



Analyse des Données Météorologiques en France

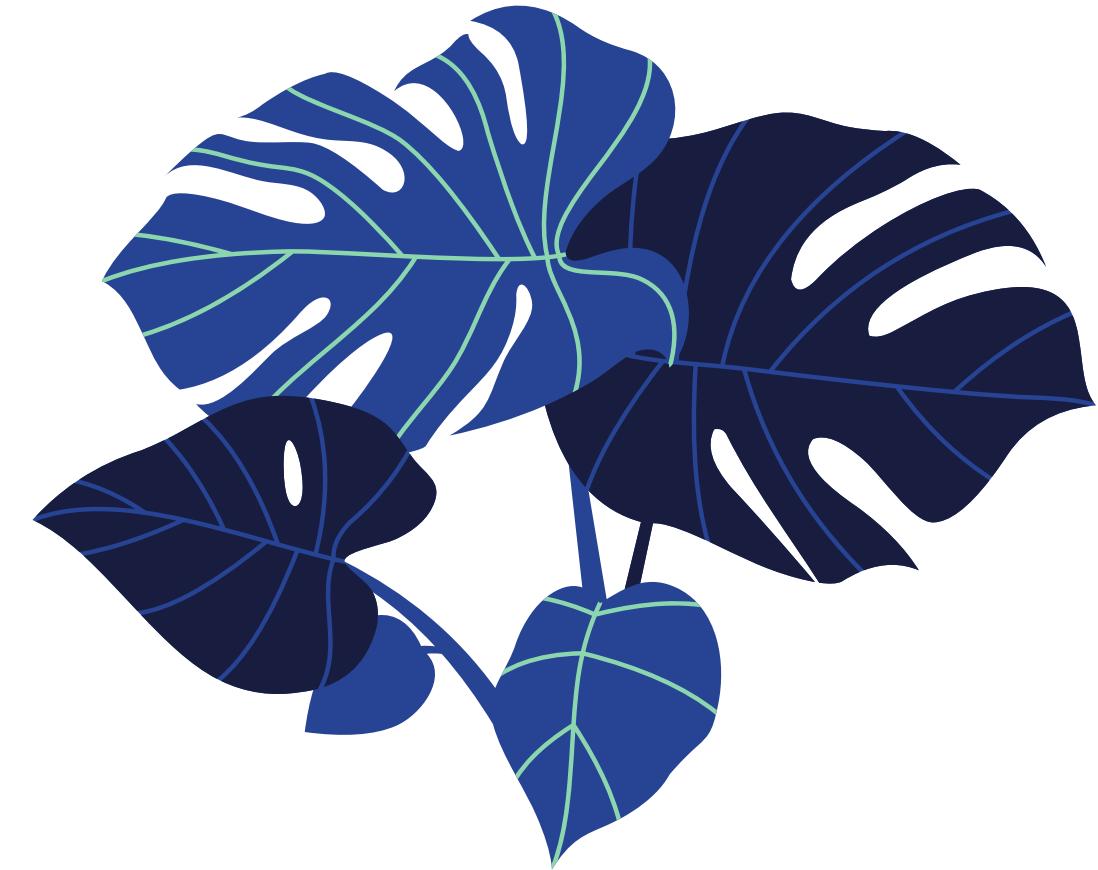
Groupe G10

Yuefan LIU, Mouzheng LI et Lianghong LI

Sommaire



1. Introduction
2. Statistiques descriptives
3. Analyse ACP
4. Régression linéaire simple
5. Régression linéaire multivariée
6. Conclusion



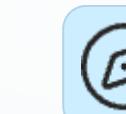
Introduction au Projet

Ce projet vise à explorer un ensemble de données météorologiques de l'année 2024, mesurées dans 25 villes françaises. L'objectif principal est de prédire la température à Paris pour janvier 2025, puis d'étendre la prévision aux mois de février à avril 2025. Nous avons utilisé deux méthodes principales pour cette étude : l'Analyse en Composantes Principales (ACP) ainsi que la régression linéaire simple et multiple pour établir des modèles prédictifs.



Description des Données

Analyse préliminaire pour comprendre la structure et les caractéristiques des données.



Identification des Variables Principales

Utilisation de l'Analyse en Composantes Principales (ACP) pour réduire la dimensionnalité.



Construction de Modèles Prédic

Application de la régression linéaire simple et multivariée pour les prévisions.

Statistiques Descriptives

Nous avons d'abord vérifié l'état des données, confirmant l'absence de valeurs manquantes sur les 25 villes. Cette étape cruciale a permis d'identifier des contrastes climatiques marqués entre les villes.

Extrêmes Climatiques

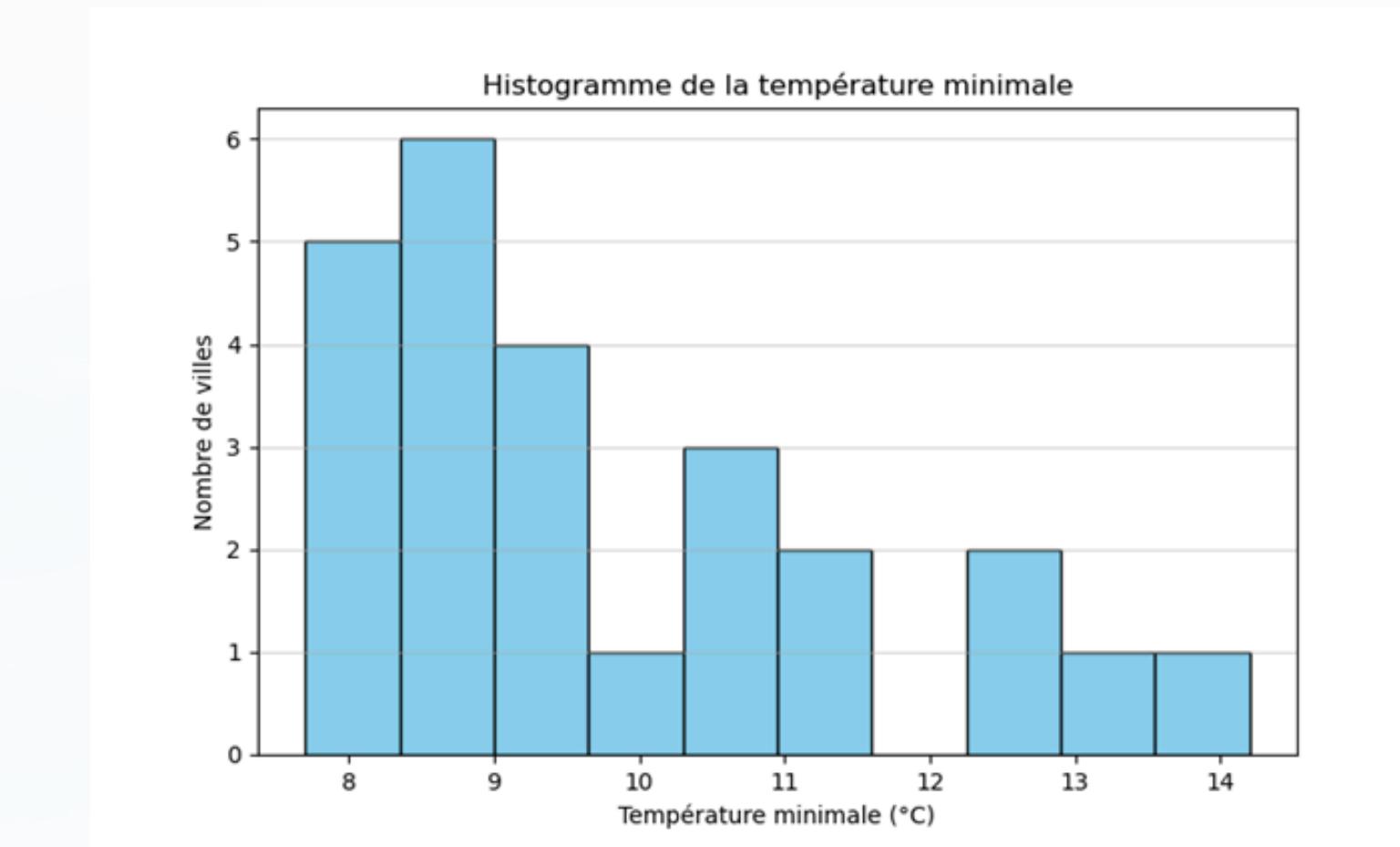
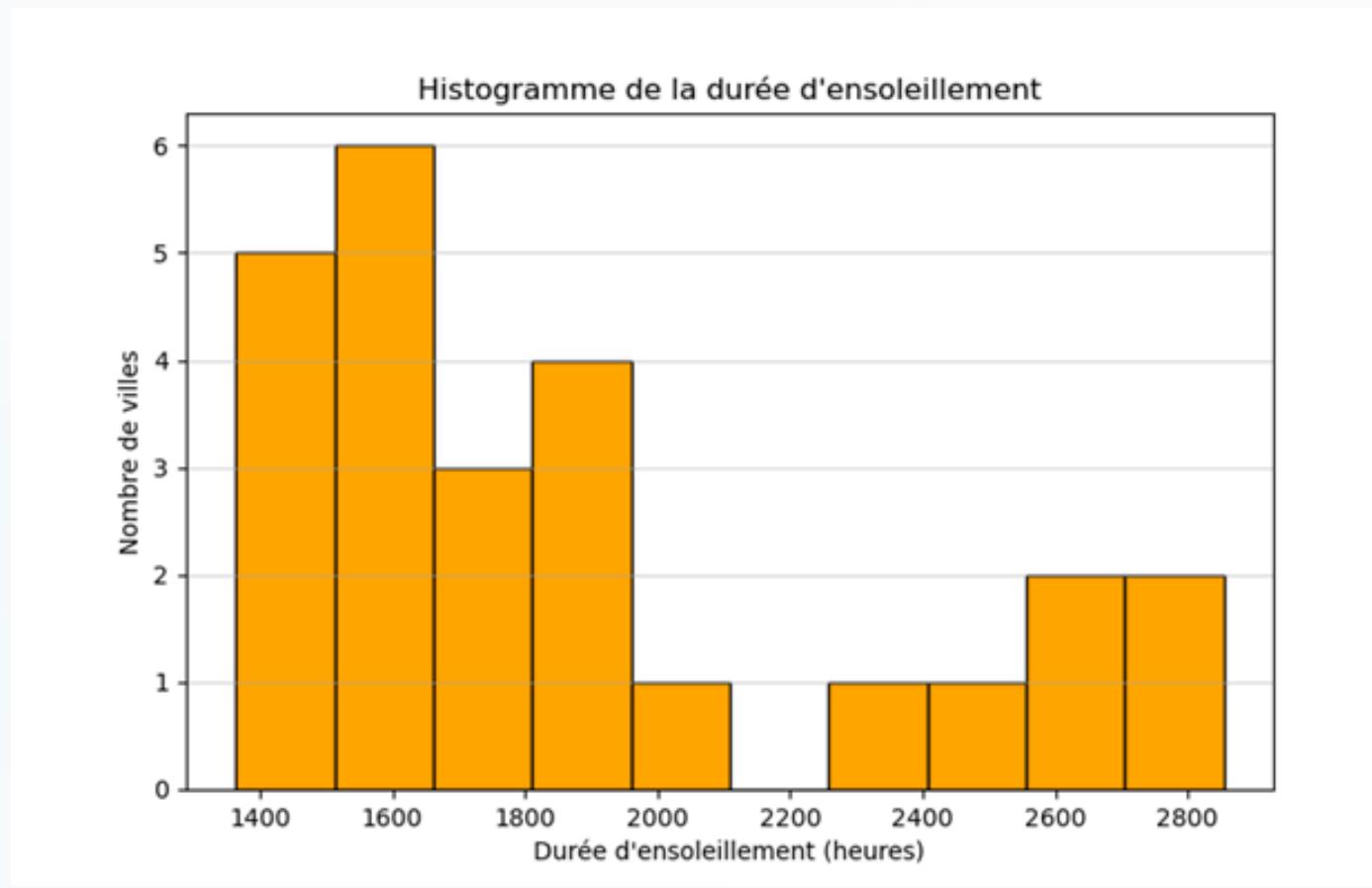
- Beauvais : La ville avec la température minimale la plus basse (7,7 °C).
- Ajaccio : La température maximale la plus élevée (22,5 °C).
- Ajaccio : La valeur la plus faible en précipitations (515,8 mm).
- Biarritz : La valeur la plus élevée en précipitations (1819,4 mm).
- Brest : La durée d'ensoleillement la plus faible (1363,2 heures).
- Ajaccio : La durée d'ensoleillement la plus élevée (2854,5 heures).



Numéro	Variable	Ville	Valeur
[q2a] ↵	Température minimale minimale	Beauvais	7.7 °C ↵
[q2b] ↵	Température minimale maximale	Nice ↵	14.2 °C ↵
[q2c] ↵	Température maximale minimale	Brest ↵	15.5 °C ↵
[q2d] ↵	Température maximale maximale	Ajaccio ↵	22.5 °C ↵
[q2e] ↵	Précipitations minimales	Ajaccio ↵	515.8 mm ↵
[q2f] ↵	Précipitations maximales	Biarritz ↵	1819.4 mm ↵
[q2g] ↵	Ensoleillement minimal	Brest ↵	1363.2 h ↵
[q2h] ↵	Ensoleillement maximal	Ajaccio ↵	2854.5 h ↵

Statistiques Descriptives

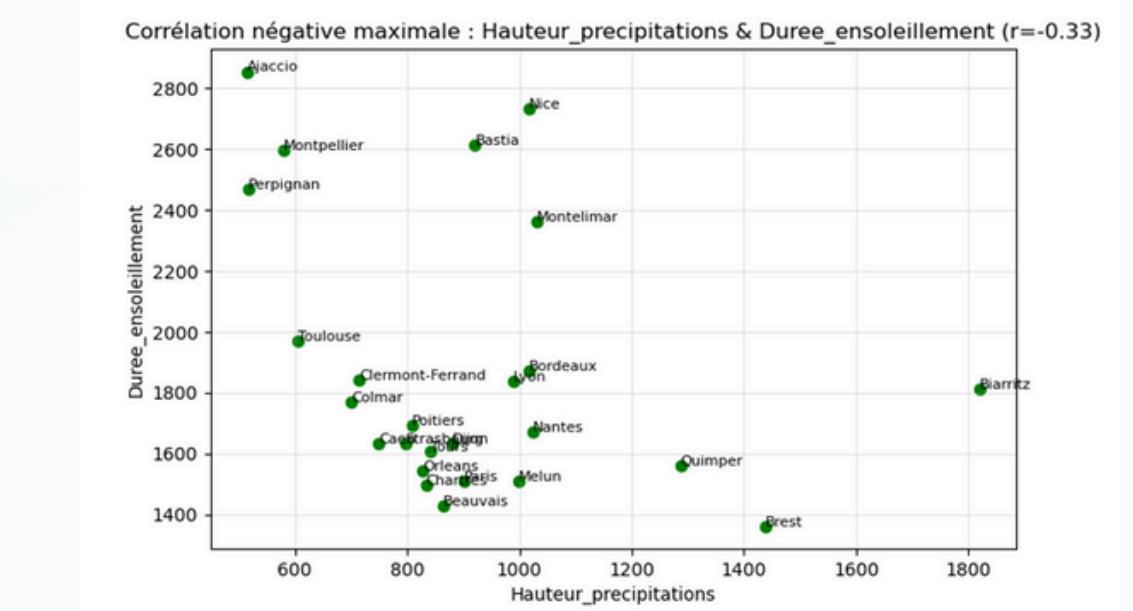
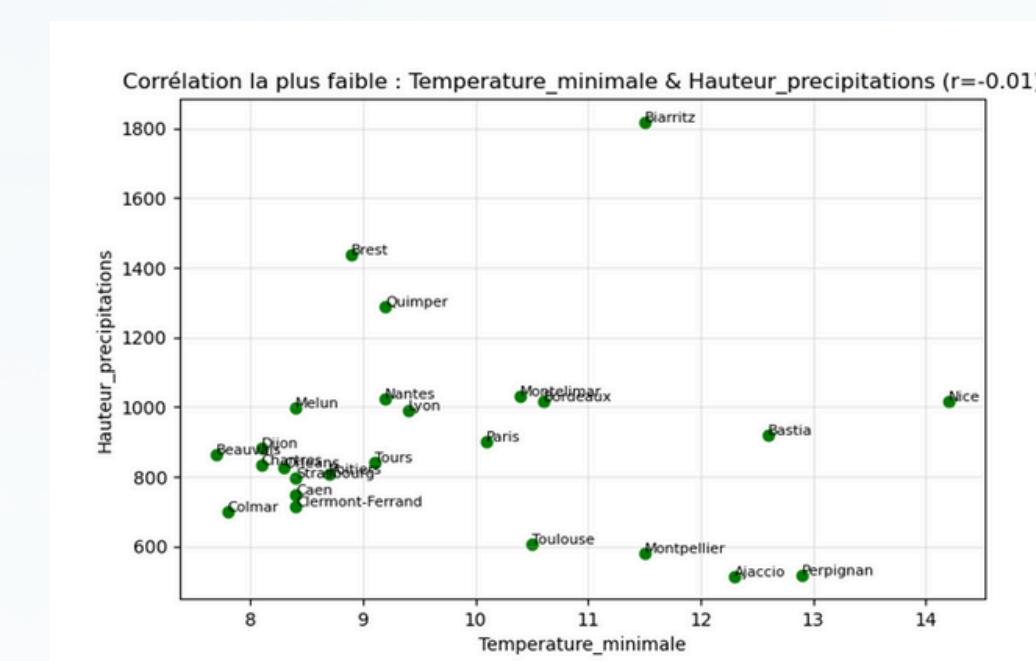
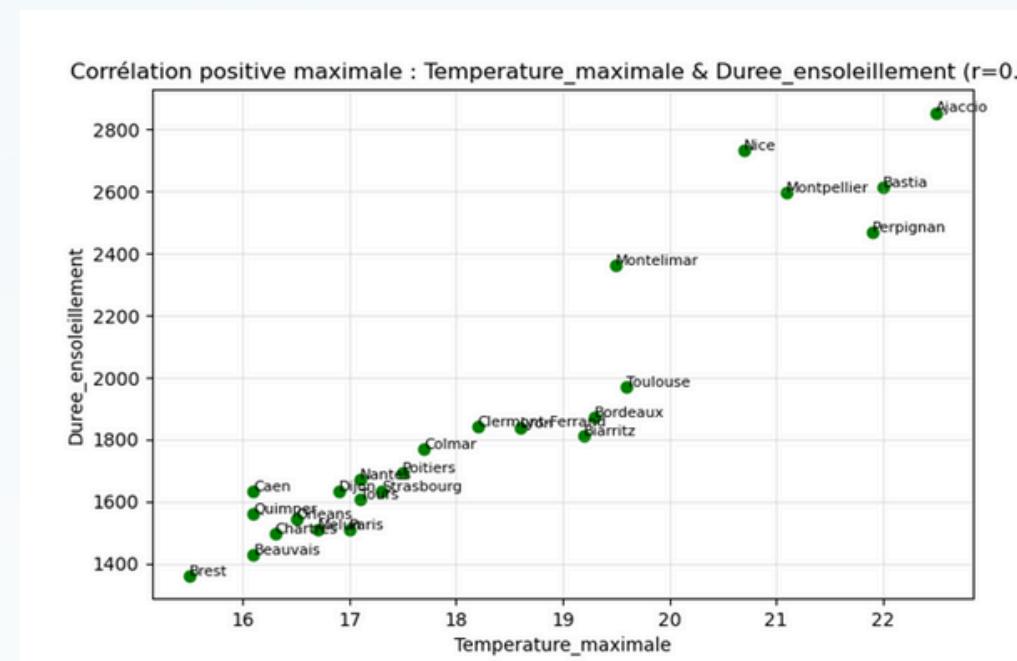
Concernant la variance des variables, nous avons observé que la température minimale présente une variance relativement faible (3,25), ce qui indique une stabilité des températures nocturnes. En revanche, la durée d'ensoleillement présente la variance la plus élevée (198 650), reflétant des différences régionales marquées.



L'histogramme montre que les températures minimales sont principalement regroupées entre 8 °C et 10 °C, tandis que la durée d'ensoleillement présente une distribution bimodale, certaines villes bénéficiant de beaucoup plus de soleil que d'autres.

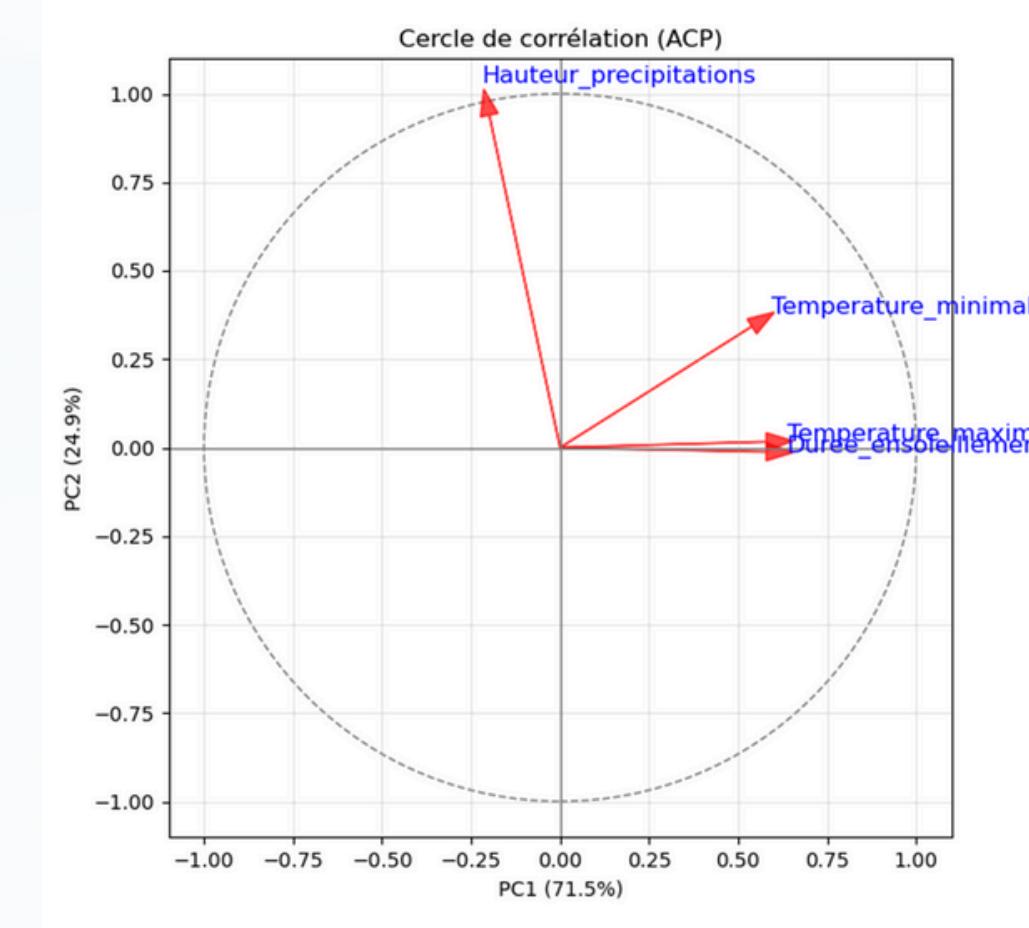
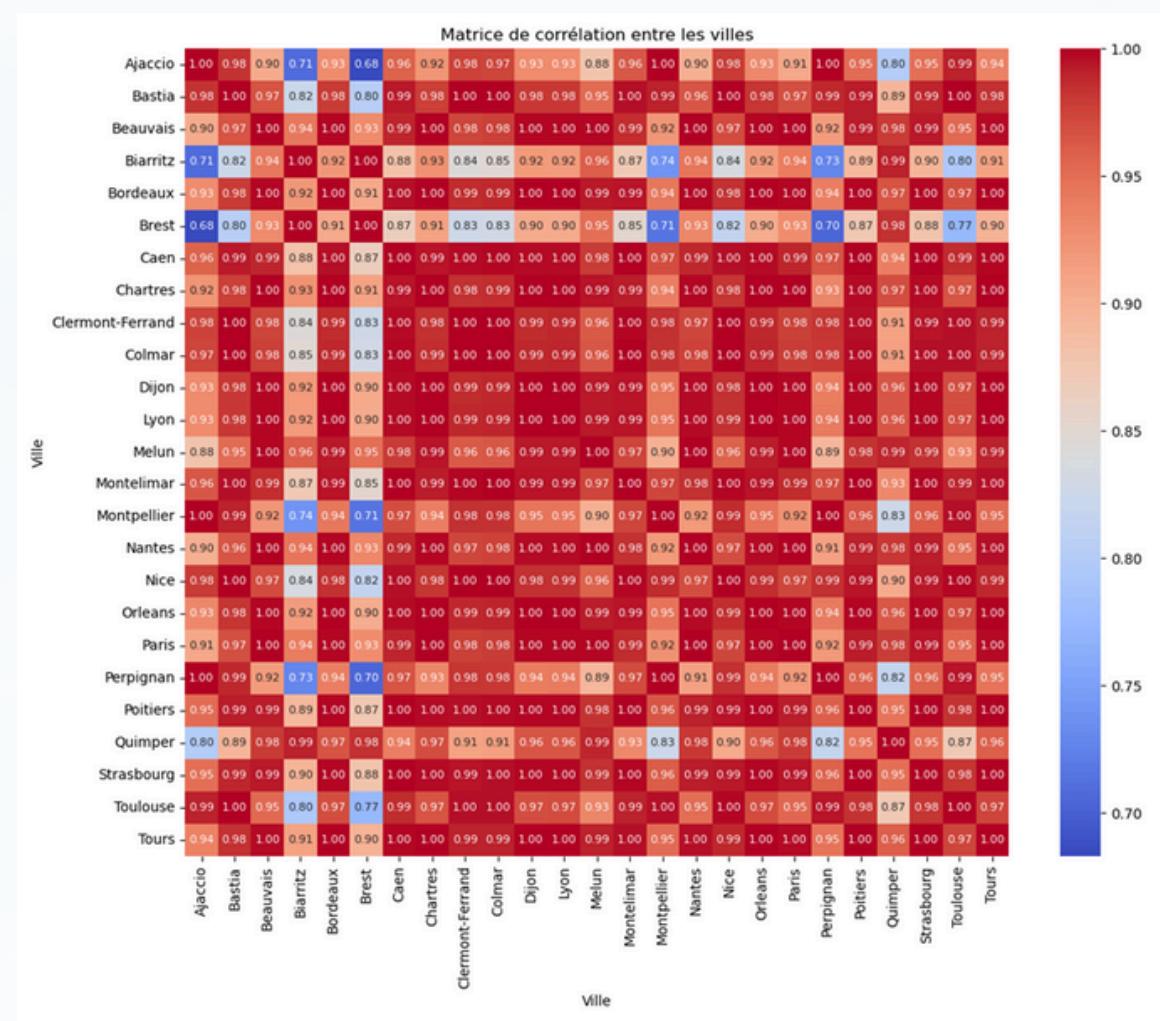
Statistiques Descriptives

- Dans l'analyse des corrélations entre les variables, la corrélation entre la température minimale et la température maximale est très forte ($r \approx 0,95$), ce qui montre une forte liaison entre les températures diurnes et nocturnes.
- En revanche, la corrélation entre la hauteur des précipitations et la durée d'ensoleillement est pratiquement nulle ($r \approx 0,00$), ce qui indique que ces deux variables sont indépendantes l'une de l'autre.
- De plus, il existe une corrélation négative modérée entre la température minimale et les précipitations ($r \approx -0,33$), ce qui signifie que les villes où les températures sont plus basses tendent à recevoir plus de précipitations.



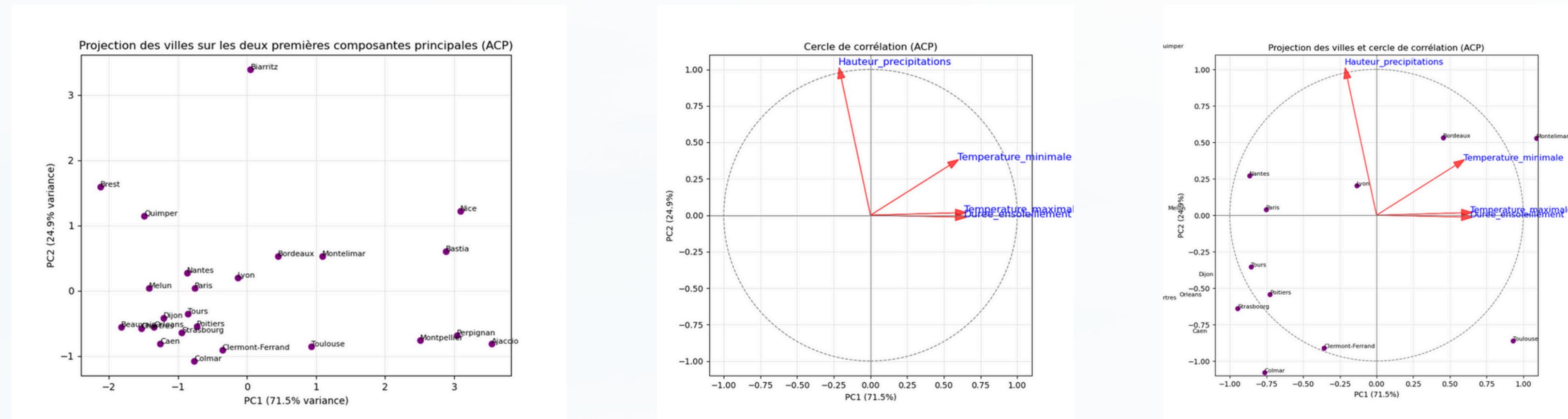
Statistiques Descriptives

- Ensuite, nous avons effectué une Analyse en Composantes Principales (ACP) pour réduire la dimensionnalité des données météorologiques. Cette méthode nous a permis d'identifier les principales caractéristiques climatiques des 25 villes.
- La première composante principale (PC1) explique 71,54 % de la variance totale et est principalement liée à la température maximale et à la durée d'ensoleillement, tandis que la deuxième composante principale (PC2) explique 24,93 % de la variance et est influencée par la hauteur des précipitations et la température minimale. Ensemble, ces deux composantes expliquent plus de 96 % de la variance des données, ce qui signifie qu'elles résument bien les principales informations climatiques.



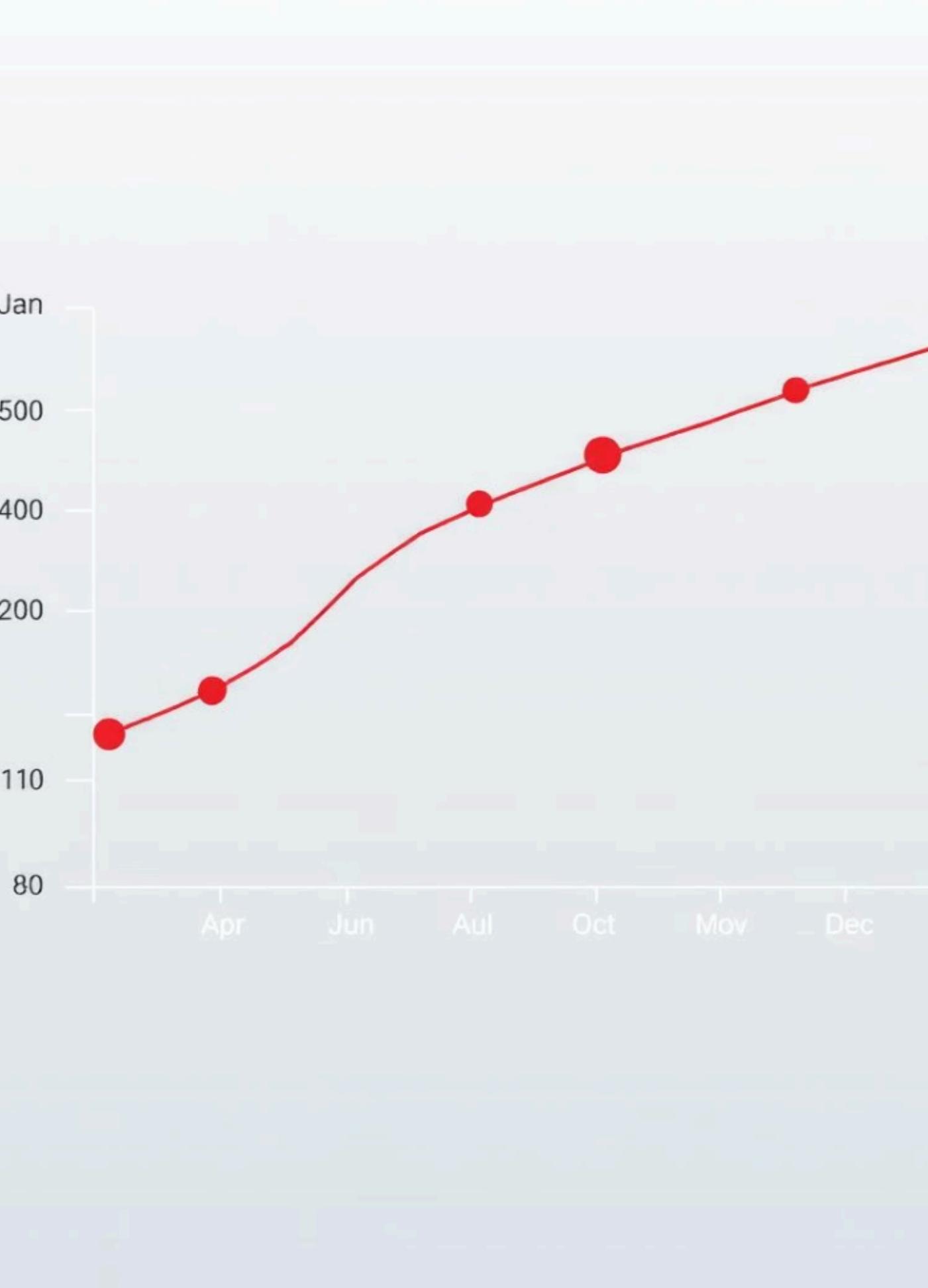
Statistiques Descriptives

- Dans le nuage de points ACP, nous avons observé que des villes comme Ajaccio et Brest occupent des positions extrêmes, mettant en évidence leurs caractéristiques climatiques spécifiques.
- Le cercle de corrélation confirme également les relations entre les variables : une forte corrélation entre la température et l'ensoleillement, et une indépendance relative des précipitations par rapport aux autres variables.



- Enfin, en superposant le nuage de points ACP avec le cercle de corrélation, nous avons pu relier ces observations aux villes extrêmes identifiées précédemment.

Régression Linéaire Simple



Nous avons analysé l'évolution mensuelle de la température maximale à Paris en 2024, qui suit un cycle saisonnier classique. Pour prédire janvier 2025, nous avons testé 12 modèles de régression.

Modèle Optimal

Utilise les 5 derniers mois de données.

- R^2 ajusté = 0.9745
- $\beta_0 = 58.02, \beta_1 = -4.58$

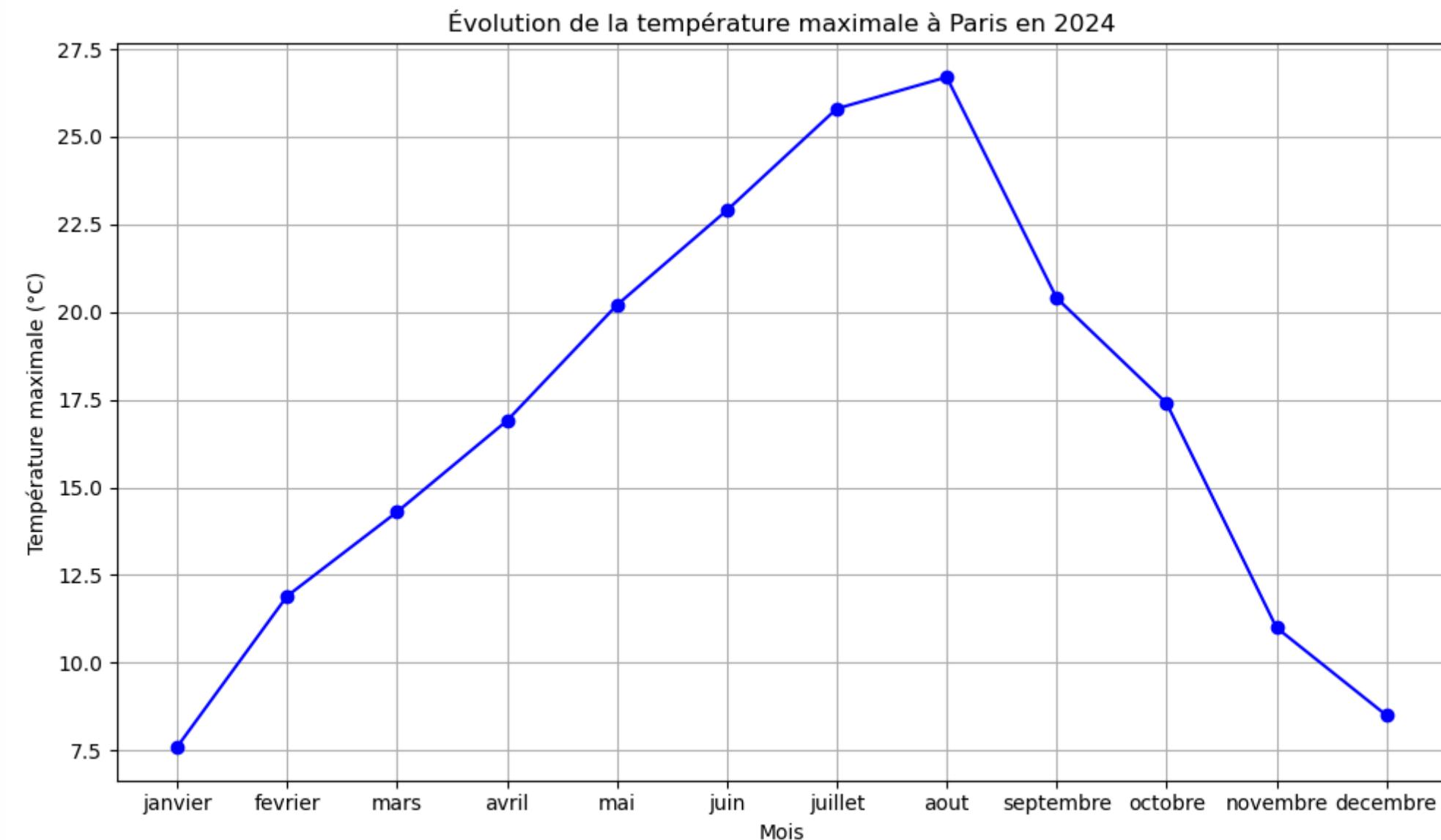
Prédiction Janvier 2025

Température prédictive: 3,06°C.
Écart par rapport à la réalité (7,5°C): -4,44°C.

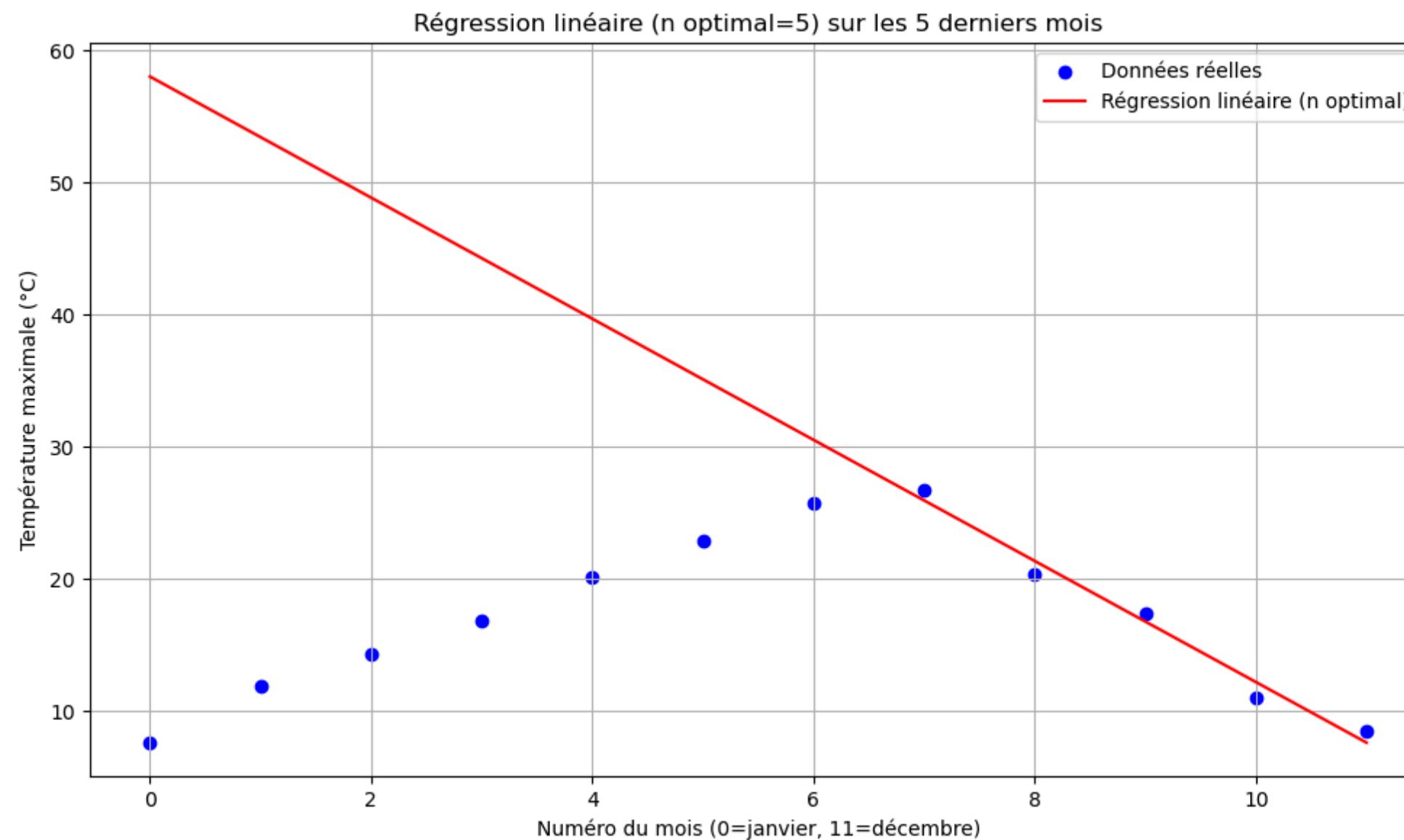
Validation Statistique

Test de significativité de la pente: $p = 0.0011$, confirmant la pertinence du modèle malgré la sous-estimation.

La courbe température 2024 :



Régression linéaire



$$\text{Température} = \beta_0 + \beta_1 \times \text{Mois}$$

Valeur optimale de n : 5

R^2 ajusté optimal : 0.9745

R^2 optimal : 0.9808

Coefficient β_0 : 58.0200

Coefficient β_1 : -4.5800

Équation du modèle :

Température = 58.0200 + -4.5800 × Numéro_mois

Conclusion ($\alpha=5\%$) : Oui, la pente est significative : il existe une relation linéaire.



Régression Linéaire Multivariée

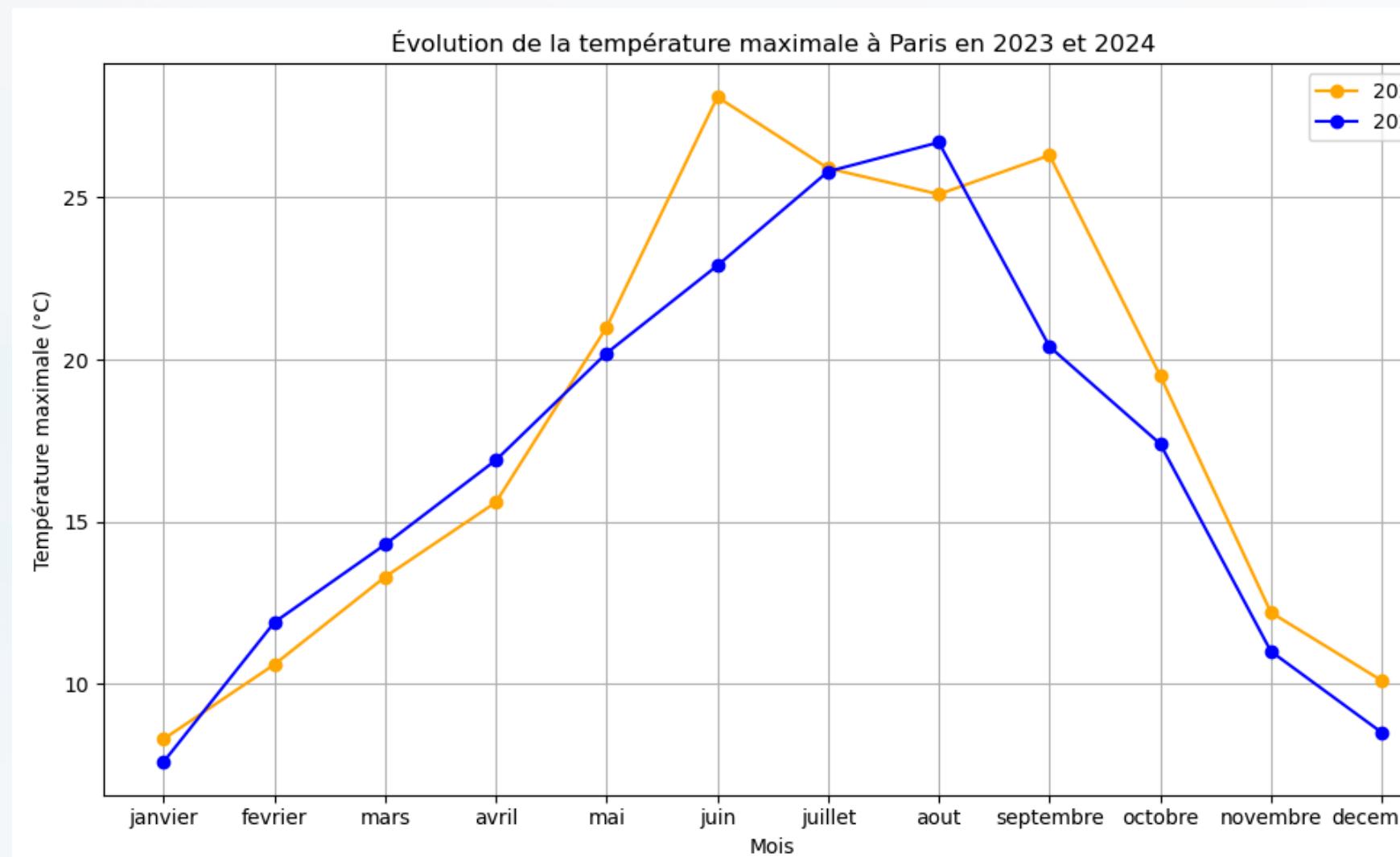
En combinant les données de 2023 et 2024 à Paris, nous avons observé une évolution parallèle des températures sur les deux années, malgré quelques écarts d'amplitude. Cette approche a permis d'explorer des modèles plus complexes.



Combinaison de Données

Intégration des données de température de Paris de 2023 et 2024 pour une analyse plus robuste.

Superposition de l'évolution des températures à Paris en 2023 et 2024



Tendances similaires pour les deux années, été 2023 plus chaud avec pics marqués, 2024 plus stable et moins extrême.



Exploration des Combinaisons

Test de 4095 combinaisons de variables, incluant jusqu'à 12 mois de retard, pour identifier le modèle le plus performant.

Notre modèle devient :

$$\text{Température} = \beta_0 + \beta_1 \times \text{Mois}$$



$$\text{Température_du_mois_i} = \beta_0 + \beta_1 \times \text{Temp}_{\{i-1\}} + \beta_2 \times \text{Temp}_{\{i-2\}} + \dots + \beta_{12} \times \text{Temp}_{\{i-12\}}$$

Donc on a nombre des combinaisons :

$$\sum_{k=1}^{12} \binom{12}{k} = 2^{12} - 1 = 4096 - 1 = \boxed{4095}$$



Modèle Optimal

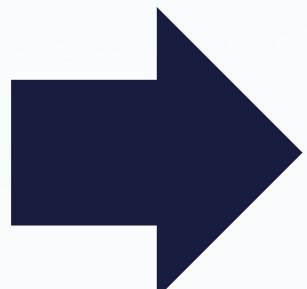
Le meilleur modèle utilise 9 mois passés (Temp_1 à Temp_12 sauf 3, 4, 6), offrant un R² ajusté de 0.9997 et une excellente significativité statistique.

Formule de résolution (équation normale) :

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Explication :

- $X^T X$: c'est la **matrice de covariance** entre les variables explicatives
- $X^T y$: c'est la **corrélation entre les variables et la variable cible**



```
== Détails du meilleur modèle ==
R² ajusté : 0.9997
Nombre de variables sélectionnées : 9
```

Détails des variables :

Nom de variable	Coefficient	p-valeur
Constante	3.6854	0.3316
Temp_1	1.4433	0.0011
Temp_2	-0.9721	0.0053
Temp_5	0.2130	0.0319
Temp_7	-0.7719	0.0014
Temp_8	1.2007	0.0013
Temp_9	-1.1211	0.0032
Temp_10	1.0561	0.0007
Temp_11	-0.8306	0.0037
Temp_12	0.5465	0.0029

Équation du modèle :

Température = 3.6854 + 1.4433 × Temp_1 + -0.9721 × Temp_2 + 0.2130 × Temp_5 + -0.7719 × Temp_7 + 1.2007 × Temp_8 + -1.1211 × Temp_9 + 1.0561 × Temp_10 + -0.8306 × Temp_11 + 0.5465 × Temp_12

Conclusion : Oui, il existe une relation linéaire significative ($\alpha=5\%$).

Prédiction Janvier–Avril 2025

Le modèle multivarié a été utilisé pour prédire les quatre premiers mois de 2025 à Paris. Bien que les prédictions soient précises à court terme, les écarts augmentent avec l'horizon temporel, soulignant les limites du modèle à long terme.

Mois	Prédite (°C)	Réelle (°C)	Écart
Janvier	7.95	7.5	+0.45
Février	8.65	8.6	+0.05
Mars	11.96	14.6	-2.64
Avril	18.91	20.0	-1.09

[q17] Prédictions pour janvier-avril 2025 :
Janvier 2025 :

Température prédite : 7.95°C
Température réelle : 7.5°C
Écart : 0.45°C

Février 2025 :
Température prédite : 8.65°C
Température réelle : 8.6°C
Écart : 0.05°C

Mars 2025 :
Température prédite : 11.96°C
Température réelle : 14.6°C
Écart : -2.64°C

Avril 2025 :
Température prédite : 18.91°C
Température réelle : 20.0°C
Écart : -1.09°C

Conclusion de l'Étude

Notre étude a démontré l'utilité des outils statistiques pour l'analyse et la modélisation des données météorologiques. La régression simple s'est avérée efficace pour les prévisions immédiates, tandis que la régression multivariée a montré sa puissance pour des horizons raisonnables.



Synthèse des Données

L'intérêt des outils statistiques pour synthétiser et modéliser les données météorologiques.



Prévisions Immédiates

L'efficacité de la régression simple pour des prévisions à court terme.

3

Puissance Prédictive

La robustesse de la régression multivariée, avec des limites sur l'horizon temporel.

En perspective, l'intégration de variables supplémentaires (humidité, vent) ou l'exploration de modèles non linéaires (réseaux de neurones) pourrait améliorer la précision des prévisions à long terme.

Merci pour votre attention !

Nous espérons que cette présentation vous a apporté un éclairage sur l'application de l'ACP et de la régression linéaire dans l'analyse météorologique. N'hésitez pas si vous avez des questions.

