Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrating Explainable Machine Learning, Extreme Value Analysis, and CMIP6 Projections for Extreme Precipitation Prediction in the Eastern Mediterranean Coastal Region of Türkiye

Fatma Deniz Cüvitoğlu
Mersin University
Ali Cüvitoğlu
acuvitoglu@atu.edu.tr
Adana Science and Technology University

Research Article

Keywords: Extreme precipitation, XGBoost, SHAP, Eastern Mediterranean, CMIP6, Return periods, Türkiye

Posted Date: June 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10004357/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Theoretical and Applied Climatology on September 15th, 2026. See the published version at https://doi.org/10.1007/s00704-026-06560-w.

Integrating Explainable Machine Learning, Extreme Value Analysis, and CMIP6 Projections for Extreme Precipitation Prediction in the Eastern Mediterranean Coastal Region of Türkiye

Fatma Deniz Cüvitoğlu[1] and Ali Cüvitoğlu[2*]

[1]Department of Environmental Engineering, Mersin University, Çiftlikköy Campus, Mersin, 33343, Türkiye.

[2*]Department of Artificial Intelligence Engineering, Adana Alparslan Türkeş Science and Technology University, Sarıçam, Adana, 01250, Türkiye.

*Corresponding author(s). E-mail(s): acuvitoglu@atu.edu.tr;
Contributing authors: fatmadeniz@mersin.edu.tr;

Abstract

Extreme precipitation events in the eastern Mediterranean coastal region of Türkiye pose severe risks to densely populated provinces including Mersin, Adana, Osmaniye, and Hatay. Despite this vulnerability, the physical drivers, statistical recurrence properties, and future trajectories of precipitation extremes in this subregion have not previously been integrated within a unified analytical framework. This study presents a hybrid machine learning and statistical pipeline combining multi-source data fusion, K-Means spatial clustering, XGBoost classification with SHAP explainability, Generalised Extreme Value and Generalised Pareto Distribution return period analysis, and CMIP6 climate projections. Quality-controlled homogeneity testing of 24 Turkish State Meteorological Service stations (2010–2024) identified a statistically significant and spatially coherent precipitation regime shift beginning in 2020, with a domain-mean anomaly of -20.5% relative to the preceding decade. K-Means clustering partitioned the 24 stations into three physically interpretable precipitation regime

zones (Coastal, Inland, and Mountainous). The XGBoost classifier achieved an AUC-ROC of 0.835 on the 2023–2024 independent test set. SHAP analysis identified near-surface relative humidity as the overwhelmingly dominant predictor (mean |SHAP| = **0.557**), with a nonlinear activation threshold near 65–70%, consistent with Clausius–Clapeyron thermodynamics. Its importance intensified progressively with elevation, reflecting the orographic precipitation mechanism. GEV fitting to 44 years of CHIRPS annual maxima yielded domain-mean 100-year return levels of 102.4 mm day^{-1}, with the highest hazard concentrated in the eastern inland corridor. CMIP6 projections under SSP2-4.5 and SSP5-8.5 indicate consistent reductions in extreme event frequency (−19.5% and −12.1%, respectively) by 2041–2060, accompanied by scenario-dependent changes in wet-day intensity, presenting a "fewer but potentially more intense" precipitation future for the region.

Keywords: Extreme precipitation, XGBoost, SHAP, Eastern Mediterranean, CMIP6, Return periods, Türkiye

1 Introduction

The Mediterranean basin is defined as a "hot spot" within the scope of global climate change studies and constitutes one of the most vulnerable regions to climate change (Bağçaci et al. 2021; Zittis et al. 2021; Deepa et al. 2024). Studies indicate that the precipitation situation in this region is a "precipitation paradox." According to this paradox, the total amount of precipitation is decreasing however, due to the increasing temperature, the atmosphere holds more moisture, which leads to precipitation severe enough to cause floods and inundations (Zittis et al. 2022, 2021). This situation is supported by the Clausius-Clapeyron (CC) relationship. CC relationship predicts that for every 1°C increase in temperature, water vapor holding capacity of the atmosphere expands by approximately 7% (Martínková and Kyselý 2020). The Mediterranean basin is a region that requires attention in climate studies due to the diversity in its topographic structure. Because even in small areas, differences in surface features cause differences in the precipitation regime. For example, the Taurus Mountains and coastal regions show very different climatic characteristics (Aksu et al. 2023; Hisam et al. 2025). The Eastern Mediterranean region, encompassing the coastal

provinces of southern Türkiye, lies precisely on the border between the arid subtropical zone and the mid-latitudes with variable air movements. This critical location makes the region extremely sensitive and vulnerable to climate change (Yilmaz and Imteaz 2014). In densely populated and agriculturally active provinces such as Mersin, Adana, Osmaniye, and Hatay, short but very intense rainfall causes loss of life, collapse of urban infrastructure, and damage to vast agricultural lands. Located between the Mediterranean Sea and the Taurus Mountains, this city corridor is more vulnerable to flood risk due to its geographical structure.

Extreme rainfall in the Eastern Mediterranean is caused by a combination of both regional and global factors. While frontal weather systems coming from the sea in winter meet the annual rainfall needs of coastal regions, uplift movements on the slopes of the Taurus Mountains (orographic effect) lead to local and intense rainfall independently of general weather conditions (Mastrantonas et al. 2022a,b, 2021). The variation in rainfall in the region over the years is controlled by large-scale atmospheric circulation patterns such as the North Atlantic Oscillation (NAO) and the Atlantic Multidecadal Oscillation (AMO) (Intergovernmental Panel on Climate Change 2023; Mastrantonas et al. 2021; Krichak et al. 2014). However, these natural cycles are currently occurring with human-induced climate change. Although current climate models (CMIP6) predict significant warming and summer drought for Türkiye by the end of the century, they are much more successful in predicting winter precipitation compared to older models (Bağçaci et al. 2021).

Machine learning (ML) methods are effective in extreme precipitation prediction studies thanks to their ability to make sense of complex relationships between input and output parameters (Hess and Boers 2022; Lin et al. 2023; Salcedo-Sanz et al. 2024; Spuler et al. 2024; Hisam et al. 2025; Senjaliya et al. 2026). Today, gradient boost tree community algorithms such as XGBoost (Chen and Guestrin 2016) and LightGBM (Ke et al. 2017) are prominent in environmental prediction studies. These algorithms

can cope with situations such as data deficiency and dataset imbalance and exhibit superior performance in meteorological datasets (Lin et al. 2023). Furthermore, in recent years, these models have been significantly correlated with spatial clustering methods, thus considering irregular rainfall regimes in heterogeneous areas such as the Mediterranean basin. It has been observed that predictions can be physically explained through post-explainability analysis performed via XGBoost (Li et al. 2020) and SHAP (SHapley Additive exPlanations; Lundberg and Lee 2017). Hisam et al. (2025) showed that more reliable results were obtained compared to a single data source with an ML-based downscaling study combining satellite data and ground station measurements with ML models such as XGBoost, LightGBM and SHAP in the Mediterranean region of Türkiye.

The Eastern Mediterranean region has not been sufficiently studied scientifically compared to similar areas in the Western Mediterranean. In particular, the physical causes triggering extreme rainfall, the frequency of these events, and what awaits us in the future have not yet been analyzed under a single scientific framework with a holistic approach. However, the integration of data from various observational sources in the coastal region of the eastern Mediterranean, regional grouping, SHAP-based explainability, gradient boosting classification, extreme value analysis and climate change projection within a single integrated system has not yet been realized. Characterizing the danger that extreme precipitation can cause is a technical challenge. However, analyses conducted for the Eastern Mediterranean region of Türkiye have been found to not account for the characteristics of precipitation regimes, not to correlate recurrence period estimates with ML-based prediction models, and to be limited to isolated station analyses. Although CMIP6 scenarios give a strong signal of drought for average precipitation (Bağçaci et al. 2021), how extreme precipitation will be received remains unclear. This situation makes it more valuable to take measurements at more points and calibrate the model with region-specific data in modeling using satellite-based

data (Aksu et al. 2023; Hisam et al. 2025). It is unclear how the difference between "change in the number of rainy days" and "rainfall intensity" will vary at the local level.

In this study, we aimed to present an integrated machine learning and statistical framework to address the shortcomings in extreme rainfall prediction, future projection, and hazard characterization in Türkiye's eastern Mediterranean coastal region (covering the provinces of Mersin, Adana, Osmaniye, and Hatay).

The main stages of the study are: (I) quality-controlled homogeneity testing of data from 2010–2024 for 24 ground stations; (II) multi-source data fusion combining ground observations, CHIRPS, and ERA5-Land with bias correction; (III) clustering of the 24 stations into three physically interpretable precipitation regime zones: coastal, inland, and mountainous; (IV) a 40-attribute predictor matrix encompassing topographic variables, local measurements, seasonal cycles, and lagged teleconnection indices; (V) XGBoost binary classification with temporal cross-validation and class imbalance correction; (VI) global and cluster-based SHAP explainability analysis; (VII) GEV (Generalized Extreme Value) and GPD (Generalized Pareto Distribution) recurrence period estimation from annual maximum values using 44 years of CHIRPS data; and (VIII) CMIP6-based extreme event frequency and wet day intensity projection under SSP2-4.5 and SSP5-8.5 scenarios for the 2041–2060 horizon using MPI-ESM1-2-LR.

This study is expected to contribute to: (1) determining the change in precipitation regime through homogeneity testing at 24 stations and its effects on the transferability of the ML model; (2) the first integrated ML–extreme value–climate projection pipeline for eastern Mediterranean coastal region of Türkiye; (3) identifying the dominant parameter in the forecast and its relationship with other parameters; and (4) the change in extreme event frequency and possible changes according to the wet day intensity scenario between 2041–2060.

2 Data and Methods

2.1 Study Area

This study covers the eastern Mediterranean basin, including the provinces of Mersin, Adana, Osmaniye, and Hatay in the Mediterranean region of Türkiye. The study area has an altitude range from sea level to 1400 m. The region exhibits Mediterranean climate characteristics, with hot and dry summers and mild and rainy winters.

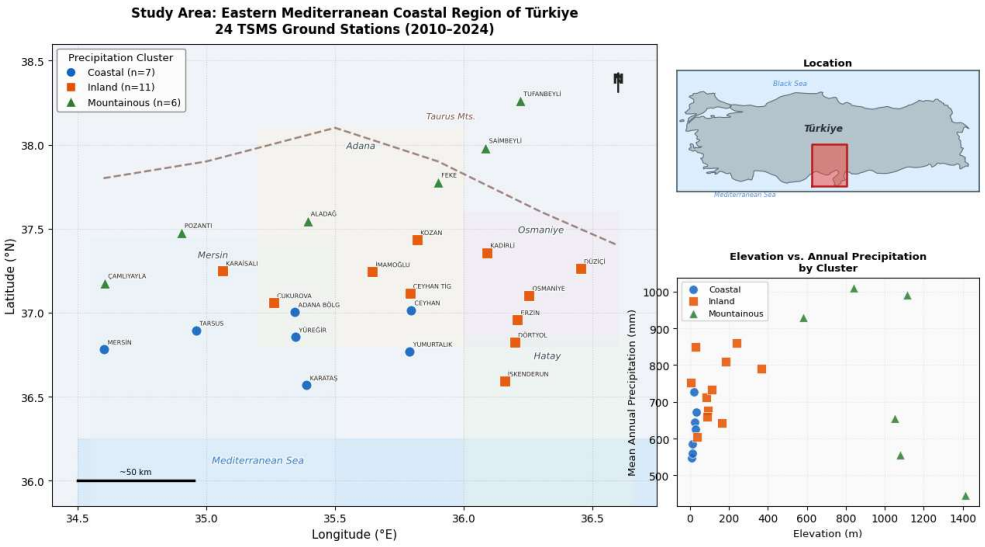


Fig. 1 Study area map showing the eastern Mediterranean coastal region of Türkiye. The 24 TSMS ground stations used in this study are displayed according to their K-Means precipitation cluster assignment: Coastal ($n = 7$), Inland ($n = 11$), and Mountainous ($n = 6$). The inset map indicates the location of the study domain within Türkiye. The scatter plot shows the relationship between station elevation and mean annual precipitation by cluster.

The spatial distribution of the 24 stations across the four provinces is illustrated in Figure 1.

2.2 Data Sources

This study aims to create a multi-source data infrastructure using multiple data sources (local station observation data, satellite-derived precipitation data, and re-analysis datasets). Details regarding the data sets used are summarized in Table 1.

Table 1 Summary of datasets used in this study.

Dataset	Variable(s)	Resolution	Period	Source	Role
TSMS ground stations	Precip., T_{mean}, RH, ET	Point (24 st.)	2010–2024	(Turkish State Meteorological Service 2024)	Ground truth
CHIRPS v2.0	Daily precip.	0.05° (∼5 km)	1981–2024	(Funk et al. 2015)	Precip. feature
ERA5-Land	2m air temp.	0.1° (∼9 km)	2009–2024	(Muñoz-Sabater et al. 2021)	Temp. feature
NOAA CPC/GML	ONI, NAO, AO, EAWR, AMO	Monthly	1981–2024	(NOAA Climate Prediction Center 2024; NOAA Global Monitoring Laboratory 2024)	Teleconnection
SRTM DEM	Elevation	30 m	—	(Farr et al. 2007)	Topography
E-OBS v30.0e	T_{mean}, precip.	0.25°	1980–2024	(Cornes et al. 2018; Copernicus Climate Data Store 2024)	Climatol. context

2.2.1 Local station observation data

Station-based daily total precipitation (mm), average, minimum and maximum temperature (°C), relative humidity (RH) (%) and evapotranspiration (ET) data were obtained from the General Directorate of Meteorology of Türkiye (TSMS) (Turkish State Meteorological Service 2024).

2.2.2 Satellite-derived data

Satellite-derived precipitation data were obtained from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) version 2.0 dataset (Funk et al. 2015), using daily precipitation data with a spatial resolution of 0.05°.

2.2.3 Reanalysis data

Hourly 2-meter air temperature (T2m) data with a spatial resolution of 0.1° was obtained from the ERA5-Land dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) (Muñoz-Sabater et al. 2021). Large-scale climate index data (Ocean Niño Index (ONI), North Atlantic Oscillation (NAO), Niño 3.4 sea surface temperature anomaly (NINO34), Arctic Oscillation (AO), Atlantic Multidecadal Oscillation (AMO), East Atlantic-West Russia pattern (EAWR)) were obtained from the records of the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC), Global Monitoring Laboratory (GML), and Physical Sciences Laboratory (PSL) for the period 1981–2024 (NOAA Climate Prediction Center 2024; NOAA Global Monitoring Laboratory 2024). 30 m DEM data was used for topographic information (Farr et al. 2007). Daily average temperature and daily total precipitation data covering the period 1980–2024 with a spatial resolution of 0.25° × 0.25° were obtained from the E-OBS v30.0e dataset (Cornes et al. 2018; Copernicus Climate Data Store 2024).

2.3 Station Quality Control and Homogeneity Testing

A two-stage quality control process was conducted to ensure the reliability of the data obtained from the ground stations.

In the first stage, datasets containing more than 5% of missing data were excluded from the study. After this filtering, stations with fewer than 10 valid years were also

excluded as they would not be sufficient for the homogeneity test. As a result of this filtering, the number of suitable stations was reduced to 24.

The second stage consists of homogeneity tests. The annual total precipitation data of the stations remaining after the first filtering were tested using the Standard Normal Homogeneity Test (SNHT; Alexandersson 1986), Pettitt test (Pettitt 1979), and Buishand Range test (Buishand 1982). The condition of obtaining statistically significant results in at least two tests was applied for heterogeneity.

Following the homogeneity test, breakpoint analysis was performed. Only stations with uninterrupted annual series (i.e., without missing years) were used in this analysis. Therefore, 8 stations with one or two missing years of data were excluded from the breakpoint analysis, and the analysis was performed for 16 stations. This approach was taken to protect the data used in model training from deviations caused by artificial factors (station location changes, equipment changes, etc.). Three tests (SNHT, Pettitt, and Buishand Range Test) used in the homogeneity analysis were applied to determine the breakpoint. Each test identifies the year in which the series experienced the greatest change as the breakpoint and reveals the statistical significance of this change through a p-value ($\alpha = 0.05$). Stations meeting the $p \leq 0.05$ condition in at least one test were classified as having a significant breakpoint.

2.4 Multi-Source Data Fusion

In this section, datasets from different sources were combined to create a new daily dataset. The process consists of four steps: spatial inference, sea-masking correction, bias correction, and missing value filling.

In the spatial inference step, the k-d tree algorithm was applied to match CHIRPS and ERA5-Land grid cells for 24 stations.

Sea-masking correction was used to adjust for the Mersin, Dörtyol, Yumurtalık, and Karataş stations, which are located near the sea in the ERA5-Land cell, a land

product. The closest land cell was assigned by applying the sea-masking k-d tree algorithm.

Additional bias correction was applied to eliminate the observed deviation between the average temperature values of TSMS and ERA5-Land.

Short-term gaps observed in the data obtained from TSMS stations were filled using linear interpolation, and long gaps were filled using forward and backward filling methods.

As a result of these processes, a new dataset was created for 24 stations covering the period 2010–2024 and 6 parameters (Table 2 and Table 3).

Table 2 Descriptive statistics of the master dataset (n = 112,586 station-day records, 24 stations, 2010–2024).

Variable	Mean	SD	Min	Max	Units
MGM precipitation	1.93	7.15	0.00	192.0	mm day^{-1}
MGM mean temperature	18.51	7.84	−17.60	46.70	°C
MGM relative humidity	63.65	15.95	2.00	100.0	%
CHIRPS precipitation	1.89	5.66	0.00	106.3	mm day^{-1}
ERA5-Land t2m (raw)	17.10	8.03	−16.35	35.74	°C
ERA5-Land t2m (corrected)	18.51	8.10	−13.59	37.54	°C

2.5 Spatial Clustering and Regionalization

The study area exhibits heterogeneous characteristics in terms of precipitation regime. To prevent errors that may arise from this, the study area (24 stations) was divided into homogeneous sub-regions using unsupervised clustering. The optimal number of clusters was validated using silhouette analysis (Rousseeuw 1987). Stations were defined with five parameters (annual total precipitation, annual average temperature, summer daily precipitation, winter/summer precipitation ratio, and average altitude). These five parameters were standardized with z-score normalization. The region was divided into three sub-regions: coastal, inland, and mountainous. These labels were included as categorical variables in the feature matrix during the ML process.

Table 3 ERA5-Land bias correction by station.

Station	ERA5 mean (°C)	TSMS mean (°C)	Bias (°C)	r	RMSE (°C)
Mersin	16.94	20.93	−3.99	0.991	4.10
Adana Bölge	19.04	20.24	−1.20	0.994	1.54
Osmaniye	16.79	19.35	−2.56	0.986	2.90
İskenderun	17.65	21.36	−3.72	0.986	3.91
Ceyhan Tiğem	19.09	18.79	+0.31	0.987	1.26
Kadirli	17.64	20.16	−2.52	0.989	2.82
Kozan	17.52	20.61	−3.09	0.981	3.52
Pozantı	11.00	13.95	−2.95	0.983	3.34
Karaisali	16.10	19.33	−3.23	0.987	3.56
Ceyhan	18.74	19.14	−0.40	0.991	1.10
Dörtyol	18.23	20.22	−1.99	0.984	2.37
Tarsus	19.32	19.22	+0.10	0.992	0.99
Yumurtalık	19.29	20.17	−0.87	0.990	1.28
Karataş	20.19	19.95	+0.24	0.990	1.02
Tufanbeyli	13.77	10.85	+2.92	0.977	3.59
Yüreğir/Çukurova	19.98	19.11	+0.87	0.987	1.46
İmamoğlu	18.43	19.26	−0.83	0.991	1.32
Saimbeyli	14.54	14.45	+0.09	0.991	1.15
Çamlıyayla	11.40	12.76	−1.36	0.990	2.11
Düziçi	14.82	18.71	−3.89	0.980	4.28
Aladağ	14.50	14.72	−0.22	0.991	1.13
Çukurova	18.72	19.95	−1.22	0.990	1.72
Feke	15.55	16.31	−0.76	0.984	1.70
Erzin	18.56	19.64	−1.08	0.987	1.66
Domain Mean	**17.10**	**18.41**	**−1.31**	**0.987**	**2.24**

2.6 Teleconnection Feature Engineering

Six basic teleconnection indices (ONI, NINO34, NAO, AO, AMO, and EAWR) were selected as predictor variables to include large-scale atmospheric and oceanic circulations in the model (NOAA Climate Prediction Center 2024; NOAA Global Monitoring Laboratory 2024). Lagged features were calculated for these indices, corresponding to 1, 2, 3, and 6 months prior to the target date. A total of 30 teleconnection features (6 simultaneous and 24 lagged) were combined with the daily dataset. To see the effect of teleconnection signals on local precipitation forecasting and to test its statistical significance, a Spearman test ($\alpha = 0.05$) was performed between the monthly total precipitation data and the indices (Zar 2005).

The Spearman correlation test is a non-parametric test that determines whether there is a strong monotonic (increasing or decreasing) relationship between variables. The Spearman correlation coefficient is calculated using the following formula:

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{1}$$

where r_s is the Spearman correlation coefficient, d_i is the rank difference for each data pair (x_i and y_i), and n is the number of observations (Hamed 2016; de Freitas and Abrahão 2025). The value of r_s varies between -1 and $+1$, where -1 and $+1$ indicate perfect negative and positive correlations, respectively (Hamed 2016; Tariq et al. 2025). If the p-value is lower than 0.05, the correlation is statistically significant (Hinkle et al. 2003; Mukaka 2012).

All teleconnection features were stored in a feature matrix to allow the ML model to learn complex and nonlinear relationships with its own internal algorithm. The final significance ranking was left to the SHAP analysis.

2.7 Extreme Precipitation Labeling

In this study, the term extreme precipitation was defined using the Peaks Over Threshold (POT) approach, defined as exceeding the 95th percentile (Klein Tank et al. 2002). According to this approach, all observations exceeding a certain threshold value are sampled, and the resulting data are fitted to exponential or GPD (Cunnane 1973; Madsen and Rosbjerg 1997; Beguería and Vicente-Serrano 2010). In this study, for the stations, the cases where the observed precipitation in the daily precipitation distribution for the period 2010–2024 is equal to or greater than the P_{95} threshold are labelled "Extreme" (1), and other cases are labelled "Normal" (0).

The 2010–2022 period was used as the training set and the 2023–2024 period as the test set in the model. To address the 1:18 class imbalance problem in the training set, SMOTE and inverse frequency class weighting (`scale_pos_weight`) methods

were compared. The class weighting strategy was adopted in the final model training because it showed more consistent and higher performance in F1 scores. The predictive feature matrix consists of a total of 40 variables derived from five main groups: (I) meteorological variables (TSMS observations and ERA5-Land data), (II) topographic and geographic variables, (III) temporal encodings, (IV) spatial cluster label (information on the climatological subregion to which the stations belong), and (V) teleconnection features: 30 different lagged (1, 2, 3, and 6 months) combinations of six basic climate indices.

2.8 ML Model

In this study, four different classification algorithms were used: three base learners — XGBoost (Chen and Guestrin 2016), Random Forest (Breiman 2001), and LightGBM (Ke et al. 2017) — and one stacking ensemble trained with a meta-learner.

XGBoost hyperparameters were adjusted with 30 iterations of RandomizedSearchCV. The optimum parameters are as follows: `max_depth`= 8, `learning_rate`= 0.03, `n_estimators`= 500, `subsample`= 0.8, `colsample_bytree`= 0.8, `min_child_weight`= 3, `reg_alpha`= 1.0, `reg_lambda`= 1.0, `gamma`= 0.

Since the dataset is a time series, model performance was evaluated with five-fold TimeSeriesSplit cross-validation. For each fold, the optimal classification threshold was determined as the value that maximizes the F1 score on the validation set's precision–recall curve, and the average cross-validation F1 score across the five temporal folds was obtained as 0.403 ± 0.040.

The prediction probabilities of XGBoost, LightGBM, and Random Forest were used as input features. A three-model stacking ensemble was created with a logistic regression meta-learner trained with `class_weight='balanced'`. The final model performance was evaluated on an independent test set (2023–2024) using four metrics: F1 score, area under the ROC curve (AUC-ROC), precision, and recall. The decision

threshold was not set to a default of 0.5; instead, it was adjusted to the value that maximizes F1 on the precision–recall curve.

2.9 Model Explainability (XAI) and SHAP Analysis

The SHAP method was applied to relate the prediction mechanisms of ML models to physical processes (Lundberg and Lee 2017). Using the TreeExplainer algorithm, which is optimized for tree-based models, the marginal contributions of features were calculated for 15,645 test records. Global feature significance was measured with the mean absolute SHAP values (mean |SHAP|) over all observations; directional effects (positive/negative contribution) were visualized through Beeswarm plots. In addition, to test the spatial sensitivity of the models, SHAP profiles were stratified based on Coastal, Inland, and Mountainous clusters. Nonlinear interactions between the two most dominant variables, relative humidity and mean temperature, were characterized with SHAP dependency plots.

2.10 Return Period Analysis

The recurrence levels of extreme rainfall were analyzed over a 44-year period. The CHIRPS dataset (1981–2024) was analyzed using two complementary methods.

Block Maxima/GEV approach: GEV distribution was fitted to the annual maximum precipitation series; 10, 20, 50, and 100-year recurrence levels were calculated (Coles 2001). The GEV distribution encompasses three limiting forms characterised by the shape parameter ξ: the Gumbel ($\xi = 0$), Fréchet ($\xi > 0$, heavy-tailed), and Weibull ($\xi < 0$, bounded) distributions.

POT/GPD approach: As a complementary method, the GPD was fitted to threshold exceedances using the POT method. GPD was applied to values exceeding the P_{95} threshold. This method was used to simultaneously assess the frequency and magnitude of rare events.

The two methods provide independent estimates of extreme rainfall return levels: the Block Maxima approach uses annual maximum observations and is more robust for long return periods, while the POT approach uses all observations above a certain threshold. Comparing the two methods provides a consistency check on the tail behaviour of the fitted distributions.

2.11 Future Climate Projections

Future extreme precipitation trends were assessed using the daily outputs of the MPI-ESM1-2-LR model from the CMIP6 archive (Mauritsen et al. 2019; Eyring et al. 2016; Google Cloud 2024). Two different scenarios were considered in the analysis: moderately forcing SSP2-4.5 and highly forcing SSP5-8.5 (O'Neill et al. 2016). For both scenarios, the r1i1p1f1 ensemble member was used. Historical simulation of the same model (1981–2014) and ensemble member formed the basis for bias correction and change detection.

CMIP6 data were subjected to bias correction using the Delta Method to align them with historical CHIRPS observations (Maraun 2016). The correction factor, found to be an average of 1.48, eliminated the model's tendency to systematically underestimate regional precipitation. Two main indicators were calculated for the 2041–2060 horizon based on the corrected data:

1. *Extreme Frequency:* The number of annual days exceeding the P_{95} threshold.
2. *Wet Day Intensity:* The average daily precipitation amount on wet days (pr $>$ 1 mm).

3 Results

3.1 Analysis of Climatic Parameters

Daily precipitation data (1981–2024) were analyzed using daily CHIRPS 0.25° data collected via `xarray` and `dask`. As a result of the analyses, significant differences were observed in the average precipitation height in the coastal and inland areas of the region (Figure 2a). The average precipitation was calculated to be approximately 300 mm yr^{-1} in low-lying areas and 700 mm yr^{-1} in mountainous regions. The average precipitation value for the entire area over 44 years is 412 mm. The difference between the average precipitation value for the 2011–2024 period and the average precipitation value for the 1981–2010 period was examined (Figure 2b). Positive values indicate that the last decades have been wetter. It was observed that the annual average precipitation was higher in coastal regions but lower in inland regions during the 2011–2024 period. The 95th and 99th percentiles of extreme precipitation were calculated as 11.3 and 29.4 mm day^{-1}, respectively (Figures 2c–d). An average value of 0.57 was obtained for the dry-day fraction, expressed as the fraction of rainless days. This ratio was calculated to be higher in inland areas and lower in coastal areas (Figure 2e). Trend analysis revealed a trend of -0.05 mm yr^{-1}, but the result was not statistically significant (Figure 2f).

Precipitation and temperature variation analysis of the study area was performed using E-OBS v30.0e gridded observation data in addition to CHIRPS data (Cornes et al. 2018; Copernicus Climate Data Store 2024). The average annual temperature in the region was calculated as 17.29°C, and the average annual total precipitation as 260 mm yr^{-1} (Figures 3a–b). The average precipitation obtained was lower than the average obtained with CHIRPS data (412 mm yr^{-1}). Since this difference may be due to station density and satellite-measurement methods, the direction of trends and relative changes were used as the primary comparison measure in the study.

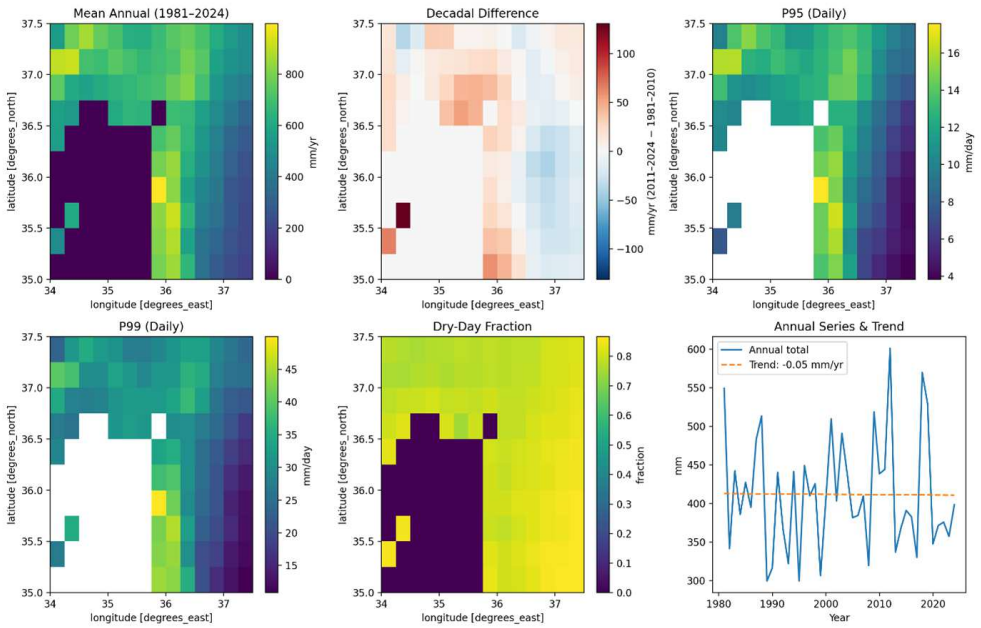


Fig. 2 Temporal and spatial analysis of daily precipitation according to CHIRPS data for the period 1981–2024 along the Eastern Mediterranean. (a) Mean annual precipitation (mm yr^{-1}); (b) decadal difference between precipitation for the periods 2011–2024 and 1981–2010 (mm yr^{-1}); (c–d) daily precipitation thresholds for the 95th and 99th percentiles (mm day^{-1}); (e) spatial distribution of the dry-day fraction; (f) variation and linear trend of annual total precipitation.

Decadal differences were examined for both temperature and precipitation (Figures 3c–d). The decadal changes in precipitation and temperature values in the study area differed within the region. Although the average temperature increased throughout the area, the increase reached 1.8°C in the south and northwest. The average precipitation value differed from the findings obtained with CHIRPS data. An increase was detected in the inner parts of the south of the study area, while decreases of up to 100 mm were detected in other areas. Trend analysis revealed a statistically significant increase in temperature (+0.06°C yr^{-1}) and a decrease in precipitation (−1.1 mm yr^{-1}). A weak negative correlation ($r = -0.21$) was found between temperature and precipitation (Figure 3f). The monthly average temperature–precipitation anomaly distribution also revealed this negative correlation (Figure 3f).

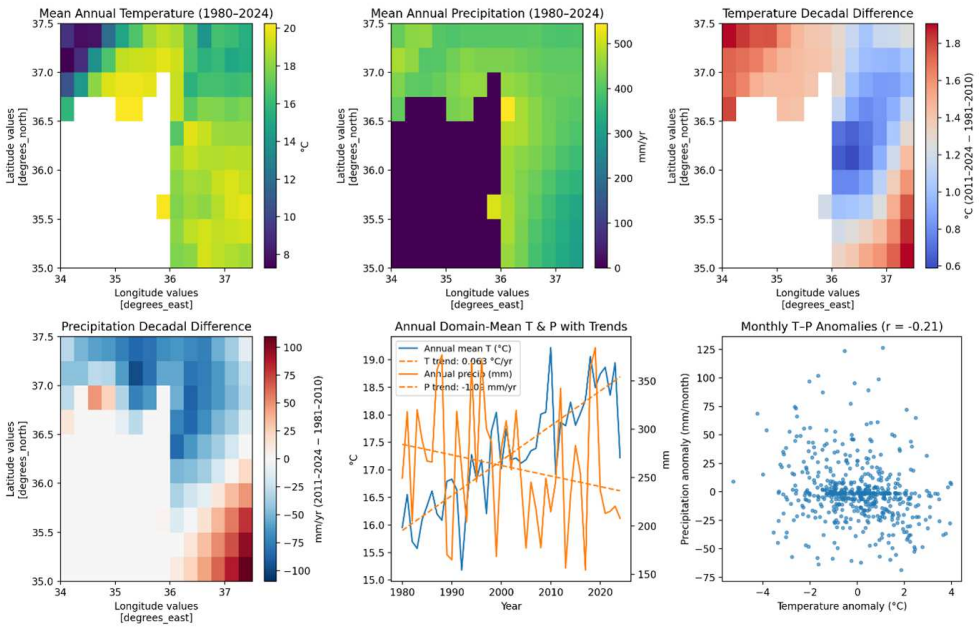


Fig. 3 Temporal and spatial analysis of daily precipitation and temperature according to E-OBS (v30.0e) data for the period 1980–2024 along the Eastern Mediterranean coastal region. (a) Mean annual temperature (°C); (b) mean annual precipitation (mm yr^{-1}); (c) decadal difference between temperature for the periods 2011–2024 and 1981–2010 (°C); (d) decadal difference between precipitation for the periods 2011–2024 and 1981–2010 (mm yr^{-1}); (e) annual domain-mean temperature and precipitation with linear trends; (f) monthly temperature–precipitation anomalies ($r = -0.21$).

The results of the analysis performed with daily T_{mean}, T_{max}, T_{min}, RH, and ET datasets from TSMS and precipitation (pr_{man}, pr_{omg}) datasets obtained from both manual and automatic observation systems (1981–2021) are given in Figure 4 and Table 4. According to local station data, the average, maximum and minimum temperature increase was 0.036, 0.042 and 0.044°C yr^{-1}, respectively (Figures 4a–c and Table 4). The average relative humidity showed an annual decrease of 0.143%. This decrease was also statistically significant (Figure 4d). The ET value, which gives the sum of evaporation and transpiration, showed a weak trend of increasing by 0.03 mm day^{-1} annually (Figure 4e). These findings indicate a warming and drying trend and increasing demand for humidity. Average rainfall from manual and automatic

measurements showed a slight decrease, with trends of −0.017 and −0.029 mm yr^{-1}, respectively (Figures 4f–g).

Table 4 Long-term annual means and linear trends of T_{mean}, T_{max}, T_{min}, RH, ET, and precipitation obtained from TSMS daily station records (1981–2021).

Variable	Mean	SD	Trend (yr^{-1})	R^2	Unit
T_{mean}	19.3	0.7	+0.036	0.54	°C
T_{max}	25.0	0.9	+0.042	0.61	°C
T_{min}	14.9	0.8	+0.044	0.59	°C
RH	67.2	4.5	−0.143	0.48	%
ET	4.7	0.5	+0.030	0.31	mm day^{-1}
pr_{man}	9.8	1.2	−0.017	0.22	mm day^{-1}
pr_{omg}	2.1	0.4	−0.029	0.25	mm day^{-1}

3.2 Station Quality Control and Homogeneity Testing

For homogeneity analysis, the Standard Normal Homogeneity Test (SNHT; Alexandersson 1986), Pettitt test (Pettitt 1979), and Buishand Range test (Buishand 1982) were applied to the precipitation data of the 24 stations remaining after filtering (Table 5). The homogeneity test results showed that all stations were homogeneous, indicating that the data from the TSMS records are reliable enough to be used in extreme precipitation analyses. The Pettitt test revealed statistically significant breakpoints in five stations (Osmaniye, Ceyhan TİGEM, Dörtyol, Düziçi, and Erzin). No statistically significant differences were observed in the SNHT and Buishand tests for the same stations. The p-values in the SNHT test ranged from 0.065 to 0.954, while the p-values in the Buishand test ranged from 0.113 to 0.907. Therefore, these five stations were not classified as heterogeneous. The indication that 2020 was a turning point showed spatial consistency for the three provinces included in the study (Adana, Hatay, and Osmaniye), which constitutes evidence for a genuine hydroclimatological change.

Breakpoint analysis revealed that in 13 out of 16 stations, none of the tests exceeded the $p \leq 0.05$ threshold. At least one statistically significant breakpoint was detected in

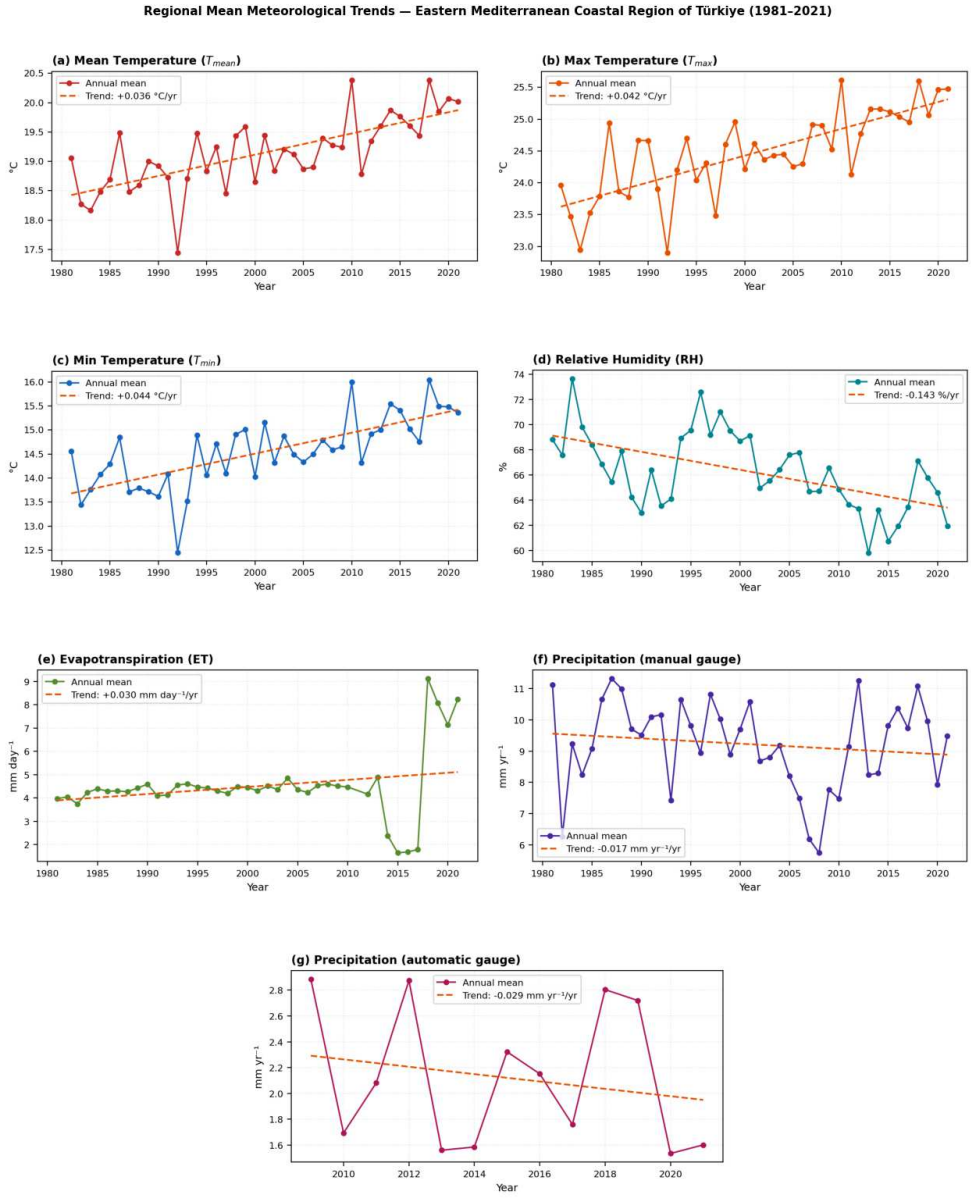


Fig. 4 Temporal analysis of (a) T_{mean}, (b) T_{max}, (c) T_{min}, (d) relative humidity (RH), (e) evapotranspiration (ET), (f) manual precipitation, and (g) automatic precipitation according to data obtained from TSMS for the period 1981–2021 along the Eastern Mediterranean coastal region. Dashed orange lines indicate linear trends.

Table 5 Homogeneity test results for the 24 TSMS stations. Asterisks (*) indicate statistically significant Pettitt test results ($p \leq 0.05$). n: number of valid years used in the analysis.

Station	Province	n	SNHT p	Pettitt p	Buishand p	Decision
Mersin	Mersin	15	0.899	0.676	0.534	Homogeneous
Adana Bölge	Adana	14	0.711	0.784	0.690	Homogeneous
Osmaniye	Osmaniye	14	0.202	0.037*	0.194	Homogeneous
İskenderun	Hatay	15	0.954	0.764	0.907	Homogeneous
Ceyhan TİGEM	Adana	15	0.087	0.028*	0.249	Homogeneous
Kadirli	Osmaniye	14	0.735	0.313	0.115	Homogeneous
Kozan	Adana	15	0.339	0.175	0.569	Homogeneous
Pozantı	Adana	13	0.251	0.216	0.637	Homogeneous
Karaisali	Adana	15	0.647	0.370	0.615	Homogeneous
Ceyhan	Adana	14	0.450	0.919	0.875	Homogeneous
Dörtyol	Hatay	15	0.078	0.012*	0.201	Homogeneous
Tarsus	Mersin	15	0.655	0.291	0.429	Homogeneous
Yumurtalık	Adana	15	0.758	0.768	0.823	Homogeneous
Karataş	Adana	14	0.907	0.740	0.524	Homogeneous
Tufanbeyli	Adana	11	0.495	0.803	0.386	Homogeneous
Yüreğir/Çukurova	Adana	11	0.762	0.227	0.159	Homogeneous
İmamoğlu	Adana	12	0.690	0.690	0.227	Homogeneous
Saimbeyli	Adana	10	0.065	0.523	0.215	Homogeneous
Çamlıyayla	Mersin	12	0.274	0.135	0.113	Homogeneous
Düziçi	Osmaniye	11	0.219	0.024*	0.272	Homogeneous
Aladağ	Adana	10	0.281	0.098	0.369	Homogeneous
Çukurova	Adana	10	0.820	0.796	0.304	Homogeneous
Feke	Adana	10	0.303	0.643	0.834	Homogeneous
Erzin	Hatay	10	0.087	0.000*	0.118	Homogeneous

* $p \leq 0.05$ (statistically significant in Pettitt test).

Ceyhan TİGEM ($p = 0.028$), Dörtyol ($p = 0.012$), and Erzin ($p = 0.000$) stations, with 2020 being the breakpoint year for all three. At Ceyhan TİGEM station, the average annual rainfall before 2020 was 672.2 mm, showing a 32.7% decrease. For Dörtyol, the average annual rainfall before 2020 was 945.1 mm, decreasing by 33.5% to 628.0 mm. Erzin showed the strongest breakpoint ($p = 0.000$), with an annual rainfall decrease from 935.7 mm to 681.6 mm.

According to the Pettitt test, the breakpoint in 8 out of 16 stations coincides with the year 2020, while according to the Buishand test, it coincides with 9 stations. Although not statistically significant for most stations, 2020 was identified as the breakpoint year; therefore, average annual rainfall values before and after 2020 were compared for all 24 stations (Table 6). A decrease ranging from 3% to 36% was

observed in all stations, indicating that the change in rainfall is regionally consistent. Furthermore, 2020 appears to be a turning point in the rainfall regime for the region and may be the beginning of a regional drought. When the entire study area is evaluated, the average rainfall showed a decrease of 20.5%. The highest decrease rates were obtained at Osmaniye (−35.9%), Dörtyol (−33.5%), and Ceyhan TİGEM (−32.7%) stations, which also yielded significant results in the Pettitt test. The lowest decrease was calculated for Tufanbeyli station, located at high altitude, with a rate of 2.9%. However, high-altitude stations also show significant decreases: Çamlıyayla (1,115 m, −12.9%), Pozantı (1,080 m, −26.2%), and Aladağ (840 m, −21.9%). This indicates that topographic features alone are not a determining factor.

Table 6 Breakpoint analysis results. Mean annual precipitation before and after 2020 and percentage changes by station. Stations marked with * showed a statistically significant Pettitt change point at 2020 ($p < 0.05$).

Station	Pre-2020 mean (mm)	Post-2020 mean (mm)	Change (%)
Osmaniye *	771.9	494.6	−35.9
Dörtyol *	945.1	628.0	−33.5
Ceyhan TİGEM *	672.2	452.6	−32.7
Tarsus	619.1	430.1	−30.5
Erzin *	935.7	681.6	−27.2
Pozantı	609.0	449.3	−26.2
Düziçi *	868.6	644.6	−25.8
Adana Bölge	693.0	518.0	−25.3
Yüreğir/Çukurova	652.6	497.2	−23.8
Karaisali	927.8	711.1	−23.4
Mersin	593.6	457.6	−22.9
Aladağ	1127.8	880.9	−21.9
Kozan	780.7	631.7	−19.1
Ceyhan	658.7	538.3	−18.3
Çukurova	695.5	577.3	−17.0
İmamoğlu	706.8	589.3	−16.6
Kadirli	751.3	629.8	−16.2
Tufanbeyli	470.6	398.5	−15.3
Feke	995.6	860.4	−13.6
Çamlıyayla	1047.8	913.0	−12.9
Karataş	747.2	653.2	−12.6
Yumurtalık	691.3	619.5	−10.4
İskenderun	773.7	710.8	−8.1
Saimbeyli	660.2	640.8	−2.9
Domain Mean	**766.5**	**608.7**	**−20.5**

* $p \leq 0.05$ (statistically significant Pettitt test).

When the annual average rainfall change is examined for the three stations showing a significant decrease as a result of the breakpoint analysis, an anomalous decrease is observed at Dörtyol in 2023 (Figure 5). Only 167.8 mm of rainfall was recorded at this station. When examined monthly, it is seen that there was almost no rainfall for a long period (April–November) in this year, indicating an extreme situation within the drought period after 2020.

3.3 Multi-Source Data Validation

CHIRPS precipitation data were validated using precipitation observation data obtained from TSMS as a reference. The validation process was performed using the Pearson correlation coefficient, mean bias, and root mean square error (RMSE). Tests were applied to both daily and annual data for the period 2010–2024. Validation statistics by station are given in Table 7.

In the Pearson test performed for daily data, the correlation coefficient (r) averaged 0.279, varying between stations from 0.184 (Çamlıyayla) to 0.379 (Çukurova). The correlation coefficient averaged 0.736 in the test performed with annual data, with the lowest and highest values calculated as 0.434 (Saimbeyli) and 0.912 (Çukurova), respectively. While a weak positive correlation was obtained for daily data, a higher correlation was observed for annual data, showing that CHIRPS reliably captures the interannual precipitation variability for the study area. The domain-mean bias was -0.4%. This value showed a positive trend at low altitudes and a negative trend at high altitudes. Çamlıyayla (1,115 m), Aladağ (840 m), and Feke (583 m) recorded biases of -32.2%, -39.4%, and -39.9%, respectively. However, Pozantı (1,080 m), also at high altitude, showed a positive bias of approximately $+40.4\%$.

The ERA5-Land temperature values, corrected for bias at the station level, were compared with TSMS observation data. The Pearson correlation coefficient was calculated as a minimum of 0.977. The findings show a high level of agreement between the

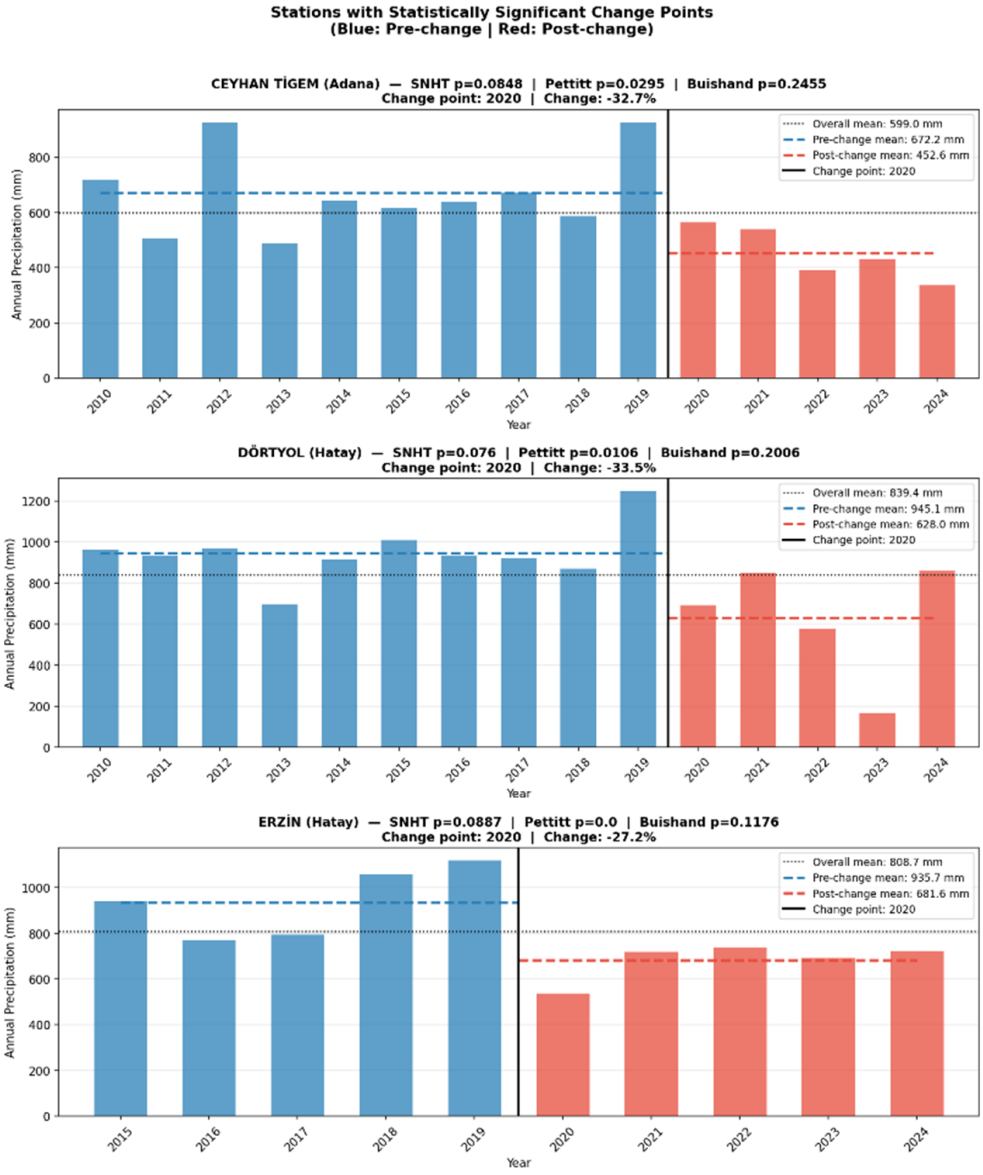


Fig. 5 Annual precipitation variation at stations with statistically significant change points (Ceyhan TİGEM, Dörtyol, and Erzin). Blue bars indicate the pre-2020 period and red bars indicate the post-2020 period. Dashed lines show pre- and post-change period means; dotted line shows the overall mean.

bias-corrected ERA5-Land and TSMS data across the study area. The RMSE value ranged from 0.99°C (Tarsus) to 4.28°C (Düziçi), with a domain mean of 2.24°C.

3.4 Spatial Clustering

K-Means clustering was applied to the feature matrix containing the variables of altitude, mean annual precipitation, mean summer precipitation, winter/summer precipitation ratio, and mean annual temperature. Three spatially consistent groups were generated (Table 8, Figure 6). The silhouette coefficient was calculated as 0.421, 0.376, and 0.435 for $k = 2$, $k = 3$, and $k = 4$, respectively. Although the coefficient obtained for $k = 4$ was slightly higher, $k = 3$ was selected because it is more consistent with the physical structure of the study area.

Three clusters consistent with the physical structure of the area were formed: the Coastal cluster ($n = 7$), the Inland cluster ($n = 11$), and the Mountainous cluster ($n = 6$). The Coastal cluster has a mean altitude of 20 m and the highest Winter/Summer precipitation ratio (17.5), indicating a strong seasonal contrast with high winter and low summer rainfall. The Inland cluster forms the transition zone between the coast and the mountains. The Mountainous cluster has a mean altitude of 1,026 m and includes the highest points of the study area, exhibiting the highest mean annual precipitation and the lowest mean annual temperature.

The box plot applied to annual precipitation by cluster shows that rainfall variability is highest in the Mountainous cluster (Figure 6b). This spatial stratification will provide the basis for cluster-resolved SHAP analysis in subsequent sections.

3.5 Teleconnection Feature Engineering

A Spearman correlation test was applied to station data with six large-scale teleconnection indices for monthly precipitation data. Statistically significant results were obtained for AMO, EAWR, and NAO (Table 9).

Table 7 Station-based validation statistics of CHIRPS daily precipitation and ERA5-Land bias-corrected daily mean temperature against TSMS ground observations. r: Pearson correlation coefficient; RMSE: root mean square error.

Station	Elev. (m)	CHIRPS r (daily)	CHIRPS r (annual)	CHIRPS bias (mm)	CHIRPS bias (%)	CHIRPS RMSE (mm)	ERA5 r	ERA5 RMSE (°C)
Çukurova	165	0.379	0.912	+0.17	+9.7	7.31	0.990	1.72
Adana Bölge	23	0.355	0.838	+0.23	+13.1	7.51	0.994	1.54
Yüreğir/Çukurova	12	0.351	0.844	+0.20	+12.3	7.45	0.987	1.46
Tarsus	12	0.343	0.907	+0.18	+11.7	7.01	0.992	0.99
Karataş	22	0.333	0.842	−0.09	−4.7	8.26	0.990	1.02
İmamoğlu	89	0.326	0.795	−0.03	−1.4	7.23	0.991	1.32
Mersin	7	0.319	0.842	+0.09	+6.0	6.83	0.991	4.10
Ceyhan	30	0.318	0.629	+0.26	+15.0	7.27	0.991	1.10
Aladağ	840	0.298	0.840	−1.09	−39.4	9.50	0.991	1.13
Ceyhan TİGEM	36	0.296	0.675	+0.30	+18.0	7.06	0.987	1.26
Kozan	112	0.292	0.756	−0.11	−5.6	7.42	0.981	3.52
Yumurtalık	34	0.286	0.658	+0.16	+8.8	8.23	0.990	1.28
Feke	583	0.277	0.635	−1.02	−39.9	8.46	0.984	1.70
Erzin	185	0.274	0.892	−0.07	−3.2	8.36	0.987	1.66
Karaisali	240	0.272	0.761	−0.32	−13.5	9.03	0.987	3.56
Kadirli	86	0.261	0.678	+0.05	+2.4	7.64	0.989	2.82
Pozantı	1080	0.244	0.593	+0.61	+40.4	7.24	0.983	3.34
Osmaniye	94	0.224	0.572	+0.27	+14.4	8.12	0.986	2.90
Tufanbeyli	1415	0.223	0.852	+0.13	+10.4	4.66	0.977	3.59
Düziçi	366	0.218	0.815	−0.18	−8.4	8.33	0.980	4.28
Dörtyol	29	0.216	0.444	−0.14	−6.1	9.17	0.984	2.37
Saimbeyli	1050	0.212	0.434	−0.19	−10.3	6.28	0.991	1.15
İskenderun	4	0.198	0.720	−0.15	−7.1	7.77	0.986	3.91
Çamlıyayla	1115	0.184	0.727	−0.87	−32.2	9.40	0.990	2.11
Domain mean		**0.279**	**0.736**	**−0.07**	**−0.4**	**7.73**	**0.987**	**2.24**

Table 8 Summary statistics of three precipitation regime clusters obtained from K-Means clustering ($k = 3$). Values represent cluster means across all stations over the 2010–2024 analysis period.

Cluster	n	Mean elevation (m)	Mean annual precip. (mm)	Winter/Summer ratio	Mean T_{mean} (°C)
Coastal	7	20	624	17.5	19.9
Inland	11	119	737	4.7	19.8
Mountainous	6	1,026	756	5.7	13.7

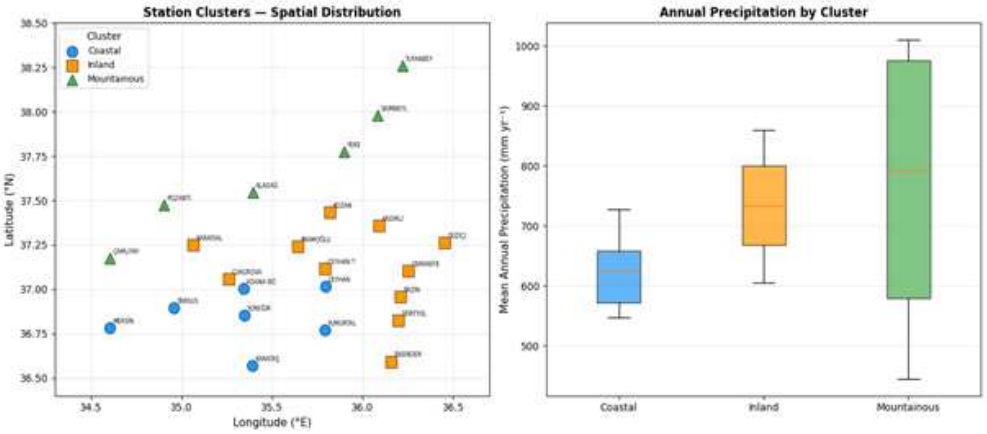


Fig. 6 Spatial distribution of station clusters (a) and inter-cluster annual precipitation variability (b) obtained from K-Means clustering ($k = 3$). Boxes indicate the interquartile range; whiskers extend to 1.5 × IQR; horizontal line indicates the median.

AMO was found to be the most dominant force influencing precipitation in the region. A strong negative concurrent correlation ($r = -0.335$) was detected between AMO and precipitation, indicating that when AMO is positive, precipitation in the study area is suppressed. Furthermore, the negative correlation coefficient ($r = -0.227$) at AMO lag 1 shows that this effect persists in the following months. However, the positive correlation coefficient at lag 6 ($r = +0.285$) is noteworthy, and can be explained by a delayed feedback loop caused by the warming ocean releasing its heat into the atmosphere.

EAWR, which represents the pressure difference between Russia and the Atlantic, can be considered the second largest force influencing precipitation in the region ($r =$

+0.299, $p < 0.001$). A positive EAWR indicates that the high-pressure system forming over Western Russia is shifting storm systems southward, increasing moisture flow to the region. A positive r at lag 2 suggests this effect continues in the following months.

NAO had a positive effect on precipitation in the region ($r = +0.208$, $p = 0.005$). A negative correlation for NAO is expected for the Mediterranean region; however, a positive correlation was observed here. It is thought that the Taurus Mountains may be responsible for this effect, as their proximity to the coast causes orographic uplift, which hinders the suppressive effect of NAO and leads to precipitation. No statistically significant relationship was found between precipitation and the ONI, NINO34, and AO indices.

Table 9 Spearman rank correlation between large-scale teleconnection indices and station-mean monthly precipitation (2010–2024, $n = 168$ months). Lag = number of months preceding the precipitation month.

Index	Lag (months)	r	p-value	Significant
AMO	0	−0.335	<0.001	Yes
EAWR	0	+0.299	<0.001	Yes
AMO	6	+0.285	<0.001	Yes
AMO	1	−0.227	0.002	Yes
NAO	0	+0.208	0.005	Yes
EAWR	2	+0.197	0.008	Yes
NAO	1	+0.166	0.026	Yes
EAWR	1	+0.145	0.052	No
NAO	3	+0.143	0.056	No
NAO	6	−0.137	0.066	No
AMO	2	−0.134	0.072	No
EAWR	6	−0.129	0.085	No
AO	6	−0.110	0.140	No
ONI	0	+0.084	0.263	No
NINO34	0	+0.070	0.350	No
AO	0	+0.081	0.277	No

3.6 Extreme Precipitation Classification Performance

3.6.1 Model Comparison

In this study, four different classification algorithms were used: XGBoost (Chen and Guestrin 2016), Random Forest (Breiman 2001), LightGBM (Ke et al. 2017), and a stacking ensemble trained with a meta-learner. The performance of the models was evaluated on an independent test set covering the period 2023–2024 (Table 10). The AUC-ROC values obtained for all models ranged from 0.816 to 0.840, indicating a moderate-to-good level of discriminatory ability. The highest F1 score (0.288) was calculated for XGB-SPW, which has an optimized decision threshold of 0.676. XGB-SPW was selected as the primary model for SHAP analysis because it offers the optimum balance between precision (0.235) and recall (0.373). Five-fold temporal cross-validation of this model produced a mean CV F1 score of 0.403 ± 0.040, indicating stable and consistent performance throughout the 2010–2022 training period.

3.6.2 Test Period Performance and Non-Stationarity

A difference was observed between the F1 score obtained through cross-validation (0.403) and the F1 score obtained on the test set (0.288). The test period (2023–2024) coincides with the post-2020 drought period, and therefore has lower annual rainfall than the training set years. The extreme event rate was 4.0% in the test set, while it was 5.2% in the training set, indicating that the distributional characteristics of the target variable changed between the two periods. However, the high AUC-ROC value (0.835) shows that the model successfully distinguishes between extreme and non-extreme events even under changing conditions.

Table 10 Classification performance of all evaluated models on the independent test set (2023–2024, $n = 15{,}645$ station-day records, extreme event rate = 4.0%). The decision threshold for each model was set to the value that maximises F1 on the precision–recall curve.

Model	F1	AUC-ROC	Precision	Recall	Threshold
XGBoost (SPW)	**0.288**	0.835	0.235	0.373	0.676
XGBoost (tuned)	0.281	**0.840**	0.206	0.438	0.644
LightGBM	0.255	0.819	0.175	**0.469**	0.104
Random Forest	0.111	0.816	0.059	0.958	0.500
Stacking (XGB+LGB+RF)	0.264	0.830	0.194	0.416	0.953

Bold = best performance per metric.

3.7 SHAP Feature Importance and Physical Interpretation

3.7.1 Global Feature Importance

SHAP analysis of the XGB-SPW model on the 2023–2024 test set showed that meteorological variables dominate the model predictions (Figures 7–8, Table 11). Relative humidity (`rh_pct`) was the most significant predictive feature, with the highest mean |SHAP| value (0.557). The Beeswarm plot (Figure 7) confirms the physical consistency of the model: SHAP values are close to zero or negative for relative humidity values below approximately 65%, while high relative humidity values strongly increase the probability of extreme precipitation with a positive SHAP contribution. This is consistent with the Clausius–Clapeyron relationship. Mean temperature (`tmean_C`) was identified as the second most significant feature with a mean |SHAP| value of 0.231.

Among the teleconnection features, `AO_lag6`, `EAWR`, and `AMO_lag3` were the highest-ranked large-scale circulation predictors. Although the mean SHAP values of teleconnection indices are low in absolute terms, these variables are physically interpretable within the feature importance hierarchy.

Table 11 Global and cluster-based SHAP feature importance ranking (mean |SHAP|) for the top 15 features of the XGB-SPW model evaluated on the 2023–2024 test set. Cluster columns show the mean |SHAP| value within the test subset of that cluster; '—' indicates the feature is not ranked in the top 5 of that cluster.

Rank	Feature	Global	Coastal	Inland	Mountainous
1	Relative humidity	0.557	0.469	0.559	0.642
2	Mean temperature	0.231	0.281	0.249	0.148
3	ERA5 t2m (corrected)	0.067	0.082	0.070	0.048
4	CHIRPS precipitation	0.044	0.048	0.044	0.042
5	Longitude	0.041	0.038	0.044	—
6	Elevation	0.035	—	—	0.055
7	Season encoding	0.025	—	—	—
8	Month	0.018	—	—	—
9	Latitude	0.010	—	—	—
10	AO_lag6	0.010	—	—	—
11	EAWR	0.007	—	—	—
12	AMO_lag3	0.005	—	—	—
13	EAWR_lag1	0.005	—	—	—
14	EAWR_lag2	0.005	—	—	—
15	NAO_lag1	0.005	—	—	—

3.7.2 Feature Importance at the Cluster Level

When the SHAP values were stratified separately for each cluster, relative humidity was the most important feature in all clusters. However, differences were observed in the feature importance profiles across the three precipitation regime regions (Figure 9).

The mean |SHAP| value for relative humidity was lowest in the Coastal cluster (0.469) and highest in the Mountainous cluster (0.642). The higher SHAP value in the Mountainous cluster indicates that the humid air mass is the decisive control mechanism in the formation of extreme precipitation at high-altitude stations, consistent with the orographic precipitation mechanism.

In the Coastal cluster, synoptic-scale frontal forcing has a more decisive effect; therefore, the model also assigns importance to variables such as temperature and longitude. That is, how cold the air is (temperature) and the direction from which the storm is approaching (longitude) are as important as humidity itself.

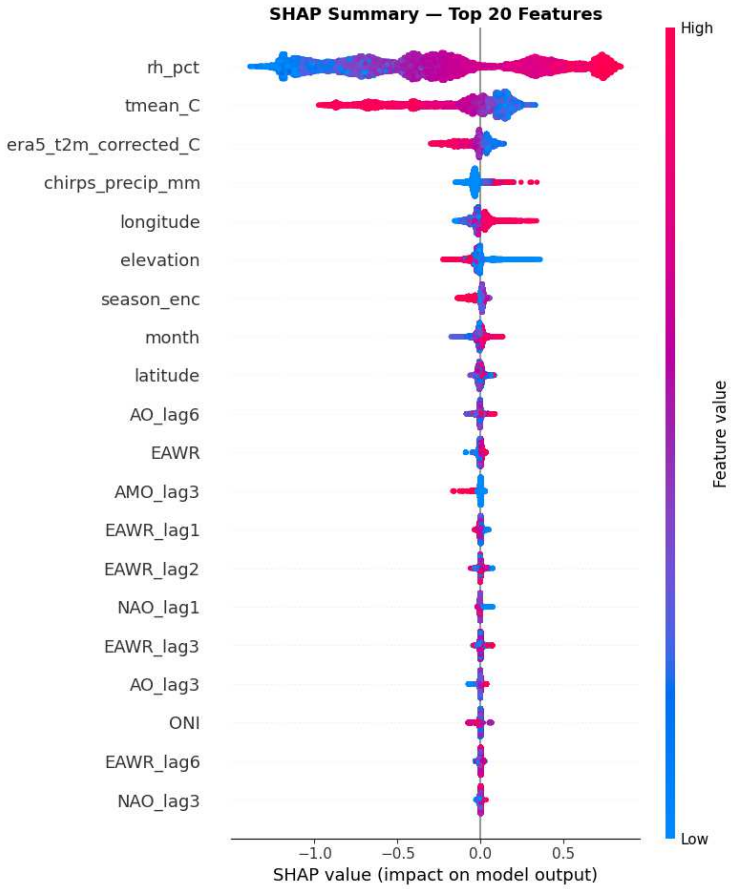


Fig. 7 SHAP summary plot (beeswarm) for the XGBoost classifier (XGB-SPW) calculated on the 2023–2024 test set. Each point represents one observation; the x-axis shows the SHAP value (impact on the log-odds of extreme precipitation); colour indicates feature value (red = high, blue = low). Only the top 20 features ranked by mean |SHAP| are displayed.

The role of mean temperature also varied among the clusters. The mean |SHAP| for the Coastal cluster was 0.281, while for the Mountainous cluster it was 0.148. This highlights the greater role of altitude-driven mechanical forcing at high-elevation stations, where the influence of temperature is reduced relative to the other two clusters. This shift in importance suggests that terrain-induced mechanical uplift becomes a significant predictive factor independent of temperature.

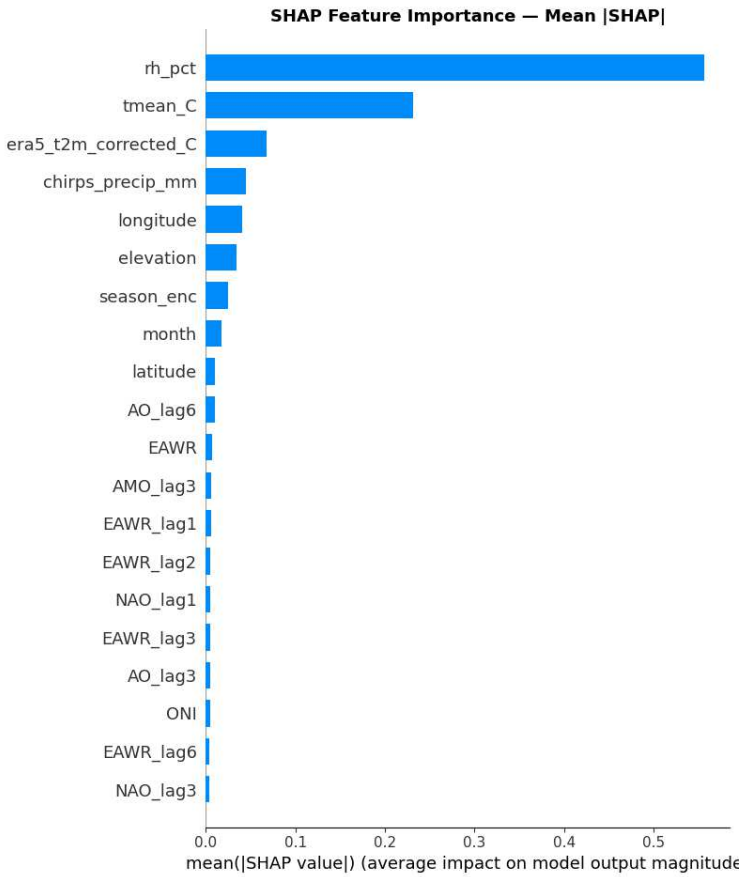


Fig. 8 Global SHAP feature importance (mean |SHAP|) for the top 20 features, averaged over all test set observations ($n = 15{,}645$).

3.7.3 Interaction of Relative Humidity and Temperature

Figure 10 shows the SHAP dependence plots illustrating the physical consistency and synergy between relative humidity and mean temperature, the two most dominant predictors. Points above the SHAP = 0 line increase the predicted probability of extreme precipitation.

When relative humidity exceeds approximately 65–70%, the positive SHAP contribution of humidity reaches its maximum under cold conditions (Figure 10a), corresponding to winter frontal events in the Eastern Mediterranean where cold, humid air masses produce the most intense precipitation.

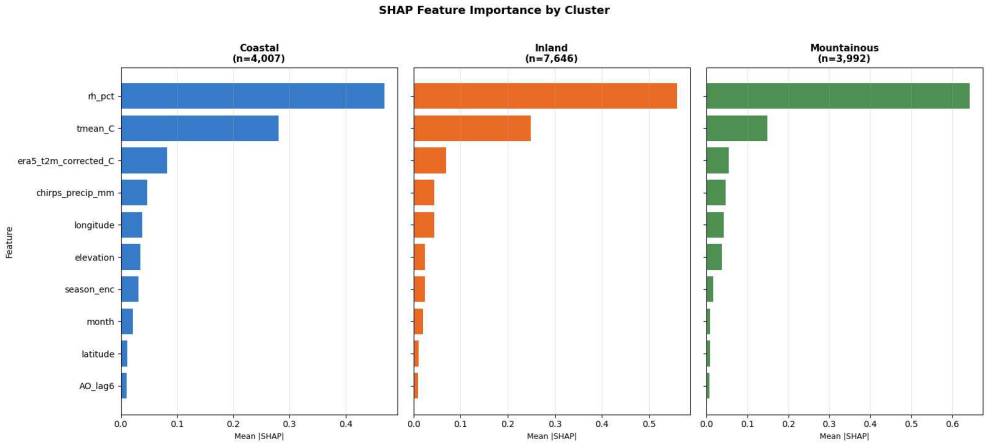


Fig. 9 Cluster-based mean absolute SHAP feature importance for the Coastal ($n = 4{,}007$), Inland ($n = 7{,}646$), and Mountainous ($n = 3{,}992$) subsets of the test set. Bars show the mean |SHAP| value for the top 10 globally ranked features within each cluster.

Mean temperature was identified as the second most important feature, with a mean |SHAP| value of 0.231. Figure 10b shows that below approximately 15–20°C, the SHAP value is positive, while above this threshold it exerts a strong negative effect. This temperature boundary coincides with the seasonal transition period (April–May) across the study area. This humidity–temperature interaction captures, in a physically interpretable and data-driven framework, the cool and humid conditions that represent the most favourable combined thermodynamic state for extreme precipitation, consistent with the theoretical predictions of Clausius–Clapeyron scaling for Mediterranean precipitation extremes.

3.8 Return Periods and Hazard Maps

3.8.1 GEV Fit

GEV distributions were successfully fitted to the 44-year annual maximum CHIRPS series across the 24 study stations (Table 12). The domain mean of the estimated shape parameter ξ is +0.016. The fact that this value is very close to zero indicates

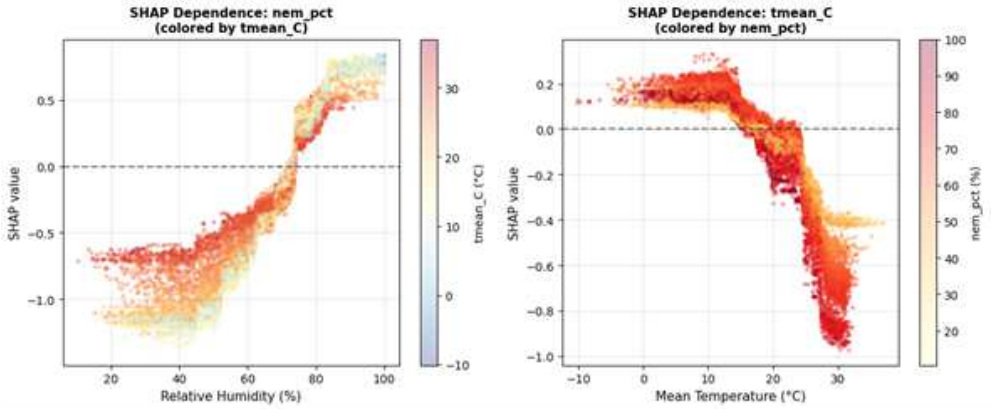


Fig. 10 SHAP dependence plots for the two dominant predictors. (a) Relationship between relative humidity and SHAP value, colour-coded by mean temperature. (b) Relationship between mean temperature and SHAP value, colour-coded by relative humidity. Dashed horizontal line indicates SHAP = 0.

that the extreme precipitation distribution in the region generally exhibits a Gumbel-type behaviour. For individual stations, ξ ranges from -0.179 (Düziçi) to $+0.228$ (Ceyhan TİGEM and Ceyhan). The positive ξ values at stations such as Ceyhan indicate a heavy-tailed distribution, suggesting that these locations may be more prone to exceptionally large extreme events.

The 50-year return level estimates range from 52.0 mm day^{-1} (Tufanbeyli) to 131.3 mm day^{-1} (Düziçi), with a domain mean of 93.5 mm day^{-1}. The 100-year return level estimates range from 54.9 mm day^{-1} (Tufanbeyli) to 153.1 mm day^{-1} (Düziçi), with a domain mean of 102.4 mm day^{-1}.

The spatial distribution of return levels reveals two high-risk zones (Figure 11). The first is the eastern coastal strip (Düziçi, Karataş, Osmaniye, Karaisali), and the second is the central coastal region (Mersin, Tarsus, Adana Region). In the eastern coastal strip, moist air from the Mediterranean Sea rises upon reaching the foothills of the Taurus Mountains, and this orographic interaction produces the most intense extreme rainfall events. A similar mechanism is observed in the central coastal region,

driven by synoptic-scale frontal systems making direct landfall. These findings are consistent with the SHAP analysis results.

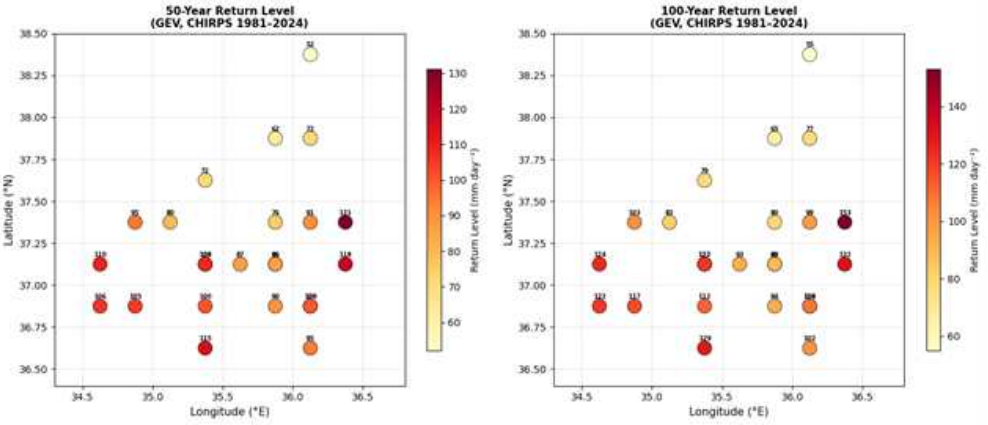


Fig. 11 Spatial distribution of extreme rainfall return levels obtained by fitting the GEV distribution to CHIRPS annual maxima (1981–2024). (a) 50-year return level (mm day^{-1}); (b) 100-year return level (mm day^{-1}). Circle colour and size indicate the magnitude of the return level estimate.

3.8.2 Comparison of GEV and GPD

GPD-based return level estimates are approximately 10–15% lower than GEV estimates at most stations for both 50-year and 100-year return periods (Table 12). Despite this difference, both methods identify the same spatial "hotspots". The lowest GPD estimate was obtained at Tufanbeyli station, consistent with the GEV result. The highest GPD value was calculated for Karataş, which also ranks third in the GEV estimates. The directional consistency in station risk rankings across both methods reinforces confidence that the results reflect the underlying climatic structure rather than methodological artefacts.

At some stations (e.g., Ceyhan, Yumurtalık, and Feke), the GPD estimate exceeded the GEV estimate. This indicates that extreme events occur not only once per year but multiple times annually at these stations. Because GPD captures these clustered

extremes above the threshold, it may provide a higher and more realistic risk estimate than GEV at such locations.

Table 12 GEV and GPD return level estimates for 50-year and 100-year return periods at all 24 study stations, derived from CHIRPS daily precipitation data (1981–2024). GEV = Generalized Extreme Value distribution (Block Maxima method); GPD = Generalized Pareto Distribution (Peaks Over Threshold, $u = P_{95}$); ξ = GEV shape parameter.

Station	Cluster	GEV ξ	GEV (mm day^{-1})		GPD (mm day^{-1})	
			$T = 50$	$T = 100$	$T = 50$	$T = 100$
Mersin	Coastal	−0.148	105.9	121.5	78.2	85.5
Tarsus	Coastal	−0.049	105.1	116.9	93.4	104.0
Yüreğir/Çukurova	Coastal	−0.079	100.5	111.7	94.7	103.3
Adana Bölge	Coastal	−0.118	108.3	121.9	93.2	101.7
Ceyhan	Coastal	+0.228	85.8	89.4	91.9	102.1
Yumurtalık	Coastal	+0.208	90.1	94.2	97.1	106.6
Karataş	Coastal	−0.110	114.6	129.2	113.5	126.2
İskenderun	Inland	+0.044	94.6	102.0	81.7	88.9
Erzin	Inland	+0.007	99.8	108.5	74.8	79.7
Ceyhan TİGEM	Inland	+0.228	85.8	89.4	91.9	102.1
İmamoğlu	Inland	+0.064	86.7	92.7	89.3	98.1
Kadirli	Inland	−0.001	91.3	99.0	76.1	81.8
Osmaniye	Inland	−0.067	118.3	131.8	80.6	86.2
Kozan	Inland	+0.098	75.8	80.1	69.6	74.8
Çukurova	Inland	−0.118	108.3	121.9	93.2	101.7
Karaisali	Inland	+0.210	79.7	83.2	68.2	72.0
Düziçi	Inland	−0.179	131.3	153.1	73.8	78.3
Dörtyol	Inland	+0.007	99.8	108.5	74.8	79.7
Feke	Mountainous	+0.150	62.2	65.4	72.8	81.5
Aladağ	Mountainous	−0.056	72.0	78.9	69.6	76.0
Pozantı	Mountainous	+0.021	94.9	102.8	79.7	85.1
Saimbeyli	Mountainous	+0.035	71.5	76.8	74.1	82.3
Çamlıyayla	Mountainous	−0.118	109.5	124.5	88.6	97.3
Tufanbeyli	Mountainous	+0.127	52.0	54.9	51.3	55.8
Domain mean		+0.016	**93.5**	**102.4**	**82.2**	**89.6**

3.9 Future Extreme Frequency under Climate Scenarios

3.9.1 Changes in Extreme Event Frequency

Bias-corrected MPI-ESM1-2-LR projections indicate a consistent and spatially homogeneous decrease in extreme rainfall frequency across the study area for the period 2041–2060 under both SSP2-4.5 and SSP5-8.5 scenarios (Figure 12, Table 13).

The historical domain mean of 21.6 extreme days yr^{-1} (1981–2014) decreases by 19.5% to 17.3 days yr^{-1} under SSP2-4.5. Under SSP5-8.5, the number of extreme rainfall days decreases to 19.1 days yr^{-1}, representing a 12.1% reduction. A spatially homogeneous decrease in extreme rainfall frequency is projected across all stations under both scenarios.

The smaller decrease under SSP5-8.5 compared to SSP2-4.5 is counterintuitive but physically explicable: under higher forcing, the increase in atmospheric moisture content partially offsets the reduction in synoptic-scale cyclone activity, resulting in a more limited frequency decline under the high-emission scenario.

3.9.2 Changes in Wet-Day Intensity

Although extreme event frequency showed a decreasing trend under both scenarios, the projected change in wet-day intensity differed in direction between the two. Under SSP2-4.5, wet-day intensity decreased by -2.2%, while under SSP5-8.5 an increase of $+1.2\%$ was projected. While SSP2-4.5 suggests a decrease in both frequency and intensity, SSP5-8.5 suggests a decrease in frequency but an increase in intensity. Under high greenhouse gas concentrations, enhanced atmospheric moisture content increases precipitation intensity, partially compensating for the declining number of rainy days.

4 Discussion

4.1 2020 Regional Precipitation Regime Change and its Effects on Model Transferability

The homogeneity test applied to precipitation data obtained from TSMS for 24 stations revealed a decrease in precipitation across the area in 2020, statistically significant for some stations. This finding indicates that the regionally observed drying is a genuine drying event, not caused by instrument malfunction as the signal is consistent across all stations regardless of cluster membership. Similar findings havebeen

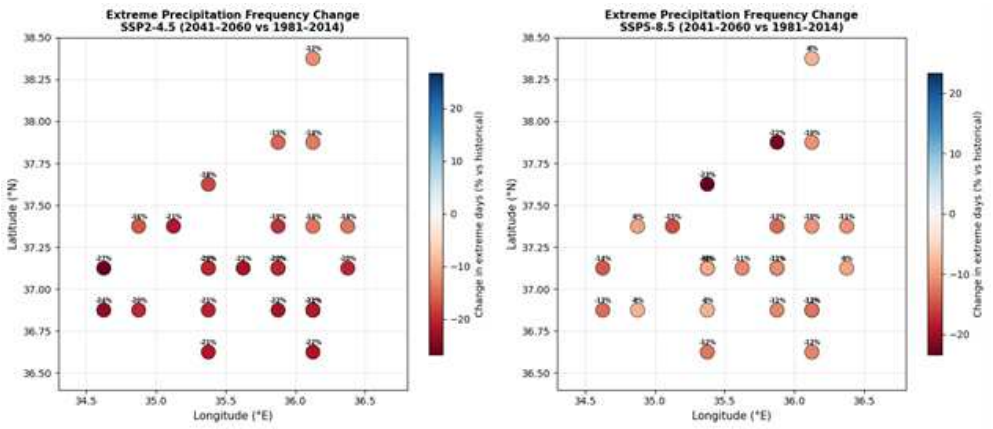


Fig. 12 Projected change in annual extreme precipitation day frequency under (a) SSP2-4.5 and (b) SSP5-8.5 scenarios relative to the historical baseline (1981–2014); MPI-ESM1-2-LR, r1i1p1f1, 2041–2060. Red shading indicates projected decreases. Values above each circle show the percentage change. All stations show negative changes under both scenarios, indicating a consistent projection of decreasing extreme event frequency across the study area.

reported by Bağçaci et al. (2021) and Zittis et al. (2022) in their studies covering the Mediterranean region of Turkey.

The break that occurred in 2020 and the subsequent decrease in precipitation are the most likely primary reasons for the difference in performance between the temporal cross-validation F1 score (0.403 ± 0.040) for the 2010–2022 training period and the F1 score (0.288) for the 2023–2024 test period. The decrease in the extreme event rate from 5.2% in the training set to 4.0% in the test set confirms that the target variable experienced a real distributional shift between the two periods, meaning that the F1 score was mathematically affected by this dilution of events. The fact that the AUC-ROC remained stable at 0.835 indicates that the model retained significant discriminatory ability despite the regime transition. The deterioration in the F1 value is not due to a breakdown in the feature-label relationship, but rather to a change in the event base ratio.

Table 13 Projected changes in extreme precipitation frequency and wet-day intensity under SSP2-4.5 and SSP5-8.5 scenarios (MPI-ESM1-2-LR, r1i1p1f1, 2041–2060 vs. 1981–2014). Extreme days defined as daily precipitation $\geq$ station-specific P_{95} threshold. Wet-day intensity = mean daily precipitation on days with pr $>$ 1 mm day^{-1}.

Station	Cluster	Hist. (days yr^{-1})	SSP2-4.5 (days yr^{-1})	Δ_{245} (%)	SSP5-8.5 (days yr^{-1})	Δ_{585} (%)
Mersin	Coastal	25.2	19.2	−24	21.9	−13
Tarsus	Coastal	28.9	23.1	−20	26.6	−8
Yüreğir/Çukurova	Coastal	26.0	20.6	−21	23.9	−8
Adana Bölge	Coastal	25.0	19.9	−20	22.2	−11
Ceyhan	Coastal	25.1	20.0	−20	22.4	−11
Yumurtalık	Coastal	23.0	17.9	−22	20.4	−11
Karataş	Coastal	19.7	15.5	−21	17.2	−12
İskenderun	Inland	18.4	14.4	−22	16.3	−11
Erzin	Inland	21.2	16.6	−21	18.5	−13
Ceyhan TİGEM	Inland	25.1	20.0	−20	22.4	−11
İmamoğlu	Inland	23.0	17.9	−22	20.4	−11
Kadirli	Inland	19.2	16.4	−14	17.2	−10
Osmaniye	Inland	26.6	21.2	−20	24.2	−9
Kozan	Inland	16.7	13.5	−19	14.5	−13
Çukurova	Inland	26.8	21.4	−20	24.4	−9
Karaisali	Inland	15.6	12.3	−21	13.2	−15
Düziçi	Inland	19.5	16.8	−14	17.4	−11
Dörtyol	Inland	20.8	16.2	−22	18.0	−13
Feke	Mountainous	8.6	7.3	−15	6.7	−22
Aladağ	Mountainous	9.9	8.1	−18	7.5	−23
Pozantı	Mountainous	32.4	27.1	−16	29.4	−9
Saimbeyli	Mountainous	20.4	17.6	−14	18.2	−10
Çamlıyayla	Mountainous	17.3	12.7	−27	14.8	−14
Tufanbeyli	Mountainous	23.4	20.4	−13	21.5	−8
Domain mean		**21.6**	**17.3**	**−19.5**	**19.1**	**−12.1**

4.2 Dominance of Near-Surface Relative Humidity and its Physical Interpretation

The SHAP analysis revealed that near-surface relative humidity was the dominant predictor across all clusters. The mean |SHAP| value for relative humidity was 0.557, more than double the value calculated for the second-ranked predictor, mean temperature (0.231). The dependence plot showed a significant nonlinear threshold around 65–70% where SHAP values were close to zero or negative (Figure 10). This value coincides with the theoretical threshold for deep convective triggering in the boundary layer

(Martínková and Kyselý 2020). The findings show that the model empirically recovers the Clausius–Clapeyron relationship from the training data without any explicit physical equation input (Zittis et al. 2021).

The mean |SHAP| value for relative humidity increased from 0.469 in the Coastal cluster to 0.559 in the Inland cluster and 0.642 in the Mountainous cluster. This progressive increase indicates that the role of humidity in the precipitation-triggering mechanism changes with altitude. While synoptic-scale forcing and sea surface temperature anomalies exert a stronger influence in the Coastal cluster (Krichak et al. 2014), orographic uplift provides an almost constant dynamic lifting mechanism in the Mountainous cluster, making the available precipitable water — represented by near-surface relative humidity — the dominant control over extreme precipitation probability.

4.3 Teleconnection Signals and Prediction Contributions

Spearman correlation analysis identified seven statistically significant teleconnection predictors. AMO ($r = -0.335$) and EAWR ($r = +0.299$) emerged as the strongest large-scale drivers of regional precipitation variability.

The stronger correlation of AMO compared to NAO is noteworthy relative to Western Mediterranean precipitation studies (Düzenli et al. 2018). This can be explained by the modulation of the North Atlantic sea surface temperature gradient on the position and intensity of the Azores High (Krichak et al. 2014), which directly suppresses the frontal systems carrying winter precipitation to southern Türkiye. Notably, AMO at a 6-month lag showed a positive correlation ($r = +0.285$), suggesting a delayed ocean–atmosphere feedback mechanism.

The positive correlation of precipitation with EAWR reflects the blocking pattern that forms over Western Russia, which redirects extratropical cyclones southward and increases moisture transport towards the eastern Mediterranean (Krichak et al. 2014). Although the absolute SHAP values of teleconnection features are lower than those of

humidity or temperature, they consistently appear in the global feature importance ranking, demonstrating that large-scale atmospheric circulation provides statistically detectable predictive information beyond contemporaneous surface meteorological observations.

4.4 Recurrence Period Estimates and Infrastructure Impacts

The 50-year and 100-year return levels based on GEV show significant spatial variability across the study area. The highest values were obtained at eastern corridor stations such as Düziçi (153 mm day^{-1}) and Osmaniye (132 mm day^{-1}), where humid air from the Mediterranean Sea encounters the Taurus Mountain foothills, producing the most intense orographic precipitation. Both methods (GEV and GPD) identified the highest risk areas at the same stations (Düziçi, Osmaniye, Karataş). The consistency of both methods shows that the hazard pattern found is not dependent on the method chosen but rather on the local climate. It is also noted that Tufanbeyli had the lowest T100 (54.9 mm) due to the rain shadow effect created by the inner Taurus mountain range, despite having a very high elevation (1,415 m). This finding reveals that high altitude does not always mean a high risk of extreme rainfall.

The domain-mean T100 estimate of 102.4 mm day^{-1} is broadly consistent with values reported for nearby Mediterranean stations in comparable GEV analyses (Bağçaci et al. 2021), providing independent cross-validation of the CHIRPS-based frequency analysis despite the known orographic underestimation bias of satellite precipitation products at high-elevation stations.

Many urban drainage systems in Türkiye are designed for thresholds of 50–100 mm day^{-1}. The estimated 100-year return levels at several eastern corridor stations substantially exceed these design thresholds, indicating a structural vulnerability of existing infrastructure to century-scale precipitation extremes in the region.

4.5 Future Projections: Frequency–Intensity Separation

Both CMIP6 scenarios project a consistent decrease in extreme precipitation frequency across all 24 stations during the 2041–2060 period, with domain means of −19.5% and −12.1% under SSP2-4.5 and SSP5-8.5, respectively. This projected decrease is physically consistent with the broader Mediterranean drying signal: the northward shift of mid-latitude storm tracks reduces the frequency of precipitation-bearing synoptic systems over Türkiye, while the weakening of Mediterranean cyclogenesis under warming further suppresses extreme event occurrence (Chericoni et al. 2025).

The smaller frequency decrease under SSP5-8.5 compared to SSP2-4.5 is apparently counterintuitive but can be explained through the competing dynamic and thermodynamic effects on precipitation under high forcing. Enhanced atmospheric moisture content under stronger warming partially offsets the dynamically driven reduction in cyclone activity, resulting in a less pronounced frequency decline.

Although extreme rainfall frequency decreases under both scenarios, the projected changes in wet-day intensity differ in direction: a decrease of −2.2% under SSP2-4.5 and an increase of +1.2% under SSP5-8.5. These projections present a "fewer but potentially more intense" future for extreme precipitation in the eastern Mediterranean (Zittis et al. 2021). While the annual number of extreme rainfall days declines, the intensity of individual events may increase, particularly under the high-emission scenario. It should be noted that these projections are based on a single CMIP6 model (MPI-ESM1-2-LR) and that multi-model ensemble analysis will reduce projection uncertainty.

4.6 Limitations and Future Research Directions

Several limitations of this study should be explicitly acknowledged. First, the use of a single CMIP6 model (MPI-ESM1-2-LR) precludes quantification of inter-model projection uncertainty; a multi-model ensemble analysis encompassing 5–10 CMIP6 models

would substantially improve the robustness of frequency and intensity projections. Second, the Delta Method bias correction preserves the model's projected relative change but does not correct for potential errors in simulated precipitation variability or the shape of the extreme value distribution; quantile mapping approaches would provide greater distributional fidelity for extreme event analysis. Third, the relatively short study period (2010–2024) is insufficient for robust extreme value analysis at the MGM station network; future studies should incorporate extended historical records where available to constrain GEV parameter uncertainty. The ML classification model does not include real-time atmospheric predictors such as higher-level divergence or precipitation. Adding these variables could improve operational forecasting capabilities. Finally, integrating social vulnerability and exposure data into a composite flood risk framework — building upon the extreme event frequency and intensity metrics developed here — would substantially enhance the practical applicability of these results for climate adaptation planning in the region.

5 Conclusion

This study applies a hybrid machine learning approach combining multi-source data fusion, K-Means spatial clustering, XGBoost classification with SHAP explainability, Generalized Extreme Value (GEV) and Generalized Pareto Distribution (GPD) return period analysis, and CMIP6 climate projections for extreme precipitation analysis and prediction in the Eastern Mediterranean coastal region of Türkiye (Mersin, Adana, Osmaniye, and Hatay).

Homogeneity analysis revealed a statistically significant change in the precipitation regime in the region after 2020, with a decrease of approximately -20.5% observed across the region compared to the preceding decade. The study area was divided into three homogeneous sub-regimes (Coastal, Inland, and Mountainous) using K-Means spatial clustering.

The XGBoost binary classification model, which addressed the class imbalance problem using inverse frequency weighting (`scale_pos_weight`), achieved an AUC-ROC of 0.835 on the independent test set, demonstrating strong discriminatory performance in predicting extreme rainfall events. In the SHAP analysis, near-surface relative humidity (mean $|\text{SHAP}| = 0.557$) showed overwhelming dominance in triggering extreme rainfall events, with its importance increasing progressively from coastal to mountainous regions. The influence of mean temperature, the second most important predictor, decreased from low-lying coastal regions to mountainous regions. The large-scale teleconnection indices AMO ($r = -0.335$) and EAWR ($r = +0.299$) emerged as the dominant drivers of interannual rainfall variability.

GEV analysis applied to the annual maximum series of the CHIRPS dataset yielded domain-mean 50-year and 100-year return levels of 93.5 and 102.4 mm day^{-1}, respectively. The highest hydrological risk was concentrated in the eastern inland corridor, where orographic and synoptic interactions converge. CMIP6 projections under SSP2-4.5 and SSP5-8.5 scenarios indicate decreases in extreme event frequency of -19.5% and -12.1%, respectively, by 2041–2060. A decrease in wet-day intensity was projected under SSP2-4.5, while an increase was projected under SSP5-8.5, indicating that the high-emission scenario will bring fewer but potentially more severe extreme rainfall events.

Based on these findings, revision of urban drainage design standards in the eastern Osmaniye–Hatay corridor is considered necessary, as recurrence level estimates substantially exceed the capacity of current infrastructure systems. Furthermore, the development of adaptive ML retraining protocols that account for non-stationary conditions arising from hydroclimatological regime changes, and the integration of teleconnection lag signals with humidity thresholds into early warning systems, are recommended. Future studies may incorporate additional atmospheric variables to improve ML classification capability and expand the analysis to a multi-model CMIP6 ensemble.

References

Aksu H, Akgül MA, Sönmez İ (2023) Evaluation of spatial and temporal variability of precipitation in the Mediterranean region of Turkey. Theoretical and Applied Climatology 152:1187–1202. https://doi.org/10.1007/s00704-023-04503-3

Alexandersson H (1986) A homogeneity test applied to precipitation data. Journal of Climatology 6:661–675. https://doi.org/10.1002/joc.3370060607

Bağçaci SÇ, Yucel I, Duzenli E, et al (2021) Intercomparison of the expected change in the temperature and the precipitation retrieved from CMIP6 and CMIP5 climate projections: A Mediterranean hot spot case, Turkey. Atmospheric Research 256:105576. https://doi.org/10.1016/j.atmosres.2021.105576

Beguería S, Vicente-Serrano SM (2010) SPEI: a multiscalar drought index sensitive to global warming. Journal of Climate 23:1696–1718. https://doi.org/10.1175/2009JCLI2909.1

Breiman L (2001) Random forests. Machine Learning 45:5–32. https://doi.org/10.1023/A:1010933404324

Buishand TA (1982) Some methods for testing the homogeneity of rainfall records. Journal of Hydrology 58:11–27. https://doi.org/10.1016/0022-1694(82)90066-X

Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 785–794, https://doi.org/10.1145/2939672.2939785

Chericoni M, Fosser G, Flaounas E, et al (2025) Unravelling drivers of the future Mediterranean precipitation paradox during cyclones. npj Climate and Atmospheric Science 8. https://doi.org/10.1038/s41612-025-01121-w

Coles S (2001) An Introduction to Statistical Modeling of Extreme Values. Springer, London, https://doi.org/10.1007/978-1-4471-3675-0

Copernicus Climate Data Store (2024) E-OBS gridded land-only observational dataset. https://cds.climate.copernicus.eu, accessed: April 2025

Cornes RC, van der Schrier G, van den Besselaar EJM, et al (2018) An ensemble version of the E-OBS temperature and precipitation data sets. Journal of Geophysical Research: Atmospheres 123:9391–9409. https://doi.org/10.1029/2017JD028200

Cunnane C (1973) A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. Journal of Hydrology 18:257–271. https://doi.org/10.1016/0022-1694(73)90051-6

Deepa R, Kumar V, Sundaram S (2024) A systematic review of regional and global climate extremes in CMIP6 models under shared socio-economic pathways. Theoretical and Applied Climatology 155:2523–2543. https://doi.org/10.1007/s00704-024-04872-3

Düzenli E, Tabari H, Willems P, et al (2018) Decadal variability analysis of extreme precipitation in Turkey and its relationship with teleconnection patterns. Hydrological Processes 32:3513–3528. https://doi.org/10.1002/hyp.13275

Eyring V, Bony S, Meehl GA, et al (2016) Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geoscientific Model Development 9:1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Farr TG, et al (2007) The shuttle radar topography mission. Reviews of Geophysics 45:RG2004. https://doi.org/10.1029/2005RG000183

826 de Freitas ITA, Abrahão R (2025) Study of global solar irradiance in north-
827 east Brazil with considerations on superirradiance and extreme superirradiance
828 events. Theoretical and Applied Climatology 156:1–12. https://doi.org/10.1007/
829 s00704-024-05193-8

830 Funk C, Peterson P, Landsfeld M, et al (2015) The climate hazards infrared precipita-
831 tion with stations — a new environmental record for monitoring extremes. Scientific
832 Data 2:150066. https://doi.org/10.1038/sdata.2015.66

833 Google Cloud (2024) CMIP6 climate model output. https://console.cloud.google.com/
834 storage/browser/cmip6, accessed: April 2025

835 Hamed KH (2016) The distribution of Spearman's rho trend statistic for persis-
836 tent hydrologic data. Hydrological Sciences Journal 61:214–223. https://doi.org/10.
837 1080/02626667.2015.1006985

838 Hess P, Boers N (2022) Deep learning for improving numerical weather prediction of
839 heavy rainfall. Journal of Advances in Modeling Earth Systems 14:e2021MS002765.
840 https://doi.org/10.1029/2021MS002765

841 Hinkle DE, Wiersma W, Jurs SG (2003) Applied Statistics for the Behavioral Sciences.
842 Houghton Mifflin, Boston

843 Hisam E, Sertel E, Seker DZ (2025) Precipitation downscaling with the integration of
844 multiple precipitation products, land surface data and gauge stations using explain-
845 able machine learning algorithms: A case study in the Mediterranean region of
846 Türkiye. Science of The Total Environment 1002:180540. https://doi.org/10.1016/
847 j.scitotenv.2025.180540

848 Intergovernmental Panel on Climate Change (2023) Climate change 2023: Synthe-
849 sis report. Tech. rep., IPCC, Geneva, Switzerland, https://doi.org/10.59327/IPCC/

AR6-9789291691647

Ke G, et al (2017) LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems 30:3146–3154

Klein Tank AMG, et al (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. International Journal of Climatology 22:1441–1453. https://doi.org/10.1002/joc.773

Krichak SO, Breitgand JS, Gualdi S, et al (2014) Teleconnection–extreme precipitation relationships over the Mediterranean region. Theoretical and Applied Climatology 117:679–692. https://doi.org/10.1007/s00704-013-1036-4

Li X, Zheng W, Liu X, et al (2020) Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. Journal of Hydrology 586:124901. https://doi.org/10.1016/j.jhydrol.2020.124901

Lin Y, Fan K, Hou Z, et al (2023) Machine learning of key variables impacting extreme precipitation in various regions of the contiguous United States. Journal of Advances in Modeling Earth Systems 15:e2022MS003304. https://doi.org/10.1029/2022MS003304

Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems 30:4765–4774

Madsen H, Rosbjerg D (1997) The partial duration series method in regional index-flood modeling. Water Resources Research 33:737–746. https://doi.org/10.1029/96WR03847

Maraun D (2016) Bias correcting climate change simulations — a critical review. Current Climate Change Reports 2:211–220. https://doi.org/10.1007/

s40641-016-0050-x

Martínková M, Kyselý J (2020) Overview of Observed Clausius-Clapeyron Scaling of Extreme Precipitation in Midlatitudes. Atmosphere 11(8):786. https://doi.org/10.3390/atmos11080786

Mastrantonas N, Furnari L, et al (2022a) Extreme precipitation events in the Mediterranean: Spatiotemporal characteristics and connection to large-scale atmospheric flow patterns. International Journal of Climatology 42:4479–4498. https://doi.org/10.1002/joc.7482

Mastrantonas N, Magnusson L, et al (2022b) Forecasting extreme precipitation events in the Mediterranean: Insights from the evaluation of selected global and regional climate models. Weather and Forecasting 37:211–229. https://doi.org/10.1175/WAF-D-21-0078.1

Mastrantonas N, et al (2021) Potential of statistical and machine learning approaches for the seasonal prediction of Mediterranean precipitation. Meteorology and Atmospheric Physics 133:1149–1164. https://doi.org/10.1007/s00703-021-00787-8

Mauritsen T, et al (2019) Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO_2. Journal of Advances in Modeling Earth Systems 11:998–1038. https://doi.org/10.1029/2018MS001400

Mukaka MM (2012) Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Medical Journal 24:69–71

Muñoz-Sabater J, et al (2021) ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. Earth System Science Data 13:4349–4383. https://doi.org/10.5194/essd-13-4349-2021

NOAA Climate Prediction Center (2024) Climate indices: Monthly atmospheric and ocean data. https://www.cpc.ncep.noaa.gov, accessed: April 2025

NOAA Global Monitoring Laboratory (2024) Climate forcing data. https://gml.noaa.gov, accessed: April 2025

O'Neill BC, et al (2016) The scenario model intercomparison project (ScenarioMIP) for CMIP6. Geoscientific Model Development 9:3461–3482. https://doi.org/10.5194/gmd-9-3461-2016

Pettitt AN (1979) A non-parametric approach to the change-point problem. Applied Statistics 28:126–135. https://doi.org/10.2307/2346729

Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20:53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Salcedo-Sanz S, et al (2024) Machine learning for extreme precipitation events: A review. Wiley Interdisciplinary Reviews: Climate Change 15:e878. https://doi.org/10.1002/wcc.878

Senjaliya J, Vaghasia V, Shah SM (2026) A comprehensive review of machine learning and deep learning approaches for rainfall forecasting: current progress, challenges, and future directions. Theoretical and Applied Climatology 157:177. https://doi.org/10.1007/s00704-026-06051-y

Spuler FR, et al (2024) Climate change and machine learning-based prediction of extreme precipitation in Europe. Geophysical Research Letters 51:e2023GL107893. https://doi.org/10.1029/2023GL107893

Tariq M, Aruffo E, Chiacchiaretta P, et al (2025) Hydro-climatic trends in central Italy: A case study from the Aterno-Pescara river watershed. Sustainability 17:1–17. https://doi.org/10.3390/su17020465

Turkish State Meteorological Service (2024) Daily meteorological observation data. https://www.mgm.gov.tr, accessed: April 2025

Yilmaz AG, Imteaz MA (2014) Investigating homogeneity of spatial snowmelt data in mountainous catchments. Hydrological Processes 28:4264–4274. https://doi.org/10.1002/hyp.9946

Zar JH (2005) Spearman rank correlation. In: Encyclopedia of Biostatistics. Wiley, https://doi.org/10.1002/0470011815.b2a15150

Zittis G, Bruggeman A, Lelieveld J (2021) Revisiting future extreme precipitation trends in the Mediterranean. Weather and Climate Extremes 34:100380. https://doi.org/10.1016/j.wace.2021.100380

Zittis G, Almazroui M, Alpert P, et al (2022) Climate change and weather extremes in the Eastern Mediterranean and Middle East. Reviews of Geophysics 60(3):e2021RG000762. https://doi.org/10.1029/2021RG000762

Funding

This research received no specific funding.

Author contributions statement

F.D.C. conceived the study, performed data curation, formal analysis, methodology, visualization, and writing – original draft and review & editing.

A.C. performed data curation, formal analysis, methodology, software development, visualization, and writing – original draft and review & editing. All authors reviewed the manuscript.

Additional information

Competing interests: The authors declare no competing interests.

Data availability

The CHIRPS dataset is publicly available at https://www.chc.ucsb.edu/data/chirps. ERA5-Land data were obtained from the Copernicus Climate Data Store (https://cds.climate.copernicus.eu). Teleconnection indices were obtained from NOAA CPC (https://www.cpc.ncep.noaa.gov) and NOAA GML (https://gml.noaa.gov). CMIP6 model outputs are available via the Google Cloud CMIP6 repository (https://console.cloud.google.com/storage/browser/cmip6). TSMS ground station data are available upon request from the Turkish State Meteorological Service (https://www.mgm.gov.tr).