Spatio-Temporal Forecasting of Weather Variables Using Vector AutoRegression (VAR)

Dusara Gamidu Team - Xenv

April 3, 2025

github repository

1 Introduction

In this study, we analyze multivariate time series data of weather variables collected from 30 different geographical locations. Given the structured nature of the dataset, we employ a Vector AutoRegression (VAR) model to forecast future values of key meteorological factors: Average Temperature, Radiation, Rain Amount, Wind Speed, and Wind Direction.

2 Methodology

2.1 Vector AutoRegression (VAR) Model

VAR is a powerful statistical method for modeling the linear interdependencies among multiple time series. It extends the univariate autoregressive (AR) model by considering multiple variables that influence each other over time. The general form of a VAR(p) model is given by:

$$Y_t = c + A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + \varepsilon_t, \tag{1}$$

where:

- Y_t is a vector of the observed variables (weather parameters) at time t.
- c is a constant vector.
- A_i are coefficient matrices capturing lagged dependencies.
- ε_t is a vector of white noise error terms.

2.2 Justification for Using VAR

The choice of VAR is motivated by several factors:

1. **Multivariate Nature of the Data:** The dataset contains several interrelated meteorological variables that evolve over time, making VAR an appropriate choice.
2. **Time Series Properties:** The dataset consists of daily records over several years, and since all time series are of equal length and contain no missing data, VAR can efficiently model their temporal dynamics.
3. **Capturing Lagged Dependencies:** Weather variables influence each other with time delays. VAR explicitly models these lag structures, making it useful for understanding the interactions between temperature, wind, and precipitation over time.
4. **Stationarity Assumption:** While VAR requires stationarity, transformations such as differencing can be applied if needed to ensure this condition is met.
5. **Forecasting Capability:** Once trained, the VAR model can generate multi-step forecasts for all variables simultaneously, providing insights into expected weather conditions for the coming months.

2.3 Model Selection and Optimization

To determine the optimal lag order (p), we use standard information criteria:

- Akaike Information Criterion (AIC)
- Bayesian Information Criterion (BIC)
- Hannan-Quinn Information Criterion (HQIC)

The model with the lowest BIC/AIC is chosen to balance model complexity and predictive accuracy.

2.4 Data Preprocessing

Before fitting the VAR model, we perform:

- **Exploratory Data Analysis (EDA):** Checking trends, seasonality, and correlations.
- **Stationarity Tests:** Conducting the Augmented Dickey-Fuller (ADF) test to verify stationarity.
- **Transformation:** If necessary, differencing is applied to remove non-stationarity.
- **Normalization:** Scaling the variables to ensure comparable magnitudes.

2.5 Model Training and Evaluation

The dataset is split into training and testing sets. The model is trained on past data and evaluated using:

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- Forecast plots against actual values

3 Conclusion and Future Work

While VAR provides a strong baseline for multivariate forecasting, future research can explore more advanced spatio-temporal models, such as Spatio-Temporal Autoregressive (STAR) models or deep learning-based approaches like LSTMs with spatial embeddings.