

---

# Temperature data generation using quantile regression forests for different climate characteristics in Türkiye

Cenk Sezen<sup>\*1</sup>

<sup>1</sup>Ondokuz Mayıs University – Turkey

## Abstract

Stochastic weather data generation is crucial for addressing dataset gaps and making future climate projections in various disciplines. Machine learning models are being used for various purposes, including the generation of synthetic weather data and the analysis of modelling or simulation. This study aims to utilise quantile regression forests (QRF) to generate synthetic daily mean temperature data across various climatic regions. In this regard, 21-year datasets were used for the synthetic mean temperature at Samsun and Alanya stations, located in the Black Sea and Mediterranean regions of Türkiye, respectively. The 16-year and 5-year datasets were used as training and testing periods. The previous lags of minimum, maximum, and mean temperatures were used as input data for generating the mean temperature. Seasonality was also considered for the temperature data generation process. Twenty synthetic temperature datasets were generated, and the success of the generation process was evaluated using statistics (i.e., mean, standard deviation, minimum, maximum, kurtosis, skewness), distribution plots, and evaluation metrics, such as the Kling-Gupta Efficiency (KGE), Nash-Sutcliffe Efficiency (NSE), and root mean square error (RMSE). The results demonstrate that the mean temperature generation process is more promising in Alanya than in Samsun. In other words, the mean temperature datasets generated under Mediterranean climate characteristics are more satisfactory than those of Black Sea climate characteristics.

**Keywords:** Black Sea, machine learning, Mediterranean, stochastic, temperature data generation

---

<sup>\*</sup>Speaker