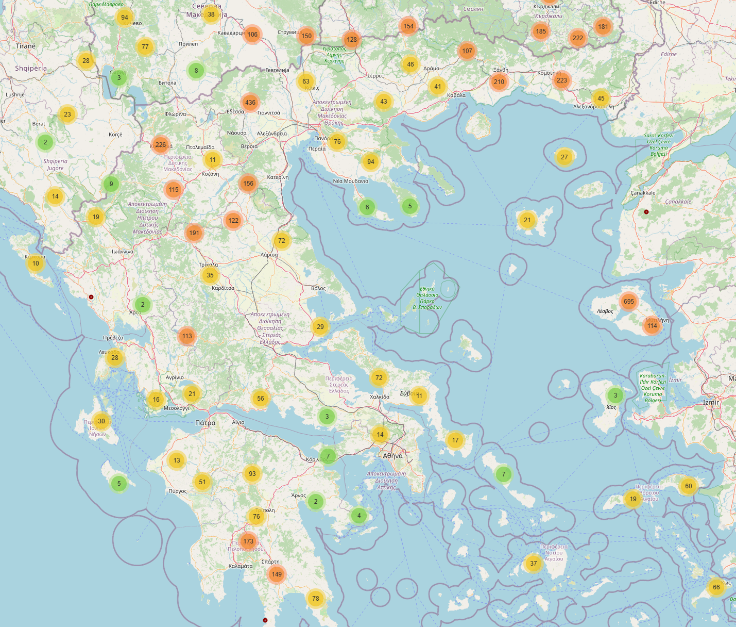
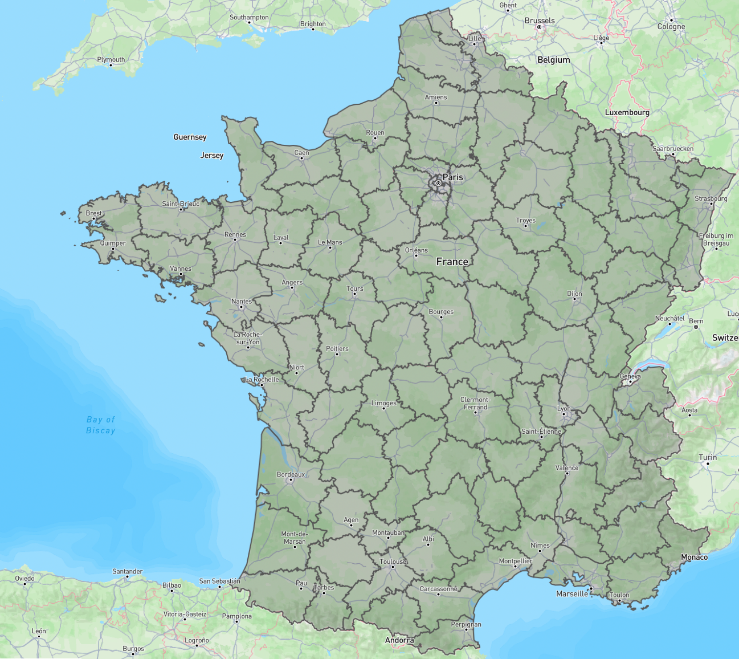
**Introduction**:

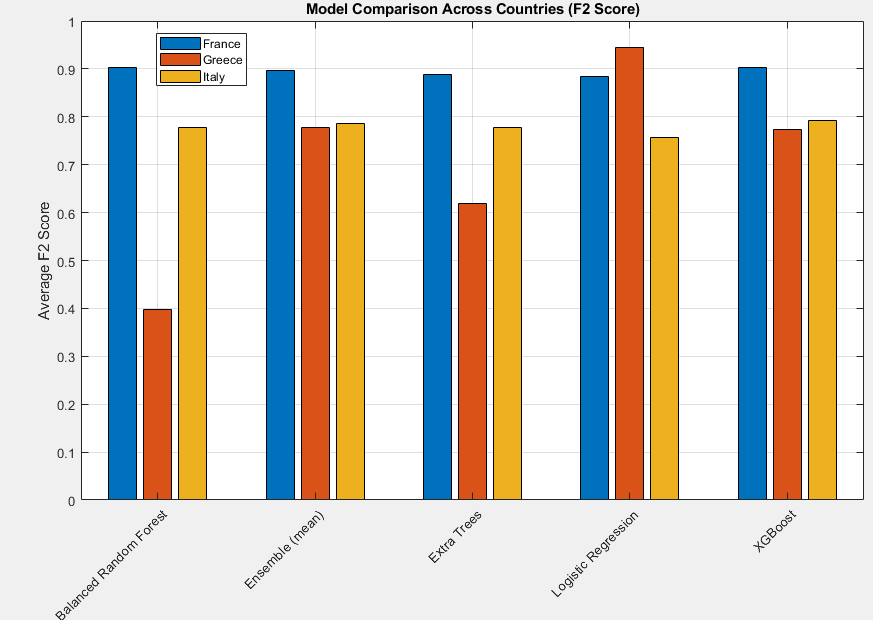
(empty)

**Data**:

* We focused on 3 countries: **Greece, Italy, and France**, to predict the **bluetongue virus outbreaks**.
  + These countries have had many bluetongue cases in the past.
  + They experienced continuous outbreaks between **2000 and 2025**.
  + Their datasets are **complete** and include **rich climate features**.
* We used **monthly data** from **January 2000 to December 2024** (25 years total).
* **Features** include:
  + Precipitation
  + Humidity
  + Surface pressure
  + Mean temperature
  + Max temperature
  + Min temperature
  + Wind speed
* **Data processing steps:**
  + Each province (area) is treated as one data point.
  + We averaged all grid data (latitude and longitude every ~10km across Europe) within each province.
  + For **precipitation**, we summed the values instead of averaging.
  + We created new features such as:
    - **Temperature range** (max - min)
    - **Rolling values** of precipitation over time
    - **Rolling outbreak history**, because past outbreaks are likely to affect future ones.

**Method**

* We used **4 popular models** to predict outbreaks for the next 6 months:
  + Balanced Random Forest
  + Extra Trees
  + Logistic Regression
  + XGBoost
* The data was split into:
  + **Training**: January 2000 – December 2021
  + **Testing**: January 2022 – December 2024
* **Prediction process:**
  + We used a **sliding window** method:
    - Each window looks back **6 months** at climate data to predict outbreaks in the **next 6 months**.
  + The model learns patterns from the training set, and then makes predictions on the testing set.
  + We evaluated and recorded the prediction results.
* Finally, we used an **ensemble model** that takes the **average prediction of all 4 models**, which gave us our final result.



**Results**

* The **ensemble model** produced strong and **stable results** across all countries and timeframes.

**Discussion**:

**Main Findings**

* Our model successfully predicts bluetongue outbreaks up to 6 months in advance using only climate and historical outbreak data.
* The ensemble model showed the most stable performance across France, Greece, and Italy, suggesting that combining multiple models helps mitigate individual weaknesses.

**Interpretation**

* The strong results of tree-based models (Balanced Random Forest, Extra Trees) suggest that the **relationship between climate variables and outbreak patterns is non-linear and complex**, which these models can capture well.
* The high F2 scores across all countries suggest the model is **especially good at detecting outbreaks**, which is great public health field where missing outbreaks is more costly than false alarms.

**Limitation**:

(work badly for data that has limited outbreak history in training data part)

**Future Work**

* Incorporate **vector ecology data** and **animal movement patterns** to improve model precision.
* Explore the use of **spatiotemporal deep learning models** (e.g., LSTMs, CNNs) that can better capture temporal dependencies and regional interactions.
* We can leverage data from countries with **rich outbreak histories** to help **predict outbreaks in countries with limited data**, improving model performance in regions where **outbreaks are rare**.

**Conclusion**:

(empty)