Application of Spatial Downscaling Methods to Climatological Data in the Pyrenees

Case study applied to the Pyrenees within the framework of the ANTICI'PYR project (2024–2027)

Salma Bensmail
M1 – HPC & Simulation

Supervised by:
Sébastien Pinel

Presented at:
AI Day – OMP
10 juillet 2025

Salma.bensmail@etudiant.univ-perp.fr

Introduction

• Climate change has a significant impact on mountain ecosystems, particularly on the flowering period of certain plant species.

• In the Pyrenees, a local species called *Delphinium montanum* has not been flowering regularly for several years, raising questions about the role of climate in this phenomenon.

• However, available climate data often have a **spatial resolution that is too coarse (8 km)** to accurately reflect the real conditions where these plants grow.

Salma.bensmail@etudiant.univ-perp.fr

Internship Context and Research Question

Missing Flowering: a Possible Impact of Climate Change?

Healthy

Deficient

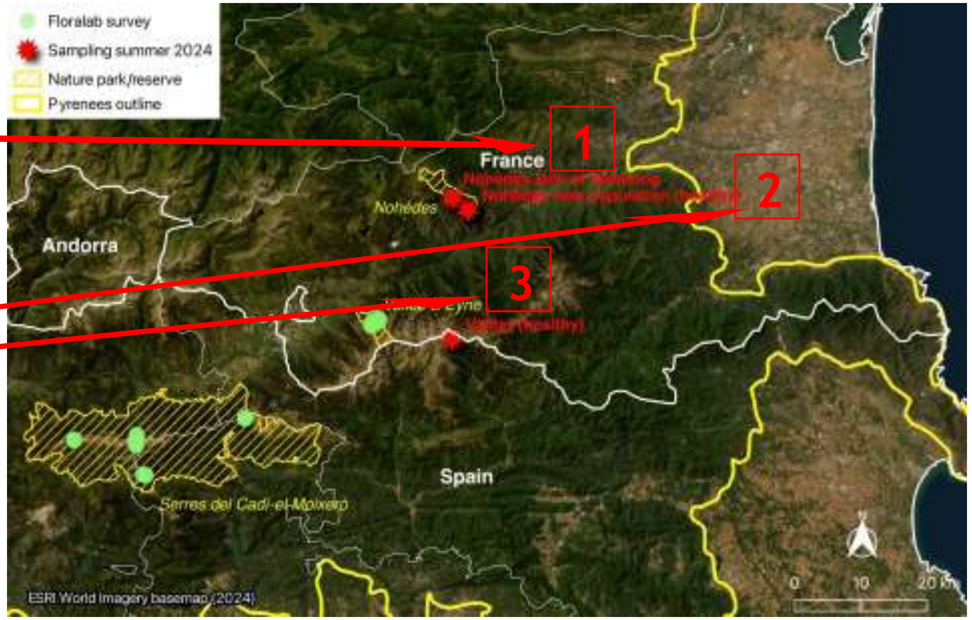


Figure 1 – *study zones used to analyze local climate and flowering behavior.*

Reference points selected for each study zone :

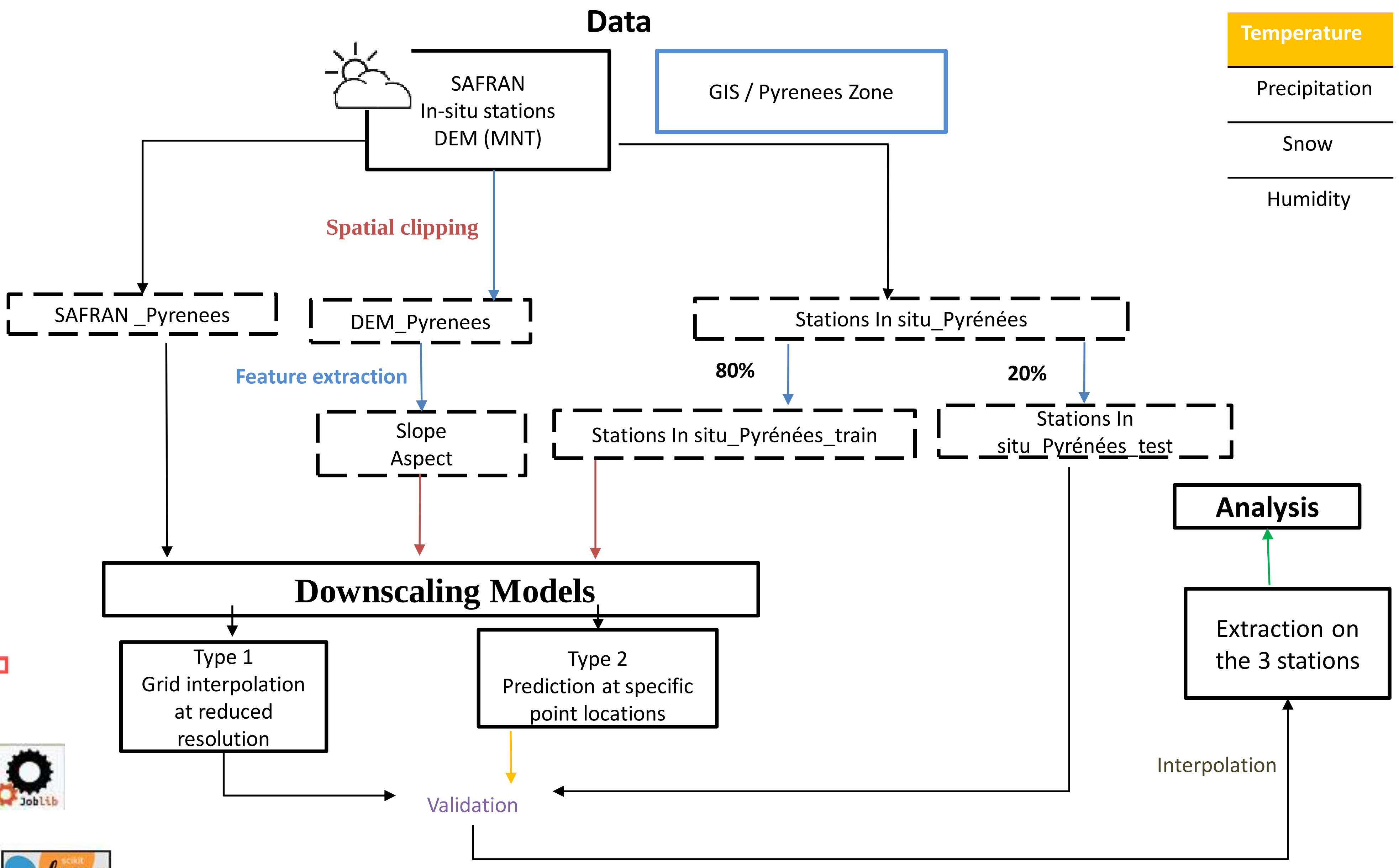
Zone	Latitude	Longitude	Altitude
Nohèdes zone 1	42.61552	2.263319	1705 m
Nohèdes zone 2	42.60009	2.291424	1585 m
Vallter zone 3	42.42589	2.264171	2140 m

Internship Objective:

To apply and test **climate downscaling methods** in order to obtain more accurate data, with the goal of better understanding local conditions and their potential link to flowering patterns.

Salma.bensmail@etudiant.univ-perp.fr

Diagram of the Climate Downscaling Process



Salma.bensmail@etudiant.univ-perp.fr

Datasets Used for Climate Downscaling

Dataset	Resolution	Period	Source
SAFRAN	8 km	1970–2025	donneespubliques.meteofrance.fr
DEM SRTM (NASA)	30 m	—	earthexplorer.usgs.gov
In-situ weather stations	Point	Varies by station	donneespubliques.meteofrance.fr

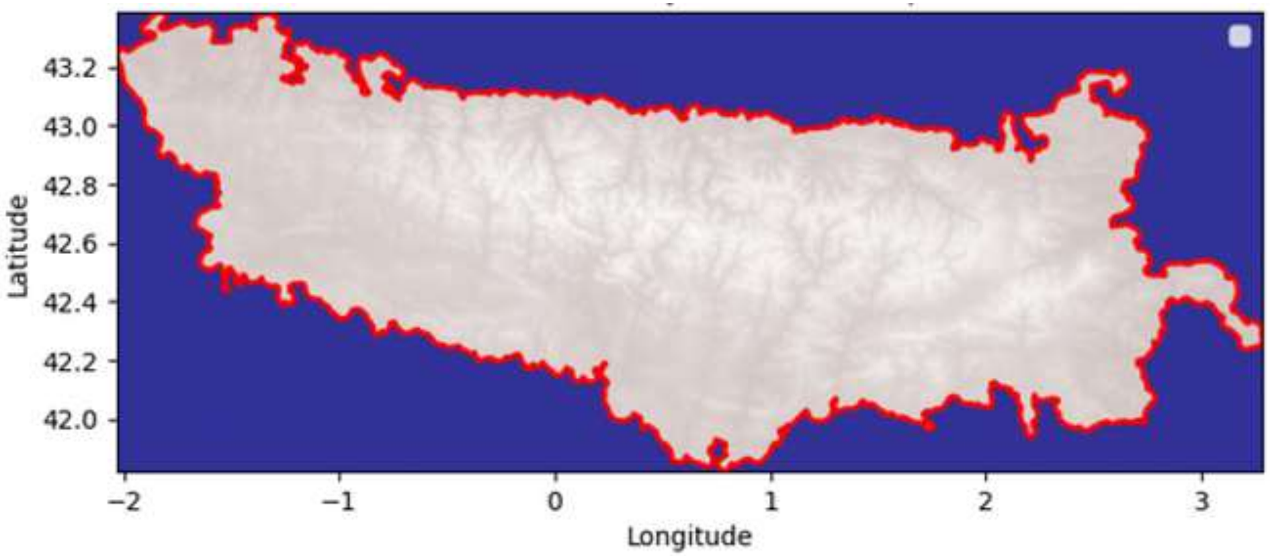


Figure 2 – *Digital Elevation Model (DEM) of the Pyrenees with red contour representing the study area*

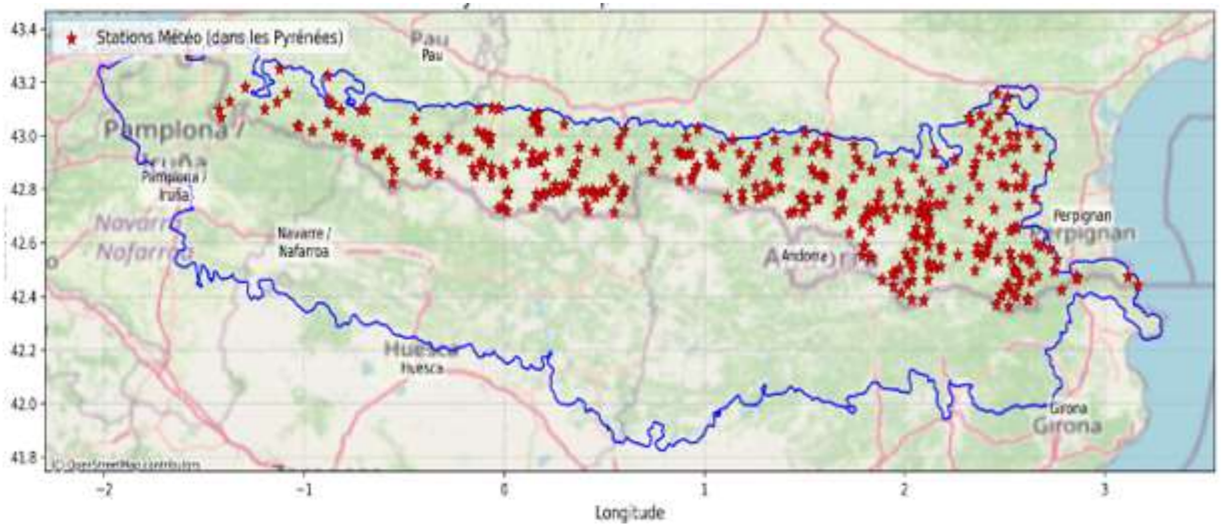


Figure 3 – *Location of in-situ weather stations in the Pyrenees*

Salma.bensmail@etudiant.univ-perp.fr

Modeling with LightGBM

LightGBM (Light Gradient Boosting Machine**)** : a supervised machine learning algorithm based on decision trees.

Model Inputs

Climatic data: temperature, precipitation, snow, humidity...
Topographic data: elevation, slope, aspect, exposure
Temporal data: day of the year

LightGBM Model

Machine Learning

Output
Prediction of a daily climatic variable (e.g., **T_obs**, **P_obs**, etc.)

Figure 4 – *Functioning of the LightGBM model applied to climate downscaling*

Why LightGBM?

- Robust to noisy data
- Optimized for large datasets
- Interpretable (feature importance easily accessible)
- Fast and flexible to train

Salma.bensmail@etudiant.univ-perp.fr

Modeling Pipeline – Maximum Temperature (Tmax)

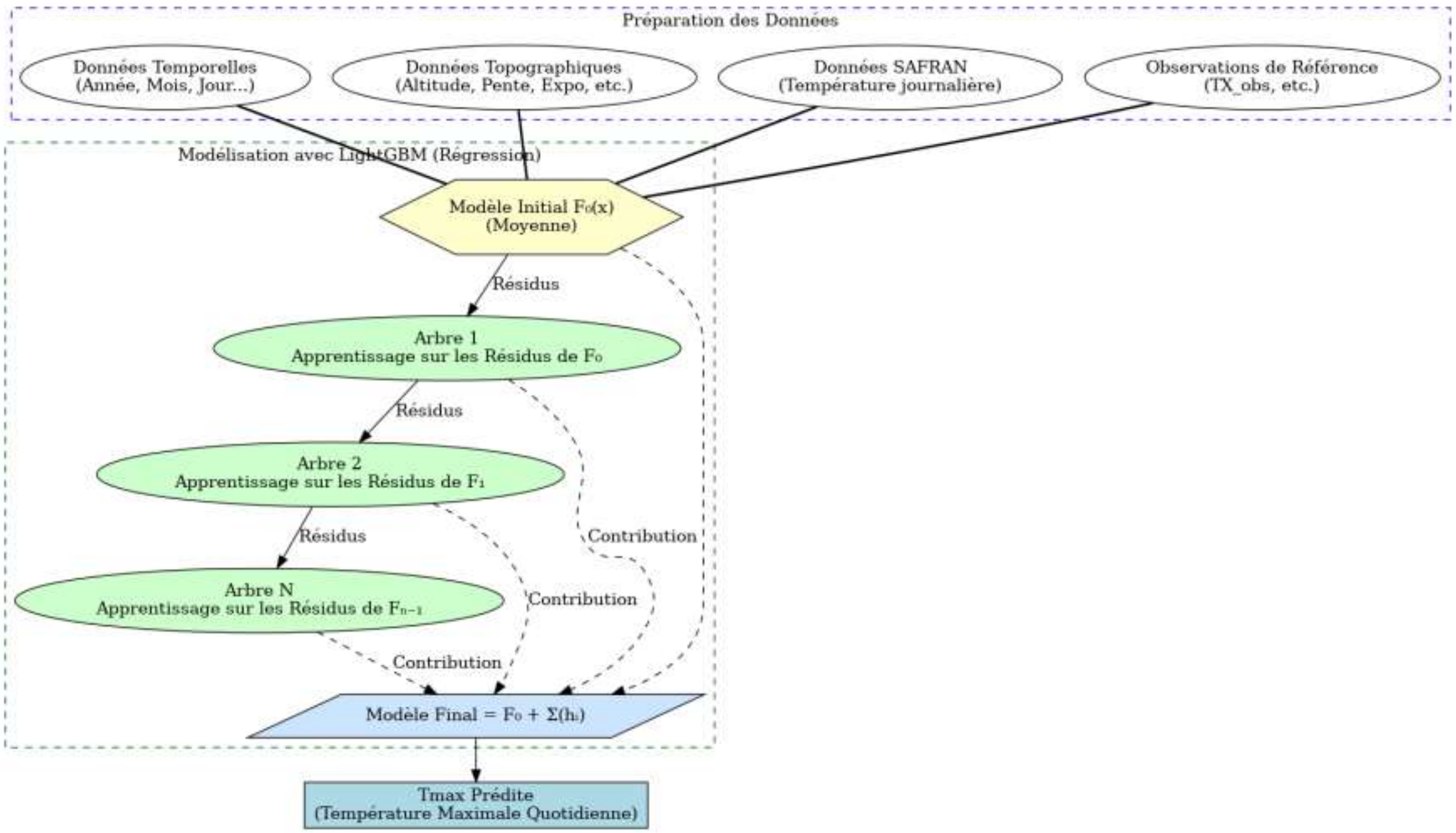


Figure 5 – *Global schematic of the LightGBM modeling process for predicting Tmax*

Salma.bensmail@etudiant.univ-perp.fr

Model Performance Evaluation – Results

Observed vs Predicted Temperatures vs SAFRAN (May 2025) :

ID Station	Date	Donnée de Référence (°C)	Température Prédite (°C)	Température SAFRAN (°C)
11260002	2025-05-09	14.1	12.77	14.2
11260002	2025-05-16	15.0	16.54	17.5
11260002	2025-05-20	14.6	13.42	15.7
66066002	2025-05-09	11.1	12.52	14.2
66066002	2025-05-16	13.0	15.5	17.5
66120001	2025-05-09	4.9	4.54	14.2
66120001	2025-05-16	6.3	9.52	17.5

Comparison of Evaluation Metrics :

Metric	LighthGBM Model	SAFRAN
MAE	✔ 1.330 °C	4.209 °C
RMSE	✔ 1.690 °C	5.550 °C
R^2	✔ 0.885	✘ 0.229

Effectiveness of LightGBM for spatial downscaling of Tmax
Provides **more reliable and accurate predictions** in locations where **raw SAFRAN data performs poorly**.

Salma.bensmail@etudiant.univ-perp.fr

Results Visualization

Correlation between observed and predicted Tmax by the model :

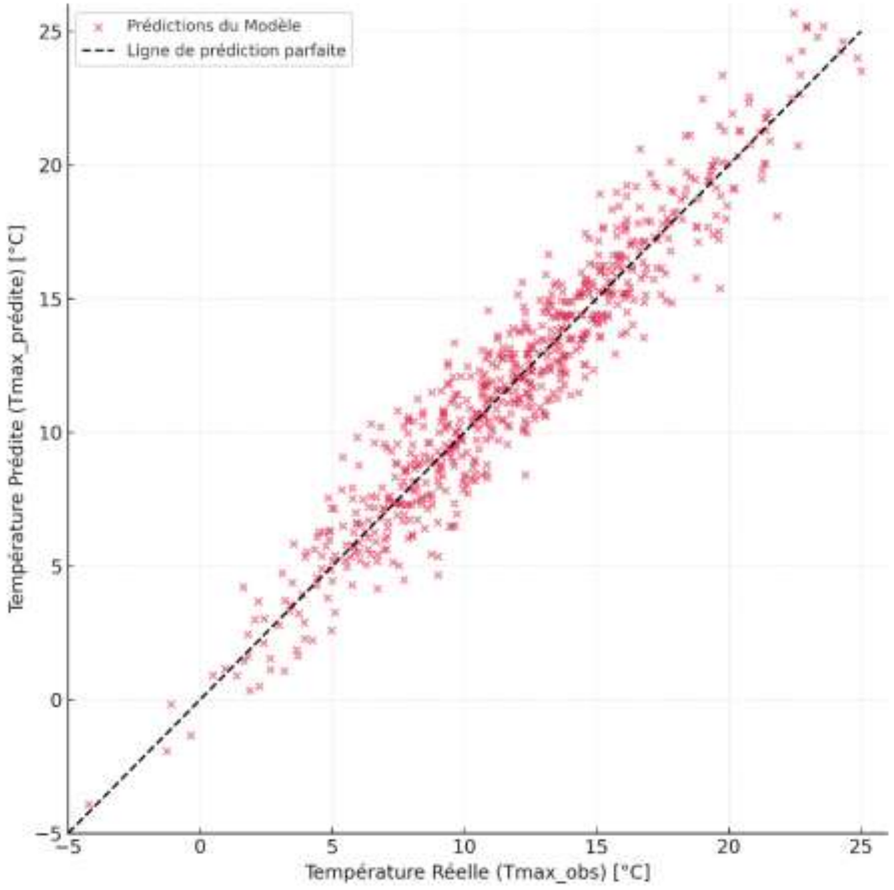


Figure 6 – *Predictions vs Observations of Tmax*

Daily evolution of maximum temperature (Tmax) – Observed, Model, and SAFRAN :

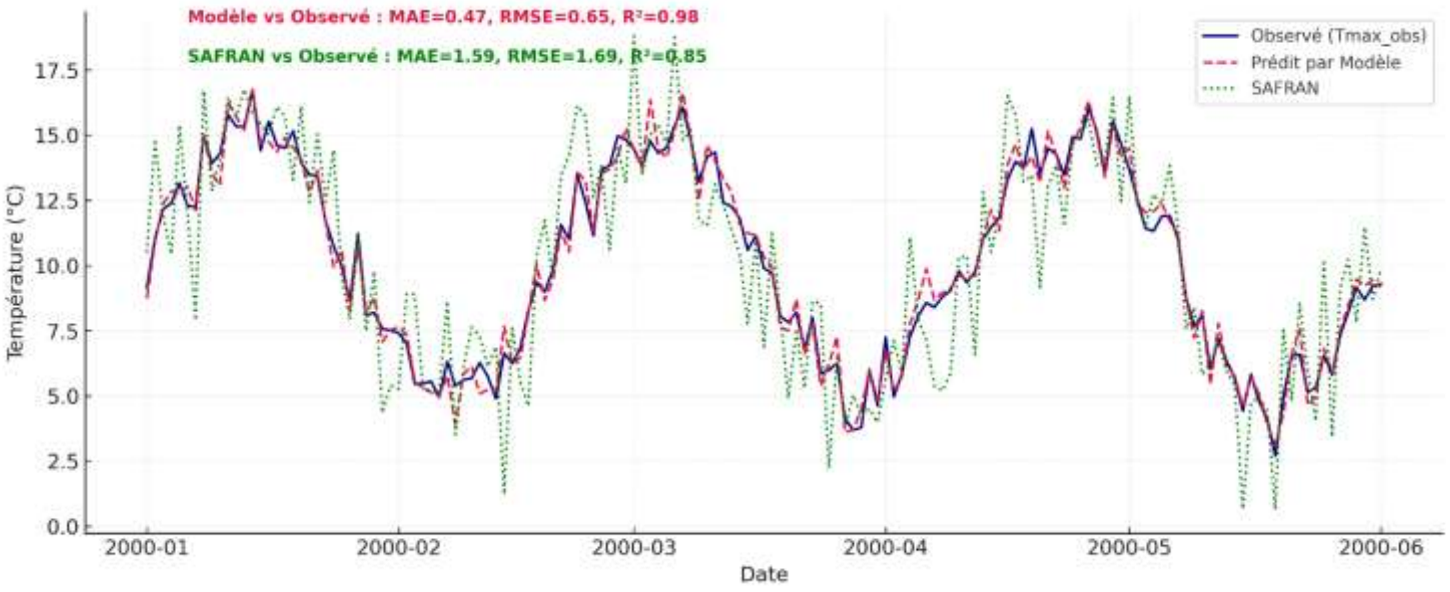


Figure 7 – *Temperatures: Observed, Predicted, and SAFRAN – Station 11260002 (Jan–May 2000)*

Salma.bensmail@etudiant.univ-perp.fr

Model Interpretability

Most Influential Variables in Tmax Prediction :

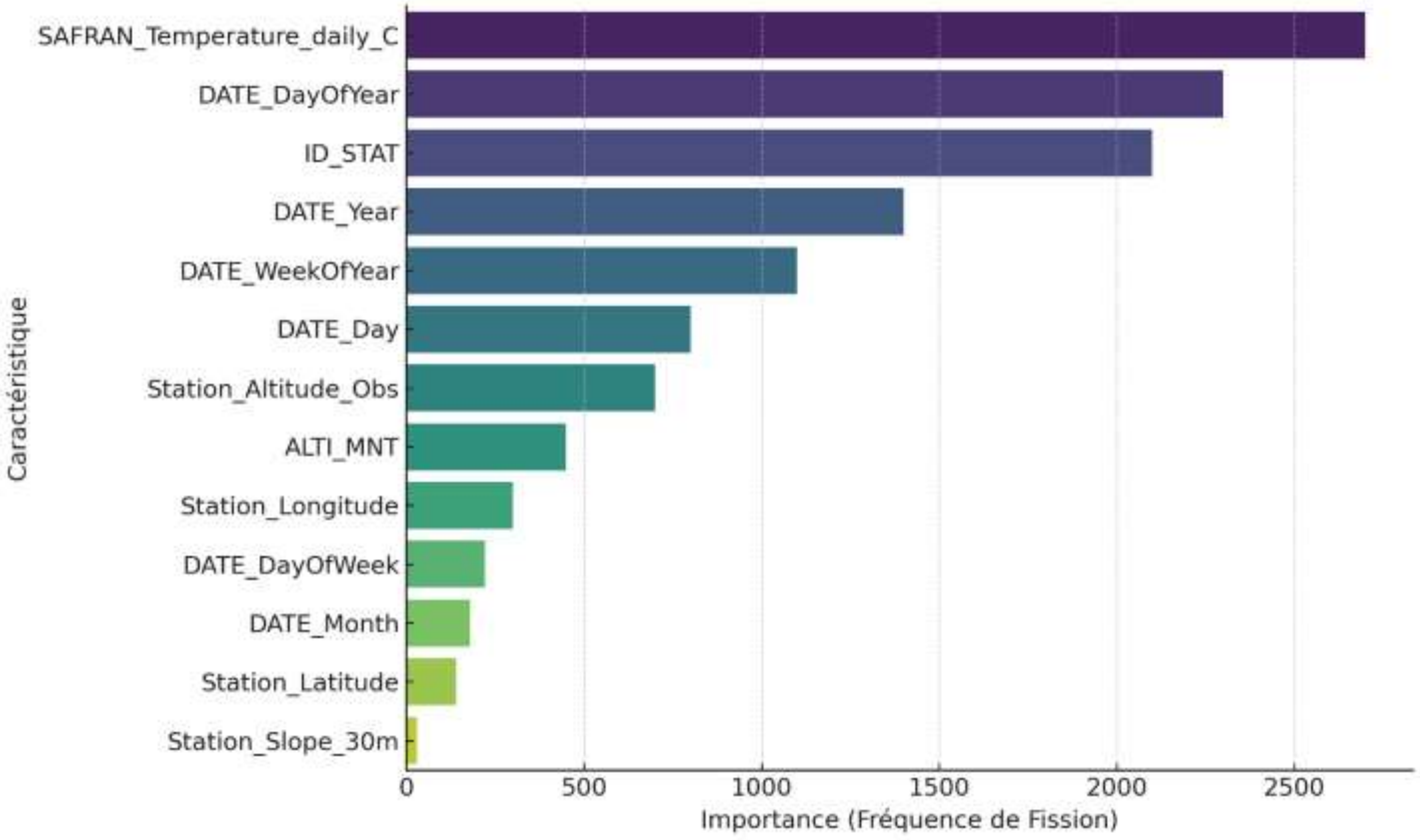


Figure 8 – *Top variables influencing Tmax prediction*

Analysis of Important Features :

• **SAFRAN_Temperature_daily_C**: most influential variable, forms the basis of downscaling.

• **DATE_DayOfYear**, **DATE_Year**: important to capture **seasonality** and annual trends.

• **ID_STAT**, **Station_Altitude_Obs**, **ALTI_MNT**: essential to distinguish **local/orographic effects**.

• **Longitude, Latitude, Slope, Aspect**: lower individual importance, but contribute to **local adjustments**.

Salma.bensmail@etudiant.univ-perp.fr

Conclusion

• This work demonstrated the effectiveness of the LightGBM approach for **downscaling daily maximum temperature (Tmax)** using SAFRAN data and topographic variables.

• Through a detailed modeling process that incorporates **altitude, seasonality**, and **geographic coordinates**, we significantly improved prediction accuracy compared to raw SAFRAN data, with consistent performance across different stations.

• These results highlight the **potential of machine learning in climate studies**, and open up perspectives for applying this method to **other environmental variables** and **other study areas** within the **ANTICI'PYR** project

Salma.bensmail@etudiant.univ-perp.fr

Discussion and Limitations

Discussion of Results:

- The LightGBM model significantly improves accuracy compared to SAFRAN.
- Seasonality and altitude effects are well captured.
- Some discrepancies remain for highly exposed or underrepresented stations.

Limitations:

- Model currently calibrated only for **Tmax**.
- Need to integrate **other climatic variables** (precipitation, snow, etc.).
- Strong dependency on **observation data quality**.

Salma.bensmail@etudiant.univ-perp.fr

Improvements and Perspectives

🛠 **Improvement Tracks:**

- Test other machine learning methods (Random Forest, XGBoost, etc.).
- Refine topographic variables (e.g., more precise DEM, true slope/exposure).
- Add external data (e.g., **cloud cover**).

📅 **Next Steps:**

- Extend the analysis to **other seasons** and **other years**.
- Apply the model to **other study areas** of the **ANTICI'PYR project**.
- Investigate the **link between modeled climatic conditions and flowering**.

Thank you for your attention!