Machine and Deep Learning-Based Drought Prediction in Northwestern Iran Using the Standardized Precipitation Index

Literature Review

Application of artificial intelligence hybrid models for meteorological drought prediction 2022

- This article investigates the use of hybrid artificial intelligence models to predict meteorological droughts using the Standardized Precipitation Evapotranspiration Index (SPEI) in northwestern Iran. It focuses on integrating wavelet transformation with advanced AI techniques to enhance forecasting accuracy over 1-, 6-, and 12-month time scales.

- **Purpose:** Improve drought prediction methods and reduce associated risks.
- **Key Indicator:** SPEI used as the main drought index.
- **Study Area:** Northwestern Iran (Ardebil Province).
- **AI Models Evaluated:** MLPNN, SVR, ANFIS, EDT.

Application of artificial intelligence hybrid models for meteorological drought prediction

The study designs a hybrid modeling framework that leverages data preprocessing techniques, such as wavelet transformation, to tackle non-stationary time series data. It explores different combinations of input parameters—including precipitation, temperature, and historical drought indices—using statistical tools to identify optimal lags before applying AI models.

Application of artificial intelligence hybrid models for meteorological drought prediction

- Input Data Processing:
  - Data normalization via Feature Scaling.
  - Use of ACF and PACF to determine significant lags.
  - Effective Factor Elimination Technique (EFET) to select optimal input parameters.
- Hybrid Modeling:
  - Integrates wavelet transformation (DWT) to extract trends and reduce data noise.
  - Compares both single parameter (drought index alone) and multiparameter models.
- AI Techniques Applied:
  - Multilayer Perceptron Neural Network (MLPNN) with wavelet transformation.
  - Support Vector Regression (SVR), Adaptive Neuro-Fuzzy Inference System (ANFIS), and Ensemble Decision Tree (EDT).

Application of artificial intelligence hybrid models for meteorological drought prediction

The study finds that model performance improves when the input data are carefully selected and preprocessed. Specifically, for short time scales, using the drought index alone with a wavelet-enhanced MLPNN yields superior predictions, while for longer time scales, combining multiple parameters with a wavelet-enhanced SVR delivers the best performance. These results are validated using statistical metrics such as RMSE, Correlation Coefficient (CC), and Nash–Sutcliffe Efficiency (NSE).

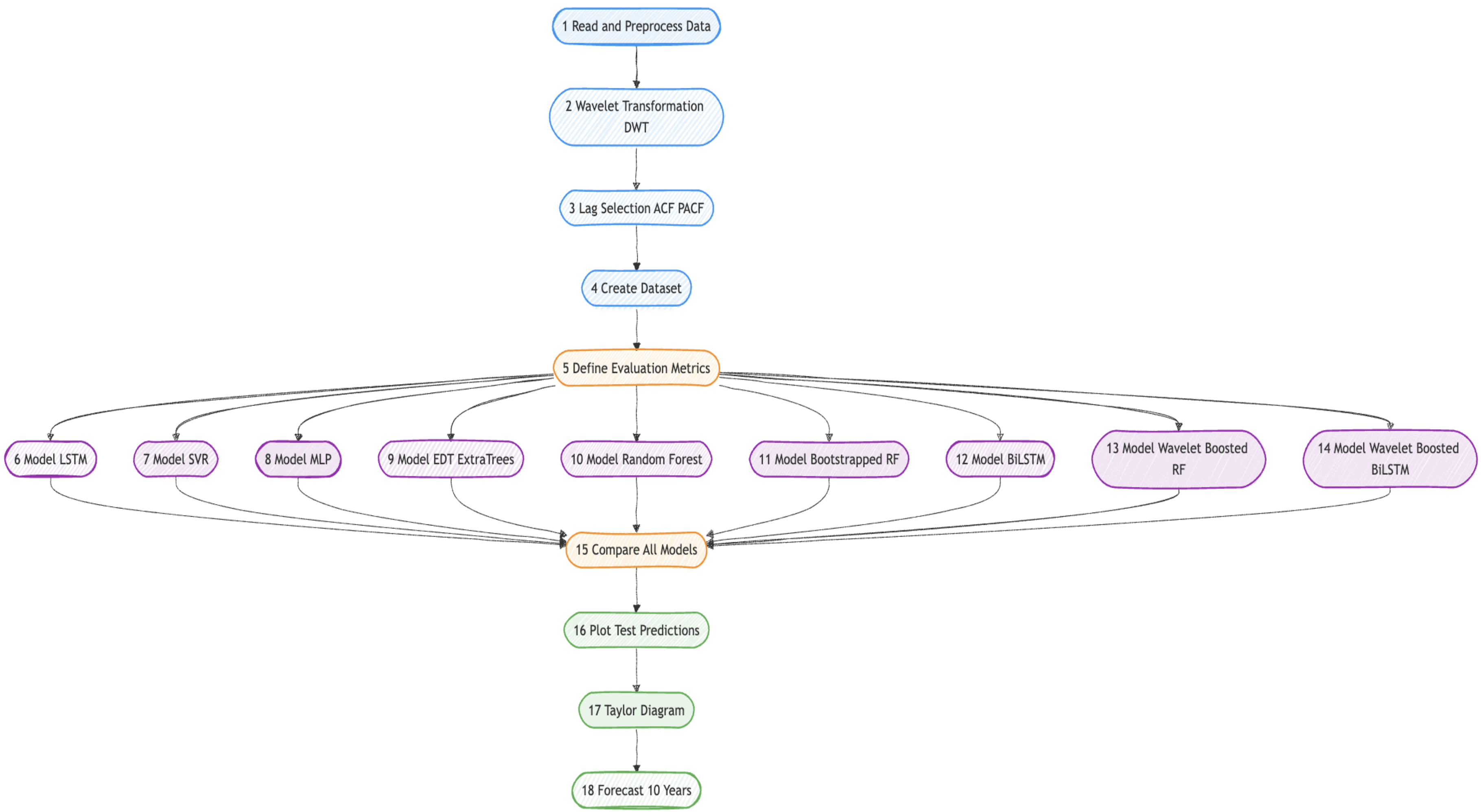
Application of artificial intelligence hybrid models for meteorological drought prediction

- Performance Metrics:
  - RMSE, CC, and NSE used to evaluate each model's accuracy.

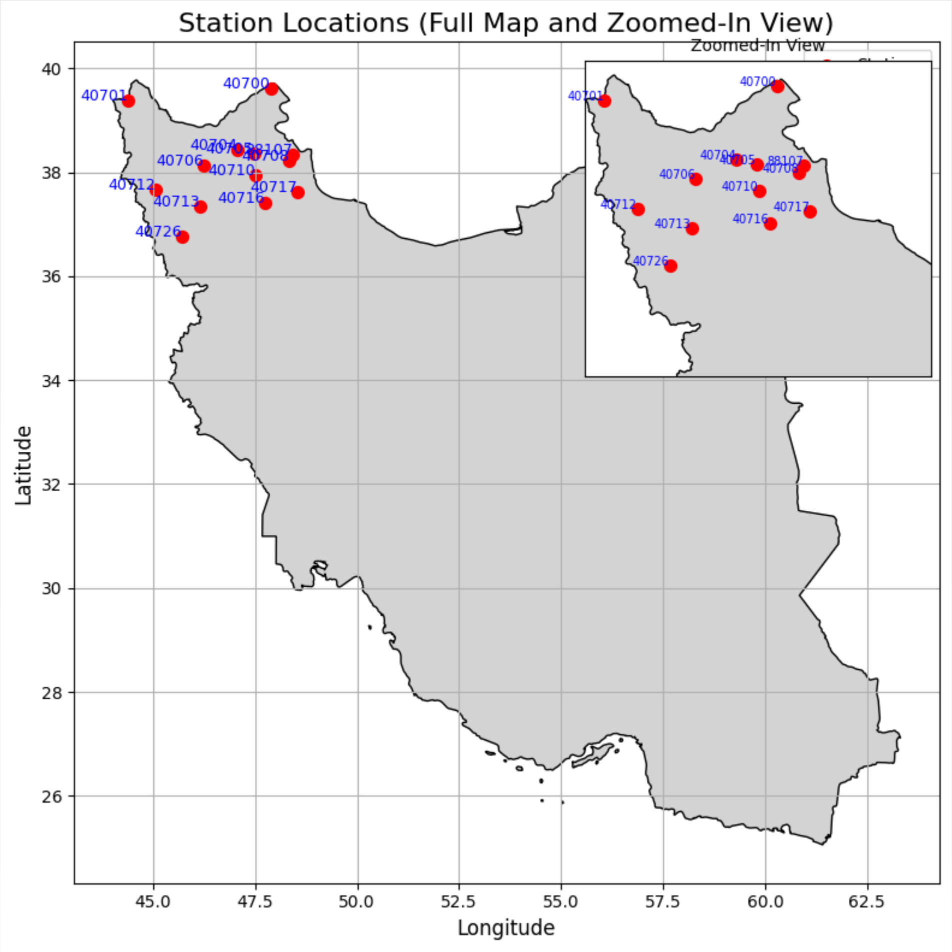
- Key Outcomes:
  - 1- & 6-Month Forecasts: Drought index-only inputs with wavelet-MLPNN perform best.
  - 12-Month Forecast: Multiparameter inputs (drought index, precipitation, temperature) with wavelet-SVR offer highest reliability.

- Implications:
  - Hybrid modeling notably enhances prediction skill, with implications for planning and mitigating drought impacts.

Proposal Flowchart


1 Read and Preprocess Data
2 Wavelet Transformation DWT
3 Lag Selection ACF PACF
4 Create Dataset
5 Define Evaluation Metrics
6 Model LSTM
7 Model SVR
8 Model MLP
9 Model EDT ExtraTrees
10 Model Random Forest
11 Model Bootstrapped RF
12 Model BiLSTM
13 Model Wavelet Boosted RF
14 Model Wavelet Boosted BiLSTM
15 Compare All Models
16 Plot Test Predictions
17 Taylor Diagram
18 Forecast 10 Years

Station Distribution



Dataset definition

Structure of Dataset

- **Dataset Overview**
  - **Time Range:** Jan 1, 1990 — Apr 1, 2024
  - **Stations:** 7 meteorological stations

Category	Features
Station Info	station_id, station_name, region_id, region_name, lat, lon, station_elevation
Temperature	tmax_m, tmax_max, tmax_min, tmin_m, tmin_min, tmin_max, ntmin_0
Rainfall & Snow	rrr24 (24-hr rainfall) , sshn (snow height)
Daily Averages	tm_m (mean temp), t_03_m, t_09_m, t_15_m (temps at 3h, 9h, 15h)

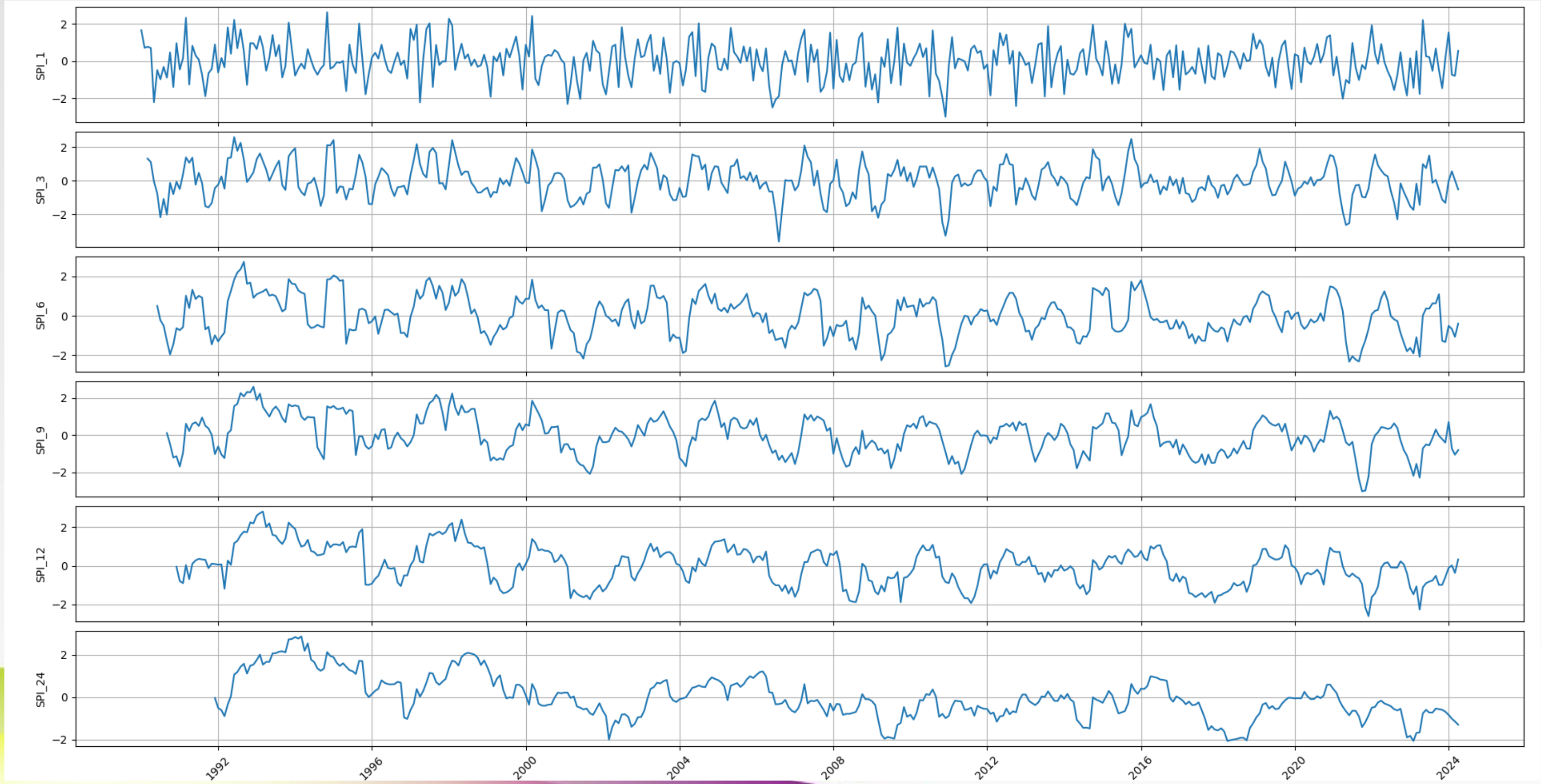
relevant columns

- rrr24: likely precipitation (24-hour rainfall in mm)
- tmax_, tmin_, tm_m, etc.: temperature indicators
- sshn: possibly snow height
- lat, lon, station_id

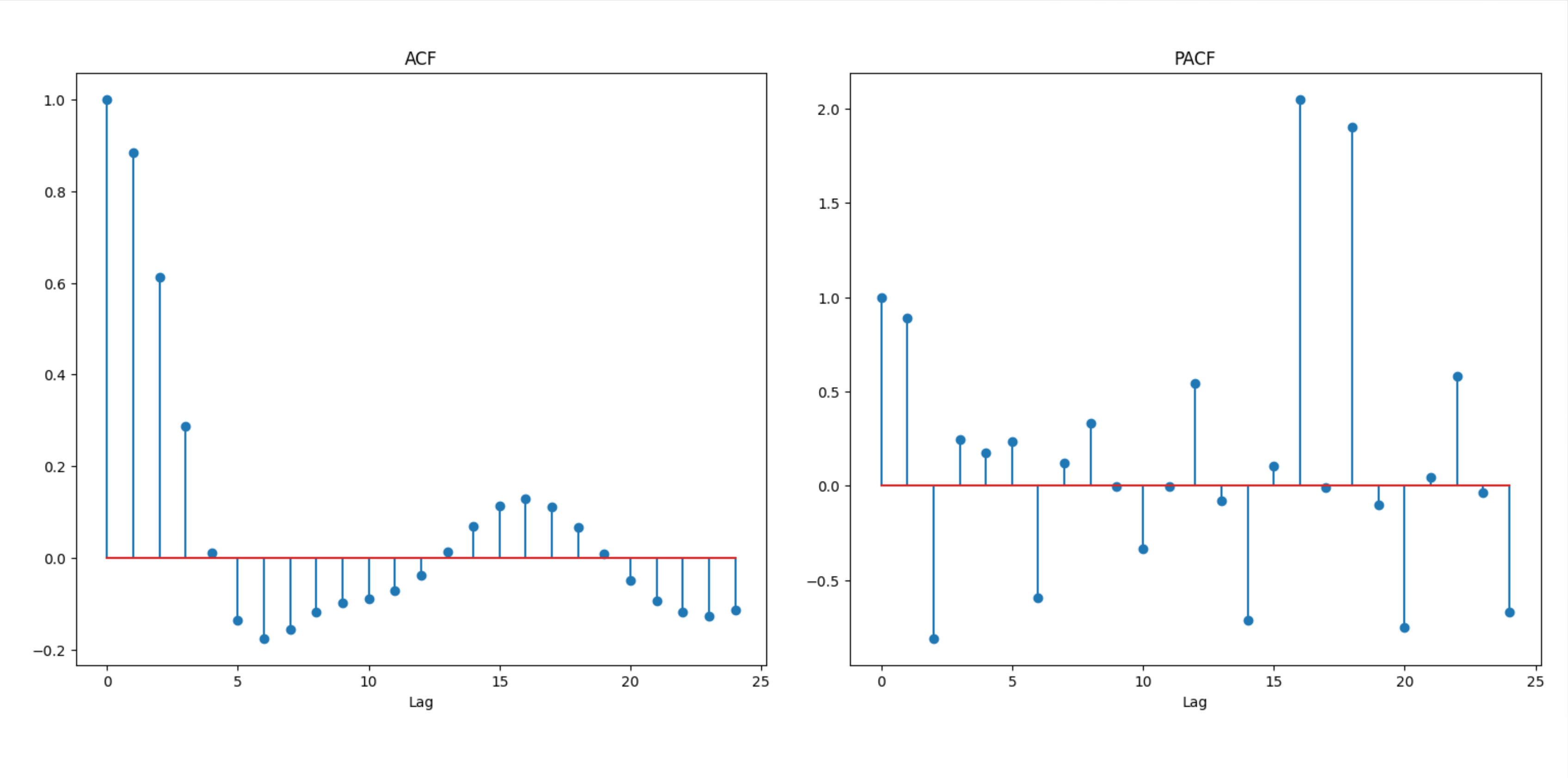
Common drought indicators

- SPI (Standardized Precipitation Index) – based only on precipitation.
- SPEI (Standardized Precipitation Evapotranspiration Index) – considers both precipitation and temperature.

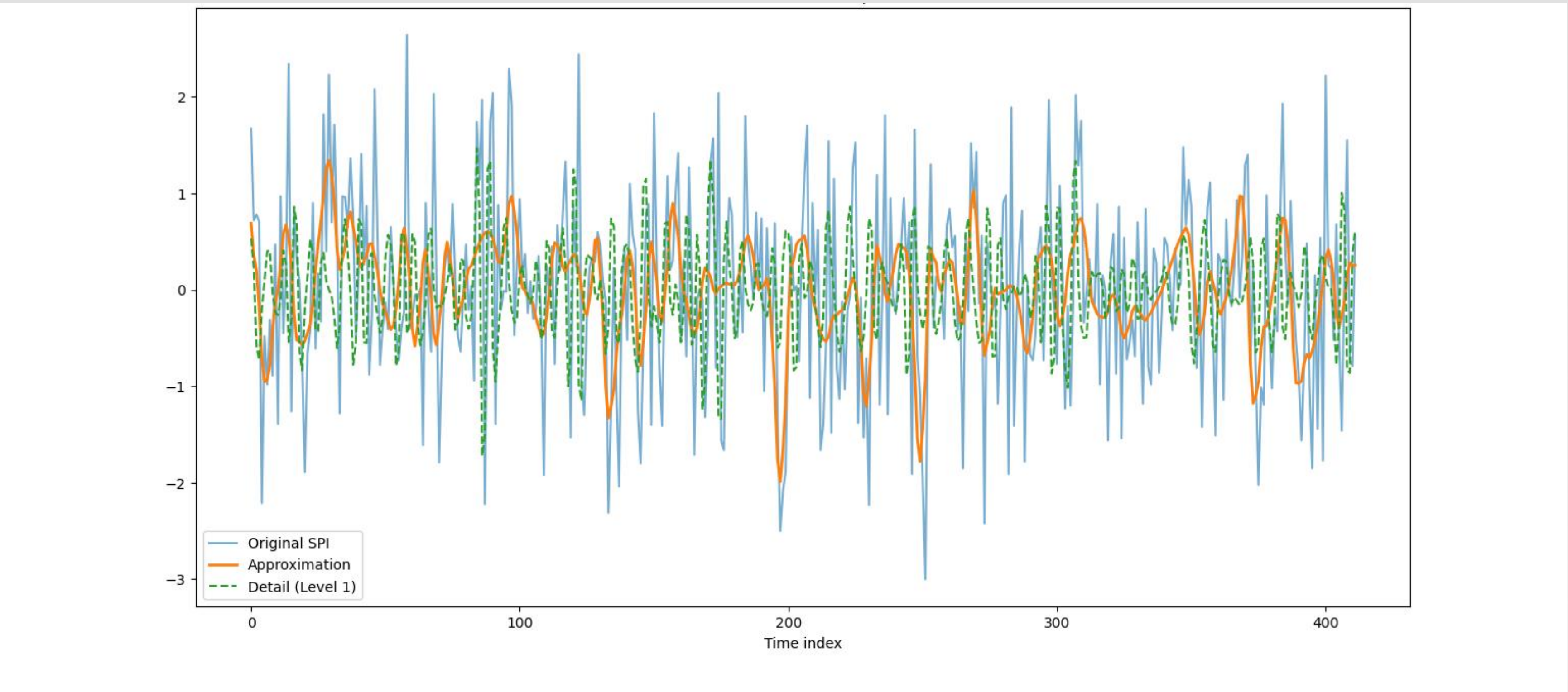
Standardized Precipitation Index (SPI) Analysis (1992–2024)



ACF and PACF



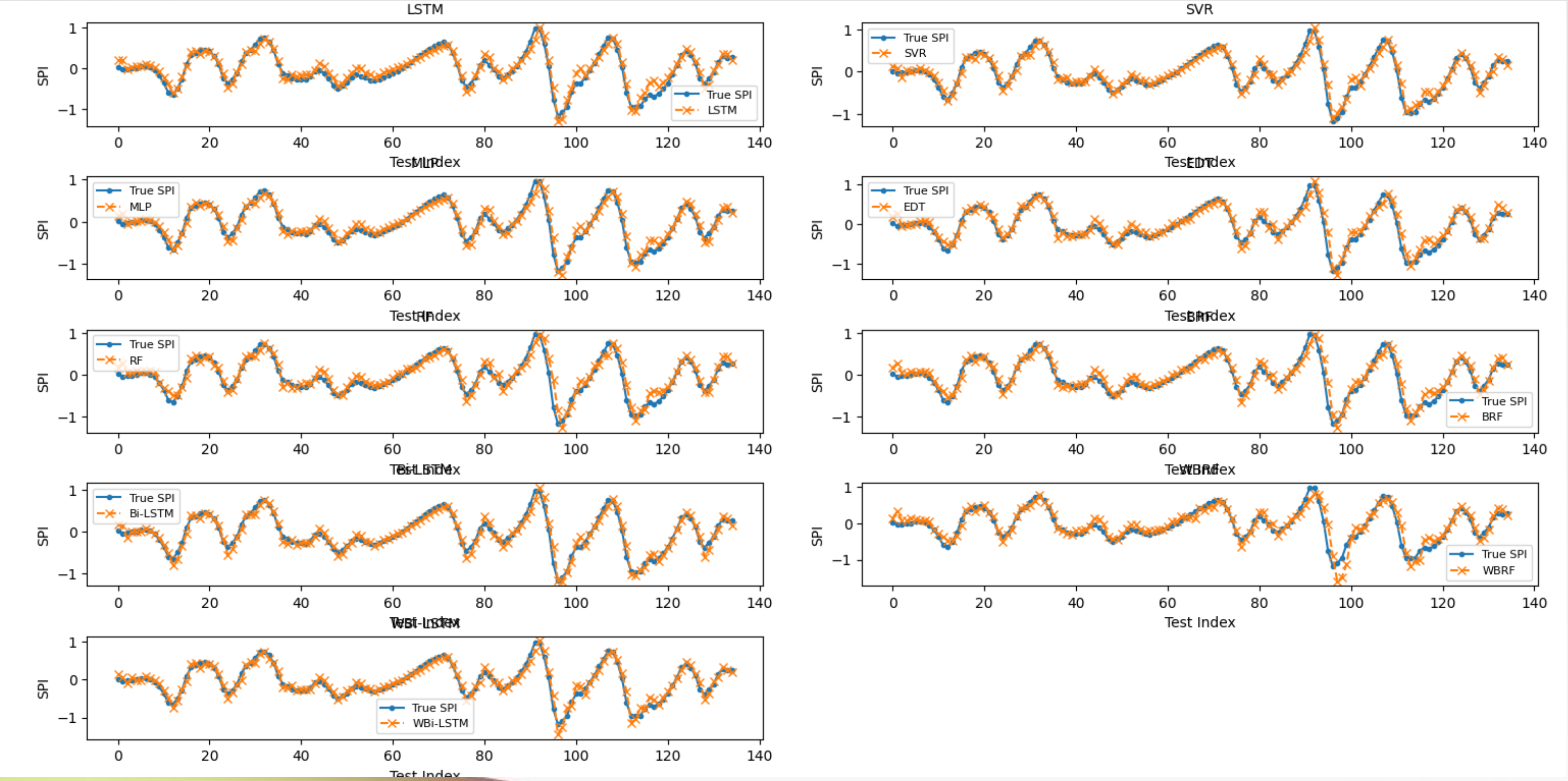
SPI with Wavelet Components



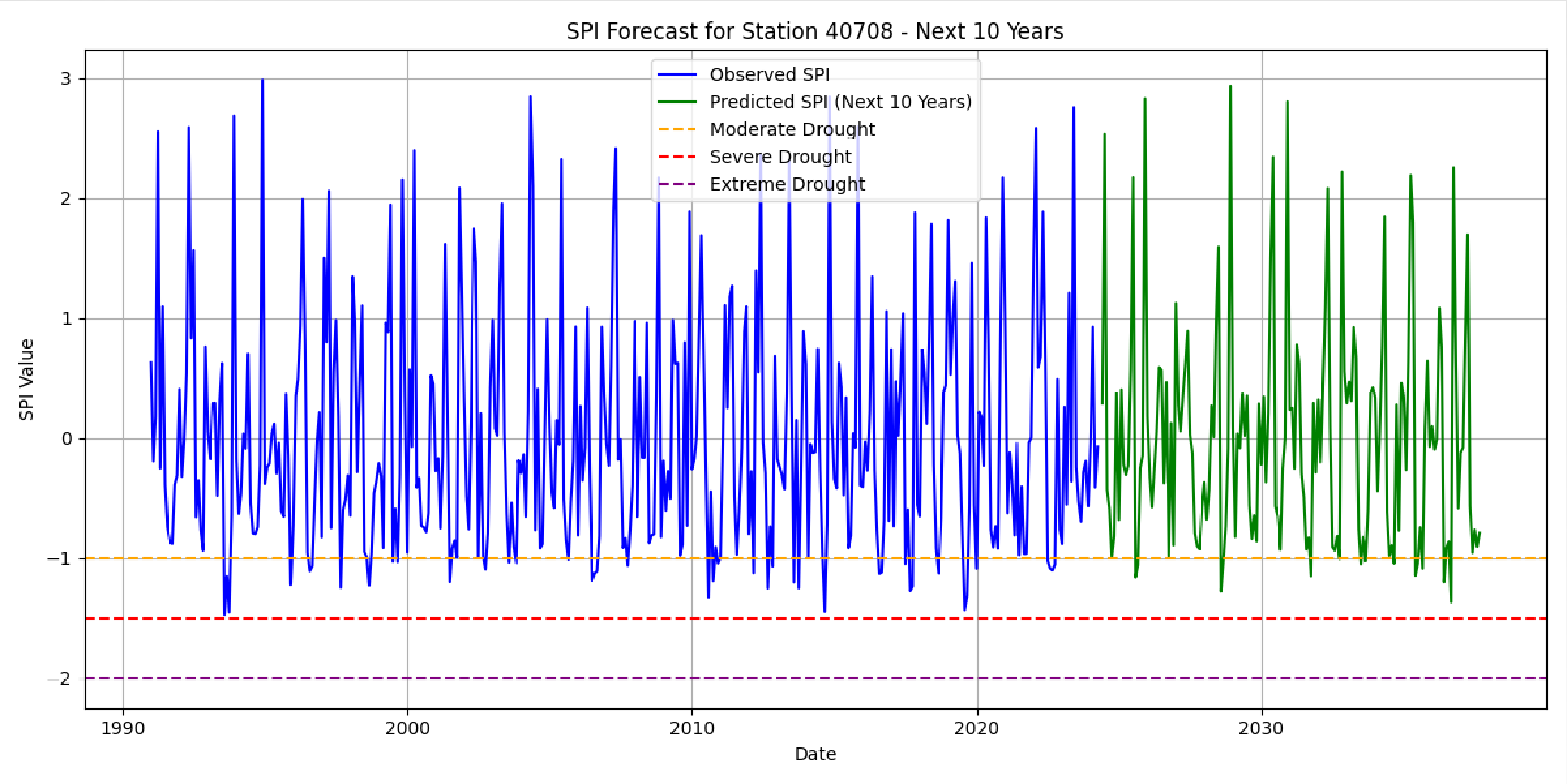
Model Comparison for SPI Prediction

- **Models Compared**
  - LSTM : Long Short-Term Memory network
  - MLP : Multi-Layer Perceptron
  - RF: Random Forest
  - Bi-LSTM : Bidirectional LSTM
  - WBi-LSTM : Weighted/Wavelet Bidirectional LSTM
  - SVR : Support Vector Regression
  - EDT : Extremely Randomized Trees
  - BRF : Boosted Random Forest
  - WBRF : Weighted Boosted Random Forest

Model Comparison for SPI Prediction



Model Comparison for SPI Prediction



Model Comparison for SPI Prediction

Among the tested approaches, Bi-LSTM, WBi-LSTM, and WBRF demonstrate strong predictive capability with close alignment to the actual SPI trends. In contrast, models like SVR and EDT show noticeable deviations, especially during rapid fluctuations. This comparison underscores the superior performance of advanced deep learning and ensemble models in capturing complex temporal dependencies in climatic data.

Taylor Diagram

It assesses machine learning and statistical models—including LSTM, SVR, MLP, EDT, RF, BRF, and Bi-LSTM—by representing three critical metrics:

- the correlation coefficient,
- the root-mean-square (RMS) difference
- the standard deviation.

The most effective model based on its proximity to the observed data, with Bi-LSTM emerging as the top performer in this case.

Taylor Diagram

- **Model Comparison**
  - Bi-LSTM: Closest to Obs (high correlation ~1.0, standard deviation ~0.4) – best performer.
  - LSTM, SVR, MLP: Lower correlations (~0.2–0.3), moderate variability (~0.3).
  - EDT, RF, BRF: Poorest performance (correlations ~0.2–0.3, standard deviations ~0.2–0.4).