

- Données mensuelles agrégées
- Peu de variables explicatives externes
- Production solaire récente

## PERSPECTIVES

- Intégrer le climat ou la consommation
- Modèles de séries temporelles (ARIMA)
- Analyse plus fine par région



**MERCI !**