Atmospheric Research 334 (2026) 108763

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmosres

A machine learning-based backward extension of IMERG daily precipitation over the Greater Alpine Region

Iman Goudarzi [a,*], Davide Fazzini [b], Claudia Pasquero [a], Agostino Niyonkuru Meroni [a], Matteo Borgnino [a]

[a] Department of Earth and Environmental Sciences, University of Milano-Bicocca, Piazza della Scienza 4, Milan, 20126, Italy
[b] Department of Physics, University of Milano-Bicocca, Piazza della Scienza 3, Milan, 20126, Italy

ARTICLE INFO

Dataset link: doi:10.5281/zenodo.16631516, github.com/imanam1371/maxm-20240516

Keywords:
Precipitation
Complex orography
ERA5
IMERG
Machine learning
XGB
Climate change

ABSTRACT

Accurate knowledge of precipitation at high spatio-temporal resolution is essential for climate studies and hydrological applications, particularly in mountainous regions where traditional models often underperform due to coarse resolution and sparse observational networks. In this study, we present a machine learning-based approach to enhance ERA5 reanalysis precipitation estimates using the satellite-derived IMERG (Integrated Multi-satellite Retrievals for GPM) product as a reference. We focus on the Greater Alpine Region (GAR), using extreme gradient boosting combined with Shapley additive explanations to identify the most influential ERA5 variables. This method enables the creation of a new daily rainfall dataset, ML-IMEX-GAR (Machine Learning IMERG backward-EXtended precipitation dataset over GAR), at IMERG's spatial resolution for the historical period 1960–2000.

Compared to ERA5, ML-IMEX-GAR reduces the spatiotemporal RMSD against IMERG by approximately 14%, and achieves strong agreement with in-situ observational monthly data, with an R^2 of 0.87. These findings demonstrate the potential of machine learning to correct reanalysis biases, improve historical precipitation reconstructions, and support climate change research in data-scarce, complex terrains.

1. Introduction

Precipitation is a fundamental element of the hydrological cycle, and accurately capturing its spatial and temporal variability is vital, as it underpins the understanding of the links with numerous hydrometeorological processes and their socioeconomic consequences (Breugem et al., 2020; Du et al., 2020; Markonis et al., 2019). Moreover, reliable precipitation estimates are key for analyzing and forecasting weather and climate extremes, such as droughts, floods, and landslides, and for ensuring effective water resource management (Van Dijk et al., 2013; Sahoo et al., 2015).

One of the most accurate and reliable measures for precipitation data is obtained with ground-based devices, such as manual and automatic rain gauges (Xu et al., 2020). However, rain gauges face several challenges including their limited number (Lanza and Stagi, 2009; Lanza and Vuerich, 2009; Schamm et al., 2014), which leads to restricted spatial coverage; their sparse global availability (Becker et al., 2013; Harris et al., 2014; Schneider et al., 2016; Zhang et al., 2021b); and the difficulty of installing them in certain regions, such as high mountains or the open ocean. Despite their limitations and inherent uncertainties, remote sensing products serve as a valuable alternative to traditional rain gauge data, offering continuous observations and broader spatial coverage. Several widely used datasets, such as the Climate Hazards group InfraRed Precipitation with Stations (CHIRPS, Funk et al., 2015), the Multi-Source Weighted-Ensemble Precipitation (MSWEP, Beck et al., 2019), the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks Climate Data Record (PERSIANN-CDR, Ashouri et al., 2015), and the Integrated Multi-satellitE Retrievals for the Global Precipitation Measurement Mission (IMERG, Huffman et al., 2020) each come with distinct spatial and temporal resolutions, geographic coverage, strengths, and limitations. These differences make them suitable for various hydrological and climatological applications depending on the study's objectives and region of interest.

Additionally, atmospheric reanalysis and numerical weather model data are valuable for analyzing the dynamical evolution of precipitation across multiple spatial and temporal scales. However, their reliability is often hindered by systematic biases, largely due to the sensitivity of

* Corresponding author.
E-mail addresses: iman.goudarzi@unimib.it (I. Goudarzi), davide.fazzini@unimib.it (D. Fazzini), claudia.pasquero@unimib.it (C. Pasquero), agostino.meroni@unimib.it (A.N. Meroni), matteo.borgnino@unimib.it (M. Borgnino).

https://doi.org/10.1016/j.atmosres.2026.108763
Received 17 September 2025; Received in revised form 10 December 2025; Accepted 7 January 2026
Available online 12 January 2026

model outputs to critical numerical parameterizations such as microphysics and turbulence schemes (Toride et al., 2019; Frei et al., 2003). For instance, reanalysis data are not suitable for climatic study because the amount and type of observational products that are assimilated change over time, making hard to distinguish trends due to real climate change from those due to the different observational inputs (Tang et al., 2020). These challenges in accurately capturing precipitation are particularly critical in regions where hydrological processes are highly sensitive to climatic variability, such as mountainous areas.

Mountain regions have recently attracted increasing scientific interest due to their pronounced sensitivity and their status as critical environmental hotspots. Mountains serve as the Earth's primary freshwater reservoirs, especially in the form of snow and ice (Viviroli et al., 2011), and they play a pivotal role in regulating the global hydrological cycle. Climate-induced reductions in the cryosphere are therefore of major concern, with potential consequences for billions of people who depend on downstream water resources (Viviroli et al., 2020). High-altitude environments are characterized by complex and heterogeneous climate dynamics, influenced by diverse topographies and landscapes across multiple climate zones. Advances in observational techniques and computational modeling have highlighted the significant role of topography in shaping regional climate responses. For instance, in the GAR, topographic features have been shown to modulate precipitation trends through mechanisms that remain only partially understood (Napoli et al., 2019). These include orographic effects such as barrier-induced shadowing and enhanced convective activity (Giorgi et al., 2016), underscoring the intricate interplay between terrain and atmospheric processes.

Given this complexity, improving the representation of precipitation in mountainous regions requires high-resolution climate data. Downscaling coarse-resolution data to finer scales is therefore essential to capture local variability. For instance, Isotta et al. (2014) emphasized the necessity of high-resolution gridded datasets from dense rain-gauge networks to adequately represent precipitation variability in Alpine terrains. Their dataset, derived from pan-Alpine rain gauges, serves as an essential resource for validating and calibrating satellite-based precipitation products. Similarly, Flaounas et al. (2013) evaluated both dynamical and statistical downscaling techniques. Their results showed that, despite limitations in capturing the full three-dimensional structure of atmospheric processes, statistical methods were effective in reproducing local-scale precipitation variability.

More recently, Machine Learning (ML) techniques have shown strong potential to identify embedded patterns and rules to represent precipitation timeseries or precipitation maps. A range of machine learning models has been used for such purposes, including Random Forests (Bhuiyan et al., 2020), Artificial Neural Networks (ANNs) (Bhuiyan et al., 2020; Wehbe et al., 2020), and Double Machine Learning (Zhang et al., 2021a). Dhinakaran et al. (2024), for instance, applied several ML algorithms obtaining seasonal climate forecasting over the Alpine region with low bias. Similarly, Lyu and Yong (2024) proposed a Double Machine Learning framework that integrates meteorological data, satellite retrievals, and reanalysis products to deliver high-precision precipitation estimates over the Tibetan Plateau. Their approach significantly outperformed conventional statistical techniques, particularly under medium-to-high rainfall scenarios. In another study, Bhuiyan et al. (2020) employed ML-based error modeling to improve the accuracy of NASA's IMERG precipitation data over the Brahmaputra basin. By incorporating meteorological and land-surface variables, both neural networks and random forests were found effective in reducing biases and enhancing the reliability of IMERG estimates. The integration of field observations with satellite-based and reanalysis datasets was proven effective in several other ML-based studies (Khan and Bhuiyan, 2021; Sachindra et al., 2018; Zhang et al., 2021a). Building on this foundation, a major objective of this study is to improve precipitation datasets, in line with previous research efforts by Beck et al. (2019, 2017), Chen et al. (2022), Guo et al. (2019), and Rozante et al. (2018).

Our research aims to merge multiple datasets and apply Machine Learning techniques to reconstruct high-resolution daily precipitation fields in the GAR. For the present study, the high resolution and purely observational multisensor IMERG product has been chosen as reference, because of its superior performances in capturing daily rainfall in complex terrain regions (Sharifi et al., 2019; Tang et al., 2020), despite its known limitations (Xiong et al., 2025). IMERG calibrates the frequent and less accurate infrared data with the direct measurements of precipitation from passive microwave sensors, with a final gauge adjustment step (Huffman et al., 2020), leading to a valuable product for scientific use, but it has a limited temporal coverage (from year 2001 to present). Here, using ERA5 variables we target the creation of a long-term dataset (1960–2000) that closely mirrors IMERG's spatial and temporal characteristics, enabling improved detection of climate change signals.

This paper is organized as follows. Section 2 describes the data used as both input and target variables, while Section 3 outlines the architecture of the ML algorithms tested. Section 4.1 presents the performances of the selected ML algorithms and identifies the most relevant atmospheric fields for the reconstruction of the high-resolution precipitation daily fields. Section 4.2 discusses the comparion of the final algorithm using two independent datasets of rain gauges. Finally, Section 5 outlines potential applications of the dataset and future research directions.

2. Data

This study uses ERA5 (Hersbach et al., 2023) which is an atmospheric reanalysis on a 0.25° × 0.25° grid that provides reliable information on the full three-dimensional dynamic and thermodynamic state of the atmosphere as a function of time.

We have selected 24 ERA5 variables to capture the dynamical and thermodynamic state of the atmosphere both at the surface and at 500 hPa (mid-troposphere), together with its vertical stability and some relevant surface forcings (see Table 1). We computed four additional variables to complement our analysis and to account for the geometrical constraints imposed by orographic forcing on surface circulation. These variables are: $q2m$, the specific humidity at 2 m above ground (Eq. (A.1)); bs, the bulk shear (Eq. (A.2)); $camr$, the column-average mixing ratio (Eq. (A.3)); and the water vapor transport wvt, defined as the product of upslope wind and $q2m$ (Eq. (A.5)). Further details are provided in Appendix. Our selection was based on the atmospheric physical reasoning and was informed by previous studies (Grazzini et al., 2021; Morgenstern et al., 2022; González-Jardines et al., 2023).

Terrain elevation as derived from the ETOPO 2022 v1 dataset (NOAA National Centers for Environmental Information, 2022; MacFerrin et al., 2024) has also been used. This dataset is part of the Earth Topography and Ocean Bathymetry (ETOPO) series and is maintained by the National Oceanic and Atmospheric Administration (NOAA). To match the desired dataset spatial resolution, the 30 arc-second topography was conservatively remapped onto IMERG's 0.1° grid. All negative values, representing ocean depths, were set to zero to focus exclusively on land elevation. An overview of the selected variables is provided in the Supplementary Material with daily maps of a randomly selected day (9th of November 2002); see Appendix Fig. A.1).

As a precipitation target product we have used IMERG (NASA Goddard Earth Sciences Data and Information Services Center (GES DISC), 2023), which is developed under the Global Precipitation Measurement (GPM) mission, launched in February 2014. The IMERG algorithm merges data from various sources, including passive microwave and infrared satellite estimates, precipitation gauges (Huffman et al., 2020; Zhang et al., 2021b). IMERG provides global precipitation data from 90°N to 90°S at a 0.1° spatial resolution and 30-minute intervals, from

Table 1
Summary of variables used.

Symbol	Unit	Source	Description
u10	m s^{-1}	ERA5	Zonal (west-east) wind speed at 10 m.
v10	m s^{-1}	ERA5	Meridional (south-north) wind speed at 10 m.
d2m	K	ERA5	Dew point temperature at 2 m.
t2m	K	ERA5	Maximum daily air temperature at 2 m.
cape	J kg^{-1}	ERA5	Potential energy available for convection.
kx	K	ERA5	Stability index indicating thunderstorm potential.
mslhf	W m^{-2}	ERA5	Latent heat transfer at the surface.
msshf	W m^{-2}	ERA5	Sensible heat transfer at the surface.
mvimd	kg m^{-2} s^{-1}	ERA5	Integrated atmospheric moisture divergence.
sp	Pa	ERA5	Atmospheric pressure at the Earth's surface.
tisr	J m^{-2}	ERA5	Solar radiation at the top of the atmosphere.
tclw	kg m^{-2}	ERA5	Liquid water in clouds in a vertical column.
tciw	kg m^{-2}	ERA5	Ice water in clouds in a vertical column.
tcwv	kg m^{-2}	ERA5	Total water vapor in a vertical atmospheric column.
totalx	K	ERA5	Combined index for convection potential.
viewv	kg m^{-1} s^{-1}	ERA5	Eastward water vapor transport.
vinwv	kg m^{-1} s^{-1}	ERA5	Northward water vapor transport.
swvl1	m^3 m^{-3}	ERA5	Volumetric water content of the top soil layer.
t	K	ERA5	Daily mean temperature at 500 hPa atmospheric level.
u	m s^{-1}	ERA5	Zonal wind speed at 500 hPa level.
v	m s^{-1}	ERA5	Meridional wind speed 500 hPa.
w	Pa s^{-1}	ERA5	Vertical air movement rate; negative means rising air.
vo	s^{-1}	ERA5	Relative vorticity
ppt	mm	ERA5	Conservatively remapped total precipitation.
q2m	kg kg^{-1}	Calculated	Specific humidity at 2 m above the surface.
camr	kg kg^{-1}	Calculated	Column averaged mixing ratio.
bs	m s^{-1}	Calculated	Maximum boundary shear.
wvt	m s^{-1}	Calculated	Water vapor transport speed.
ETOPO	m	ETOPO	Topographic elevation from ETOPO dataset.

1998 to present. It is particularly useful in regions that lack ground-based precipitation measurements, such as remote areas. There are three versions: Early Run (4-hour latency), Late Run (12-hour latency), and Final Run (available after 3.5 months), the latter incorporating gauge data from the Global Precipitation Climatology Centre (GPCC, Freitas et al., 2024). This study uses IMERG Final Run Version 07, which supersedes earlier versions of IMERG as well as datasets based on the Tropical Rainfall Measuring Mission (NASA Global Precipitation Measurement Mission, 2024).

IMERG and ERA5 represent state-of-the-art satellite and reanalysis-based precipitation datasets. Both have been extensively evaluated in the literature. While ERA5 performs well in capturing long-term variability and spatial patterns, especially over large regions with complex topography such as the Mongolian Plateau, IMERG generally offers superior accuracy in estimating, on a finer grid, monthly precipitation totals and daily intensity frequencies (Xin et al., 2022). IMERG also tends to exhibit lower error standard deviation compared to ERA5. Although ERA5 typically shows higher daily correlation with ground observations, it often overestimates precipitation, leading to a pronounced wet bias—particularly in areas with steep terrain (*e.g.*, slopes more than 5%), such as parts of Turkey (Amjad et al., 2020). Various studies have highlighted IMERG superior skills in representing daily rainfall properties in complex orography regions (Sharifi et al., 2019; Tang et al., 2020), which is the reason why we select it as a benchmark dataset. Overall, both IMERG and ERA5 are critical resources in hydrological and climatological research (Amjad et al., 2020).

One important feature of the IMERG dataset is the inclusion of a Quality Index (QI), which provides a measure of data reliability (Huffman, 2019). The IMERG Version 07 Quality Index helps users assess the trustworthiness of precipitation estimates point-wise and for each time step. This index offers a practical means to interpret the quality of IMERG data and supports informed decision-making in climate and hydrological research. There are two types of Quality Index: QI_h, which corresponds to half-hourly data, and QI_m, which pertains to monthly data. In this study, we utilized QI_h and computed its daily average, similar to the approach used in Zhao et al. (2023).

QI_h is derived based on the correlation between different satellite retrievals (*e.g.*, microwave and infrared observations), and it is used to weigh their contributions in the merged precipitation product. The QI values range from 0 to 1, where values closer to 1 indicate higher reliability, reflecting denser observational coverage and reduced random error. Fig. A.2 in the Supplementary Material illustrates the distribution of the daily Quality Index QI_d values, as well as their relationship with terrain elevation. Approximately 77% of the data points have QI_d values above 0.6, the threshold suggested for 'good' data (Huffman et al., 2015). Based on this threshold, we selected only these higher-quality data for model development and performance evaluation.

In this work, other datasets were also used for the purpose of comparing the dataset generated by the algorithm: HISTALP, EEAR-Clim and CHAPTER. HISTALP, the Historical Instrumental Climatological Surface Time Series of the Greater Alpine Region, comprises monthly in-situ records of temperature, pressure, precipitation, sunshine, and cloudiness for the GAR (Auer et al., 2007; Zentralanstalt für Meteorologie und Geodynamik (ZAMG)). From the original precipitation dataset, we have randomly selected 50 stations covering the entire study area to extract long-term, rain-gauge-based monthly precipitation records. EEAR-Clim (Extended European Alpine Region Climate) is a major, high-density dataset of daily air temperature and precipitation for the Alps from approximately 9000 stations across the entire region, with record lengths extending up to 200 years and substantially improved high-elevation coverage. This dataset provides an unprecedentedly dense and harmonized observational resource for climate research (Bongiovanni et al., 2024b,a). For our purposes, we have selected daily precipitation data from rain stations with complete time records between 1960 and 2000. In total, 295 stations were selected, covering the entire study area and spanning different elevation ranges. CHAPTER (Computational Hydrometeorology – with Advanced Processing Tools to Enhanced Realism), instead, is a high-resolution hindcast of ERA5 product generated using the Weather Research and Forecasting (WRF) model. It serves as a competitive atmospheric dataset obtained from a dynamical convection-permitting model that includes several atmospheric variables including precipitation. It provides data with a spatial resolution of 3 km×3 km over Central Europe and the Mediterranean, with a temporal coverage spanning 1981–2022 (Tartaglione et al., 2024a,b) and it represents a valuable high-resolution precipitation product over the region of interest (Bernini et al., 2025). Finally, in Section 4.2 two versions of ERA5 have been used in the comparison; the ERA5 at its original resolution and that one we refer to as ERA5 downscaled. The latter is obtained through a conservative remapping at $0.1°$ on the IMERG grid.

3. Methods

Given the use of multiple data sources with different spatial resolutions, all ERA5 variables, whose spatial resolution is 0.25°, were regridded onto IMERG's grid (0.1°) using nearest-neighbor interpolation, except for precipitation, which was downscaled with conservative remapping, computed with Climate Data Operators (CDO). The latter

operation was made to preserve the area integral, avoiding distortions of such a highly variable field. Additional tests have shown that the use of conservative remapping on the relevant input fields (as described in the next section) has a very minor impact on the final skills (not shown).

To maintain a consistent number of days per year, February 29th was removed, resulting in a total of 7300 days (365 days per year over 20 years, from January 1, 2001, to December 31, 2020). The dataset includes 150 gridded values for longitude, ranging from 4.05° E to 18.95° E, and 60 gridded values for latitude, ranging from 43.05° N to 48.95° N. This dataset was partitioned into training (60%), validation (20%), and test (20%) subsets using a random sampling approach applied at the daily level. Specifically, individual days were randomly assigned to each subset, ensuring temporal independence across the splits. This approach contrasts with sequential time-based partitioning, which was intentionally avoided to reduce potential temporal biases due to climate change signals. After partitioning, pixels associated with an IMERG daily Quality Index below 0.6 were excluded across all variables.

As shown in Fig. A.2, 23.04% of observations fall below the QI = 0.6 threshold, leaving 76.96% of the dataset (training, validation, and test) for analysis. In panel (c), within the training subset, the share of samples with $QI \geq 0.6$ is relatively high and flat at low–mid elevations (0–50th percentile: ≈ 79–82%), declines modestly through the mid–high percentiles (50–80th: ≈ 76–79%), and then drops more sharply in the upper tail (80–90th: ≈ 72–73%; highest decile: ≈ 52%). This pattern indicates that higher elevations tend to have lower QI, yielding fewer usable samples.

Among various machine learning algorithms, Extreme Gradient Boosting (XGB) (Chen et al., 2025) was selected for its high efficiency, reduced computational time, and competitive error rates when compared to alternative models.

Hyperparameter tuning is often considered a crucial step in optimizing XGB's performance. In our study, we have tested several advanced optimization techniques, including Optuna (Optuna Contributors, 2024), TPOT (Tree-based Pipeline Optimization Tool) (Olson and Moore, 2016), and Random Search (Bergstra and Bengio, 2012), across a broad range of parameters. We have adjusted the number of estimators (50 to 250), explored learning rates on a logarithmic scale (0.001 to 1), varied maximum tree depth (3 to 21), fine-tuned minimum child weights (1 to 10), and tested different subsample ratios (0.6 to 1.0). The performance improvement observed was negligible; therefore, the default settings were deemed sufficient, and further hyperparameter optimization was not pursued.

The model was trained using 100 boosting rounds (n-estimators) with a learning rate of 0.3 to regulate the contribution of each tree, and a maximum tree depth of 6 to limit model complexity. The minimum child weight was set to 1.0, ensuring that each leaf node had sufficient instance weight (based on the Hessian) to be split. To accelerate training on multi-core machines, the number of parallel threads (n-jobs) was set to −1, allowing XGB to use all available CPU cores. For reproducibility across different runs, we fixed the random seed by setting random-state to 42, ensuring consistent results.

In this study, we have employed three common metrics to evaluate the regression models: the coefficient of determination (R^2), the root mean square error/difference (RMSE/RMSD), and the mean error (ME). The R^2 value measures how well the predicted values approximate the actual data, while the RMSE/RMSD quantifies the average magnitude of the prediction errors. The ME, on the other hand, captures the average deviation between predictions and observations, indicating whether the model systematically overestimates or underestimates the values, and is used for comparing time series across different datasets. Note that performance metrics were computed using the test portion of the dataset, which was held out during model training. These metrics are defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{1}$$

$$\mathrm{RMSE} = \sqrt{(1/n)\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{2}$$

$$\mathrm{ME} = (1/n)\sum_{i=1}^{n}(y_i - \hat{y}_i), \tag{3}$$

where y_i and $\hat{y}_i$ are the observed and predicted ith rainfall values, n is the sample size and $\bar{y} = (1/n)\sum_{i=1}^{n} y_i$ is the observation sample mean. RMSD is identical to RMSE, except that it measures deviations between two model datasets instead of errors relative to observations.

To reduce computational complexity, accelerate model training, mitigate overfitting, enhance generalization and improve model interpretability, we employed SHAP (Lundberg and Lee, 2017) for feature selection. SHAP is a model feature importance framework rooted in cooperative game theory. It offers several advantages over traditional feature selection techniques as it is compatible with a wide range of ML models, provides consistent and locally accurate explanations, and offers intuitive visualizations of feature importance. The SHAP methodology proceeds in two stages. First, a predictive model is trained using the complete set of features. Then, SHAP values are computed for each feature, quantifying their individual contributions to the model's output. These values are subsequently used to rank the features by importance, highlighting those that most significantly influence the prediction of the target variable (Wang et al., 2024).

Compared to other feature selection methods, such as Recursive Feature Elimination (RFE) (Gensini et al., 2021), that may discard variables that appear weak in isolation but are actually valuable when interacting with other features, SHAP considers all possible feature combinations, providing a more comprehensive assessment of feature importance. Moreover, while RFE is computationally efficient and well-suited for small datasets or simpler models, it offers limited interpretability and is typically employed when performance optimization is the sole objective. In contrast, SHAP, although computationally demanding — particularly for deep learning models — is applicable in this study due to the lower computational cost of the XGB model relative to neural networks. Finally, since this research emphasizes explainable artificial intelligence in addition to predictive accuracy, SHAP provides the interpretability that RFE alone cannot offer.

In this study, the XGB model was initially trained using all available features in the training dataset. Then, SHAP values were then computed for each feature across the validation set, capturing their contributions to the model's predictions on a per-instance basis. By averaging these SHAP values over all samples, we obtained a global ranking of feature importance. This approach allows SHAP to assign additive importance scores to each input feature, yielding a transparent and interpretable understanding of their relevance. A SHAP value for a specific feature represents how much that feature contributed to increasing or decreasing the model's predicted precipitation (mm/d), relative to the baseline prediction, defined as the mean model output when using all features across the entire training dataset.

To determine the optimal number of input variables, corresponding to the smallest set yielding the largest improvement in performance, we computed the RMSE/RMSD for different models in which the number of input variables was progressively increased according to their SHAP-based ranking. When the change in RMSD becomes smaller than 0.5% the process was terminated.

4. Results and discussion

4.1. Model validation

Before proceeding to present and discuss the results obtained, we first introduce the most relevant variables (Fig. 1). To gain a comprehensive spatial understanding, remapped on the IMERG 0.1° resolution grid, we show time averages of: the target precipitation data (IMERG), the baseline precipitation (ERA5), the column-average mixing ratio

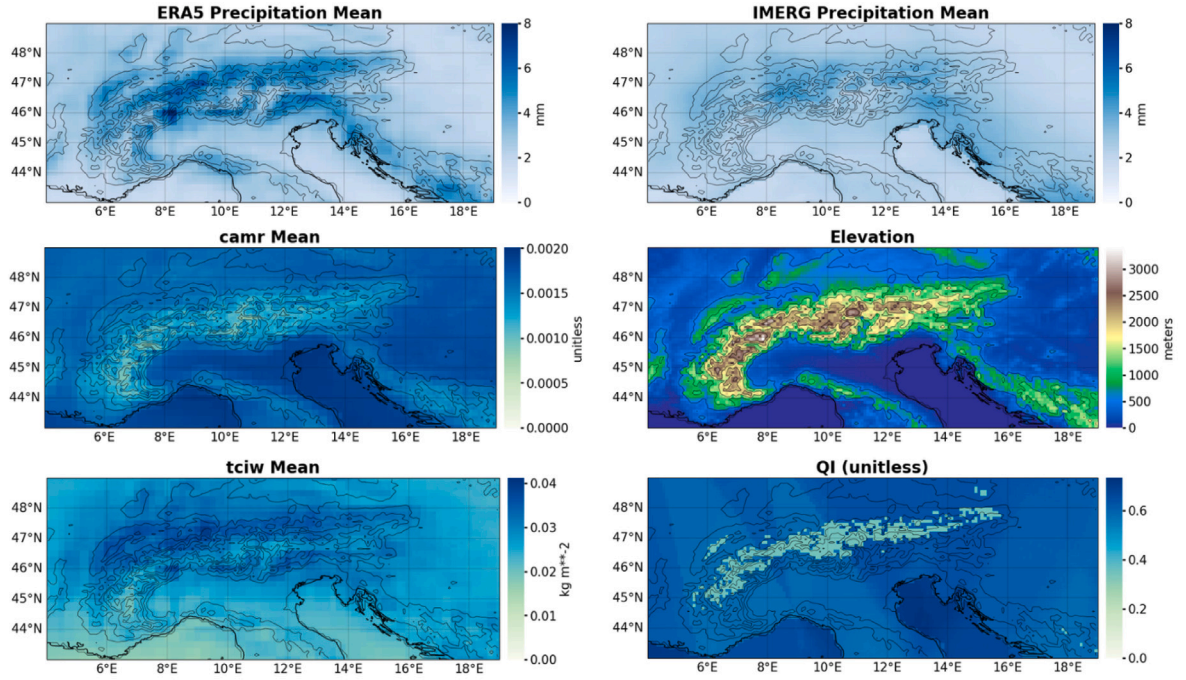


Fig. 1. Temporal daily mean (2001–2021) for (a) ERA5 and (b) IMERG precipitation, (c) camr, (d) terrain elevation, (e) tciw and (f) QI_d.

(camr), the total column ice water (tciw), topographic complexity (Elevation), and data reliability (IMERG QI_d). Since all these variables are directly or indirectly linked to terrain elevation, this overview aids in interpreting the physical and observational landscape of the study domain. In particular, it is interesting to note that QI_d is on average lower over high-elevation areas, meaning that more data are discarded in this area from the training (See also Fig. A.2). However, as discussed in Section 5, this does not compromise the skill of the model.

The ERA5 precipitation (ppt-era5), camr, and tciw emerged as the most influential variables based on the SHAP analysis (Fig. 2). For the sake of completeness, Fig. A.3 shows that the XGB feature importance method identifies similar relevant variables.These are the three variables that contribute the most to the skills of the model in terms of RMSD. In fact, as additional variables are incorporated, the RMSD decreases, indicating an improvement in the model's skills. The elbow shape of the RMSD curve reveals a sharp drop for the first three variables, where RMSD decreases from 5.1 mm/d to approximately 4.4 mm/d. Beyond this point, the reduction in RMSD becomes marginal, suggesting that additional variables have limited impact on model performance. This implies that a core subset of variables captures the majority of the model's predictive skill. From the physical point of view, it is reasonable that ERA5 coarser resolved precipitation, together with the vertical integrals of water vapor and of ice content are the strongest predictors of fine scale precipitation. It implies that ERA5 precipitation remapped onto the finer grid can valuably be corrected taking into account moisture, which affects raindrop evaporation and thus ground-reaching precipitation, as well as ice crystals, whose presence significantly accelerates the formation and the falling speed of both cold and mixed-phase clouds (Pruppacher and Klett, 2010). Therefore, the selected input features for subsequent modeling were tciw, ppt-era5, and camr. Using this minimal set of input variables and using 60% of the data the XGB training required around 2.5 min, excluding preprocessing.[1]

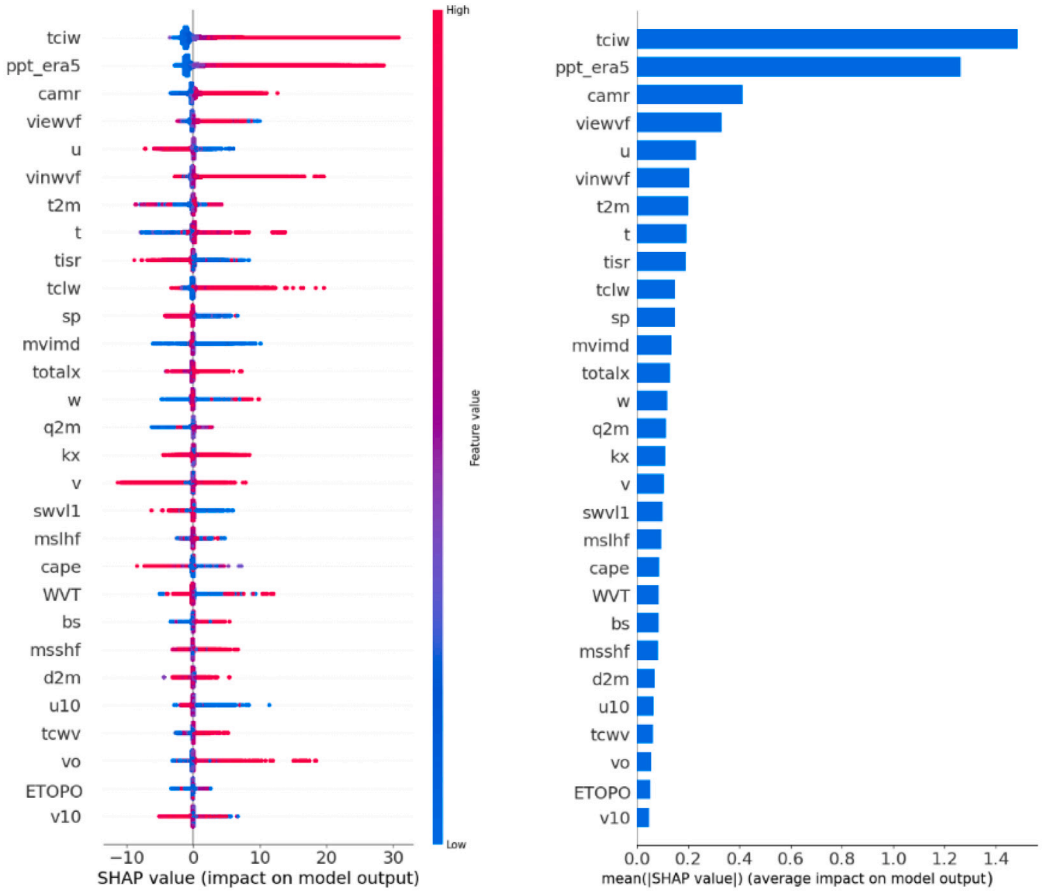
The output of our XGB-trained model, built using the three most influential ERA5 features, forms the foundation of the ML-IMEX (Machine Learning-based IMERG backward-EXtended) precipitation dataset, designed to reconstruct high-resolution daily rainfall fields for the pre-satellite era.

To evaluate its skills, we used the IMERG test subset as reference and computed both the RMSD and the R^2 of ML-IMEX. Those values were then compared with those obtained using as a predictors (1) the IMERG long-term climatology and (2) randomly selected IMERG fields, keeping IMERG as reference. The RMSD calculated using a randomly extracted precipitation map from IMERG data resulted in a $R^2 \approx 0$ and an RMSD of approximately 9.6 mm/d. Assessments based on long-term climatology, defined as the temporal average from 2001 to 2021, yielded RMSD of 6.8 mm/d and a $R^2 \approx 0$. This proves that ML-IMEX is more skillful than a random predictor or a climatological one, as it has a RMSD of 4.4 mm/d and a R^2 of 0.59.

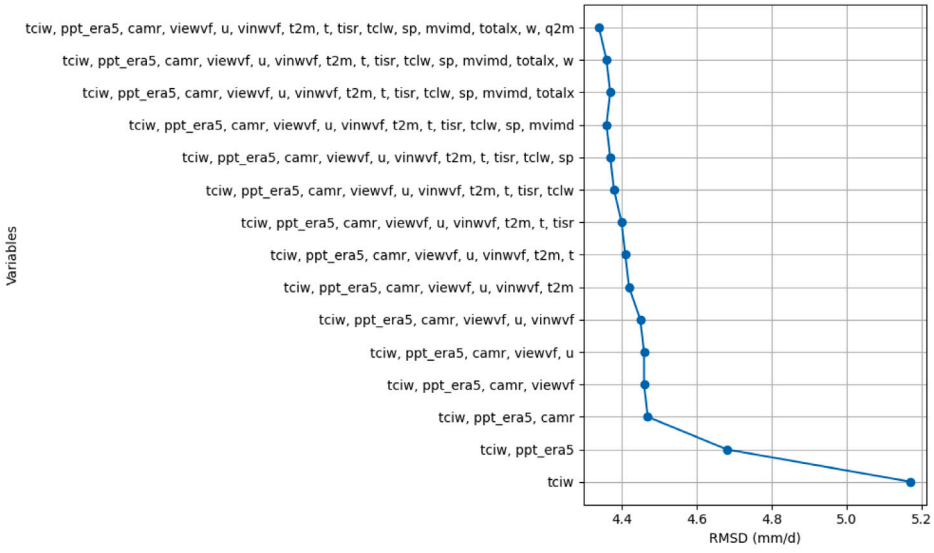
As a further step, we compare it with ERA5 skill, to identify the areas within the domain where the dataset exhibits lower RMSD and thus a better agreement with the IMERG reference. The spatial distribution of the RMSD values in the alpine region, calculated for each point over time, shows that the ML-IMEX performs well in most of the areas studied (Fig. 3). In terms of spatial–temporal RMSD between IMERG and ERA5, the total RMSD across all seasons is approximately 5.13 mm/d. Breaking this down seasonally, spring exhibits an RMSD of 4.49 mm/d, summer rises to 5.63 mm/d, fall reaches the highest value at 5.75 mm/d, and winter presents the lowest RMSD at 4.49 mm/d. The comparison between IMERG and ML-IMEX demonstrates improved accuracy, in which the overall RMSD is reduced to 4.39 mm/d. Seasonally, this model achieves an RMSD of 3.87 mm/d in MAM, 4.8 mm/d in JJA, 5.02 mm/d in SON, and 3.92 mm/d in DJF (Fig. 3).

Higher RMSD values, highlighted in red, are concentrated in specific eastern and central regions. This may be attributed to complex terrain or localized precipitation patterns, as illustrated in Fig. 3, which shows the long-term climatology of the IMERG precipitation dataset. These localized features present challenges for the model to accurately capture, highlighting the need for further investigation. In particular, the RMSD of ML-IMEX remains relatively high in summer over the central and eastern part of the Alpine arc, likely due to a mis-representation of

[1] This was done on a system with the following configuration: 13th Gen Intel(R) Core(TM) i7-1355U @ 1.70 GHz processor and 16.0 GB of installed RAM.



(a) (b)



(c)

Fig. 2. Feature Selection results including (a) SHAP values, (b) mean SHAP values, and (c) models' RMSD.

intense convective rainfall, as discussed in what follows. This should be taken into account when using this product for climate and hydrology applications, to properly propagate the uncertainty.

A comparison of RMSD (mm/d) across IMERG precipitation percentile ranges for both ERA5 and ML-IMEX indicate that ML-IMEX outperforms ERA5 across most percentiles, except for the highest percentile, which corresponds to precipitation larger than 14.5 mm/d, in which they have similar performances (Fig. 4). Overall, it is found that the RMSD tends to increase with increasing mean precipitation. Similarly, by comparing ML-IMEX and ERA5 skills as

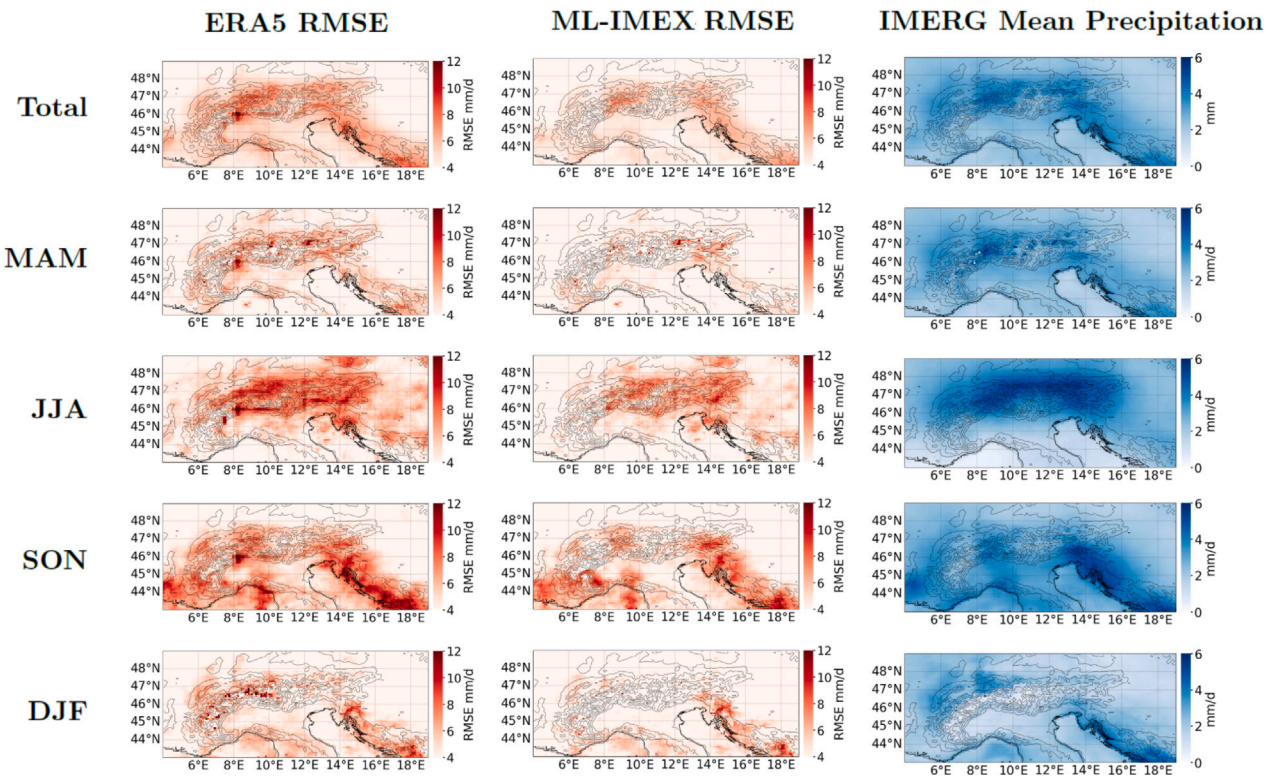


Fig. 3. Seasonal analysis of precipitation: (left) RMSD between IMERG and ERA5 on the test dataset, (middle) RMSD between IMERG and ML-IMEX, and (right) the corresponding IMERG mean precipitation (2001–2020). Rows represent annual totals and individual seasons—Spring (MAM), Summer (JJA), Fall (SON), and Winter (DJF). Terrain elevation contours are shown at 500 m intervals.

a function of different terrain elevation percentile classes, ML-IMEX outperforms ERA5. The RMSD for ERA5 increases with elevation, exceeding 6 mm/d in the highest elevation classes (namely for terrain elevation above 920 m). In contrast, the ML-IMEX shows a more gradual increase, with RMSD remaining below 5.5 mm/d even at the highest elevations. Both models achieve their lowest RMSD values at low elevations (around 100–180 m).

It is worth noting that in the XGB modeling process, training a single model on the entire year's data yielded performance that was largely comparable to models trained separately for each season. In fact, the unified annual model slightly outperformed the season-specific models in most cases. For instance, during spring, the RMSD of the annual model was 3.90 mm/d, compared to 3.93 mm/d when trained solely on spring data. In summer, however, the seasonal model performed slightly better, with an RMSD of 4.67 mm/d versus 4.82 mm/d from the annual model. Fall results were nearly identical, with RMSDs of 5.01 mm/d (annual model) and 5.02 mm/d (seasonal model). Similarly, in winter, the difference was minimal, with the annual model achieving an RMSD of 3.83 mm/d, slightly better than the seasonal model's 3.85 mm/d. Consequently, we proceeded with the first approach—training the XGB model on the full annual dataset.

4.2. Benchmarking

Up to this point, the performance of XGB was evaluated in terms of RMSD and R^2 using the test dataset (20%). After training, the model was used to generate a historical precipitation dataset for the period 1960–2000 that we named ML-IMEX-GAR. This dataset represents a machine learning reconstruction of IMERG-like precipitation on the Greater Alpine Region obtained with the ML-IMEX algorithm described above.

To compare the generated historical data, we used the HISTALP, EEAR-Clim and CHAPTER datasets. To avoid interpolation issues in the HISTALP gridded product, we have used the original station data. For each gridded dataset (ML-IMEX-GAR, IMERG and CHAPTER) we have selected the closest grid point to each HISTALP or EEAR-Clim station for comparison. The 50 randomly selected HISTALP stations used as comparison cover the entire GAR (Fig. 5).

The monthly values of the closest grid point to each HISTALP station were extracted, and the mean across all stations was computed to construct the time series (Fig. 6a) (period 1960–2021 selected for visualization purposes). We then computed the difference between all the datasets considered and HISTALP time-series (Fig. 6c); this indicates that ML-IMEX-GAR (purple line) aligns more closely with the reference HISTALP data (yellow line in panel a) than both CHAPTER (black line) and ERA5 (blue line), as indicated in the legend by the values of mean error. Not only the mean of the time series is different but also the higher moments of the distributions indicate that CHAPTER is statistically very close to ERA5, whereas ML-IMEX-GAR resembles the reference HISTALP dataset (Fig. 6b).

To quantify the performance of the ML-IMEX-GAR during the historical period (1960–2000) using HISTALP as a reference, the coefficient of determination (R^2) together with the RMSE were calculated and compared with the same from ERA5 at original resolution, ERA5 downscaled at 0.1° on IMERG grid and CHAPTER over the entire domain and for three different elevation classes: low (< 450 m), mid (450–900 m) and high (> 900 m). (Fig. 6d,e). Overall the two ERA5s and CHAPTER exhibit the lowest R^2 values, 0.44, 0.43 and 0.67 respectively,indicating weak correlations with the HISTALP data. In contrast, the ML-IMEX-GAR shows a significantly higher R^2 of 0.87, demonstrating strong correlation and improved predictive skill. Also, the RMSE is maximum for ERA5, it lowers a bit for CHAPTER and then it drops for ML-IMEX-GAR. Among all datasets, ERA5 displays the weakest agreement with HISTALP, both in terms of RMSE and R^2. CHAPTER, being a high-resolution dynamical downscaling of ERA5 shows a slightly stronger agreement with the in-situ observation. However, looking at

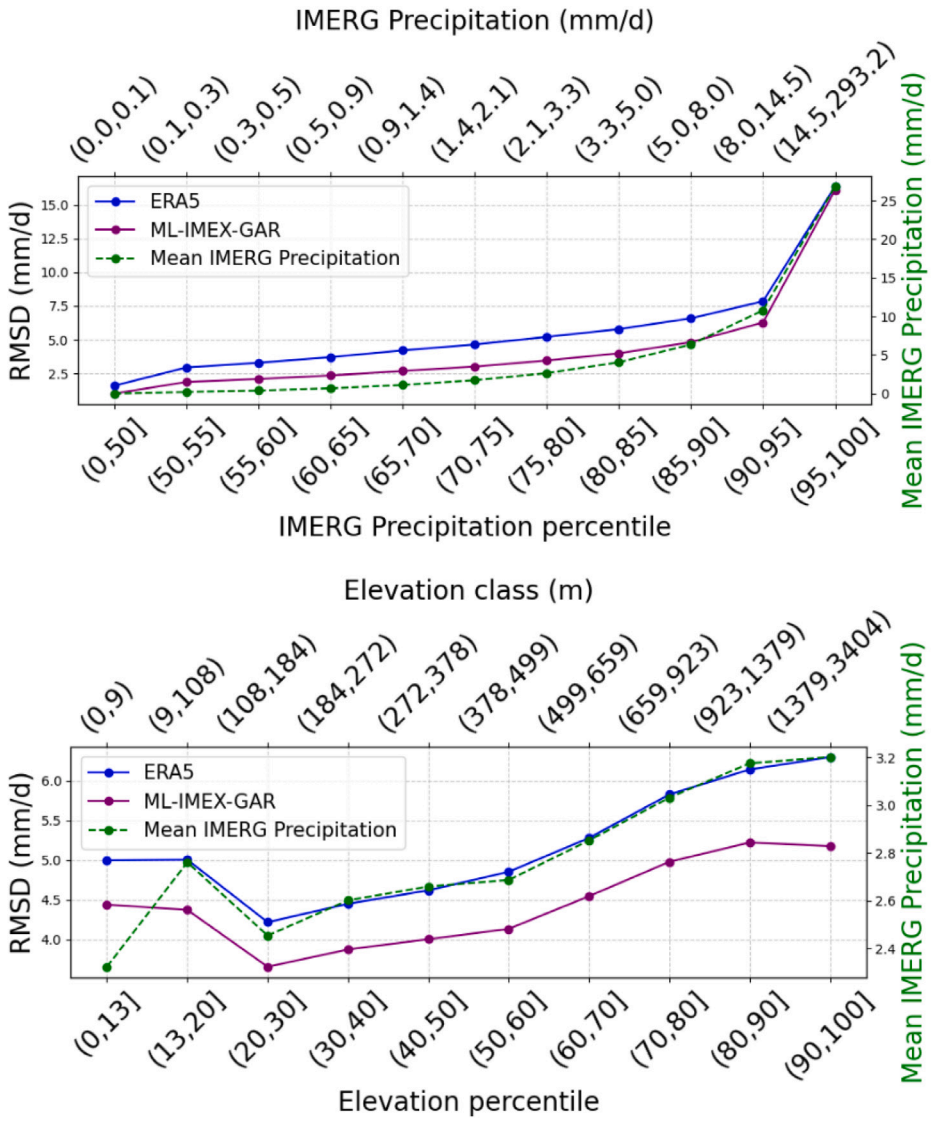


Fig. 4. ML-IMEX performances in terms of RMSD for different IMERG precipitation classes and terrain elevation classes.

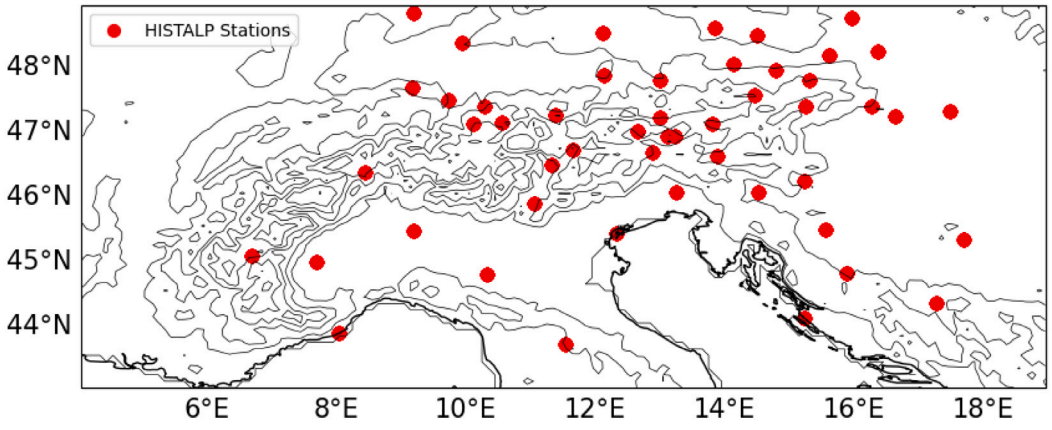


Fig. 5. HISTALP 50 randomly selected stations to validate the model performances on a monthly time scale. Elevation contours are provided every 500 m.

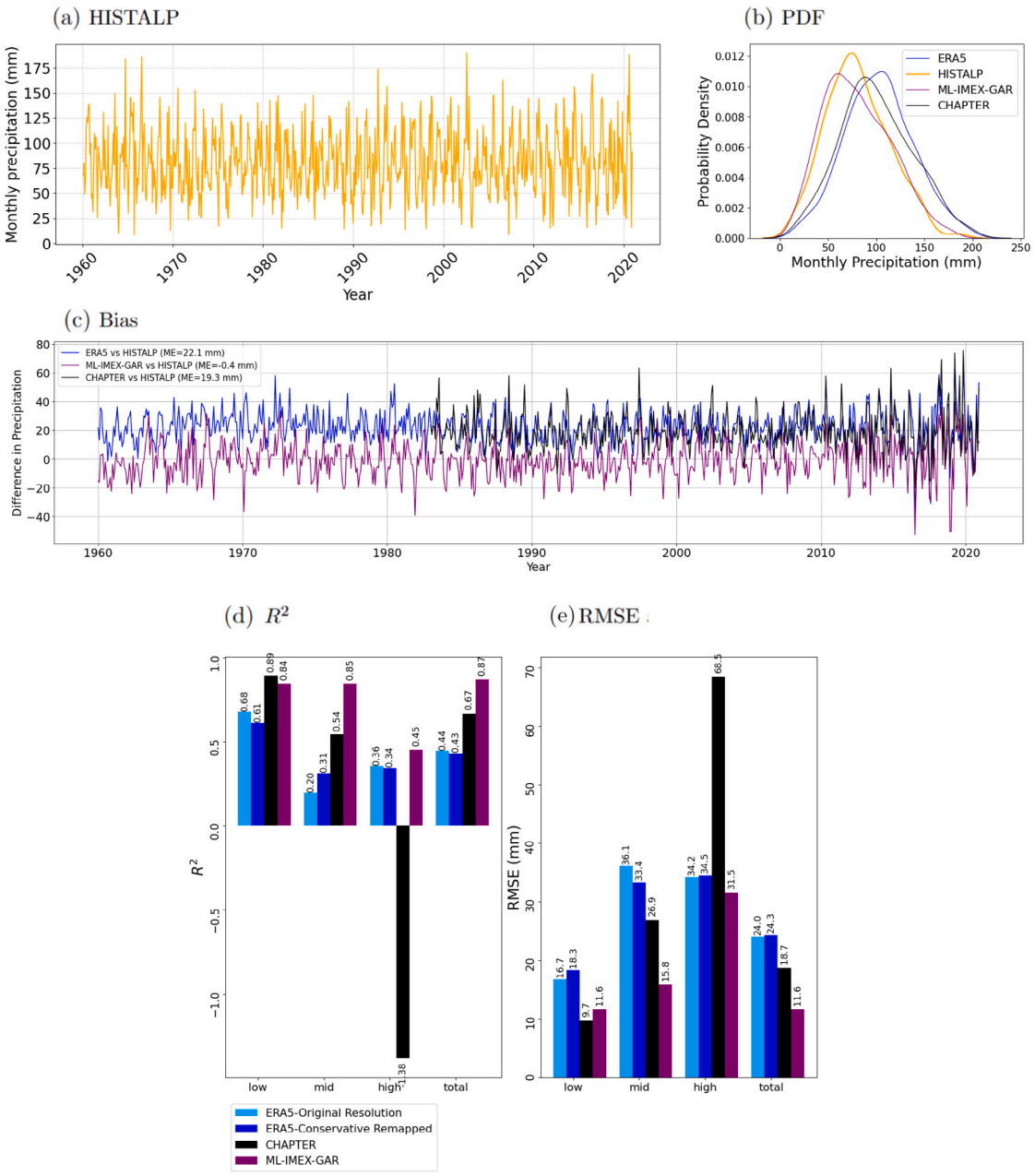


Fig. 6. Overview of (a) the HISTALP monthly precipitation time series; (b) probability density functions of monthly precipitation for datasets; (c) monthly precipitation biases relative to HISTALP; (d) R^2; and (e) RMSE across the evaluated datasets.

the elevation dependence, the high-elevation RMSE is larger than at other elevation ranges. This is probably due to the double penalty issue that typically affects high-resolution numerical models, and which is particularly impactful in mountainous regions. ML-IMEX-GAR, finally, shows the best agreement with the in-situ data at all elevation, both in terms of bias, R^2 and RMSE. It outperforms the dynamical downscaling produced by CHAPTER for all statistical metrics considered.

To extend the comparison of our dataset to the daily scale, we have repeated the analysis performed with HISTALP using EEAR-Clim as reference. We have selected daily precipitation data from 295 stations within the study domain and period (Fig. 7). As before, we have computed the PDFs, bias, R^2 and RMSE over the entire domain and for different elevation bands (Fig. 8).

In this case, the dataset that shows the closest bias to EEAR-Clim is CHAPTER (Fig. 8)c. However, ML-IMEX-GAR proves to be highly representative of the daily in-situ observations, exhibiting the highest R^2 (0.74) and the lowest RMSE (Fig. 8d-e). This result holds for both low and mid-elevation classes, whereas at high elevations the best performance is obtained by the downscaled ERA5. Unlike in the previous comparison, both downscaled ERA5 and CHAPTER show performance comparable to ML-IMEX-GAR (R^2 are 0.68 and 0.73 respectively), while ERA5 at its original resolution deviates substantially from EEAR-Clim (R^2 = 0.14).

Despite the fact that ML-IMEX was not trained to reproduce the in-situ observations from HISTALP or EEAR-Clim, it is able to generate monthly and daily precipitation data that closely match these observational records.

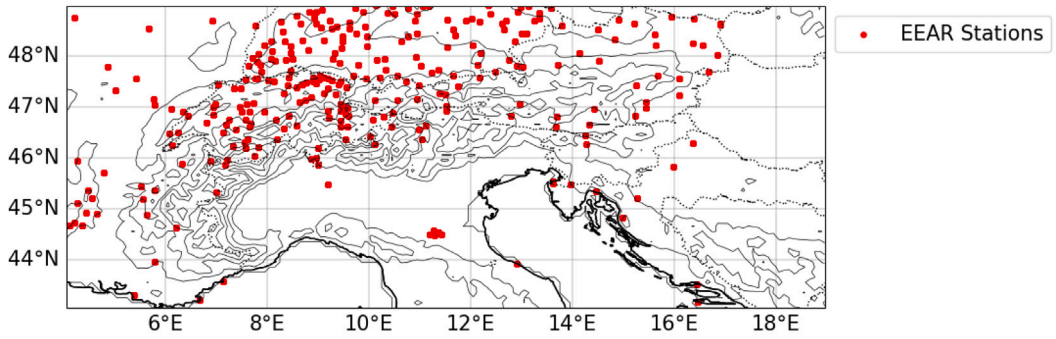


Fig. 7. Spatial distribution of the 295 EEAR-Clim stations containing full, uninterrupted records from 1960 to 2000. Elevation contours are depicted at 500-m intervals.

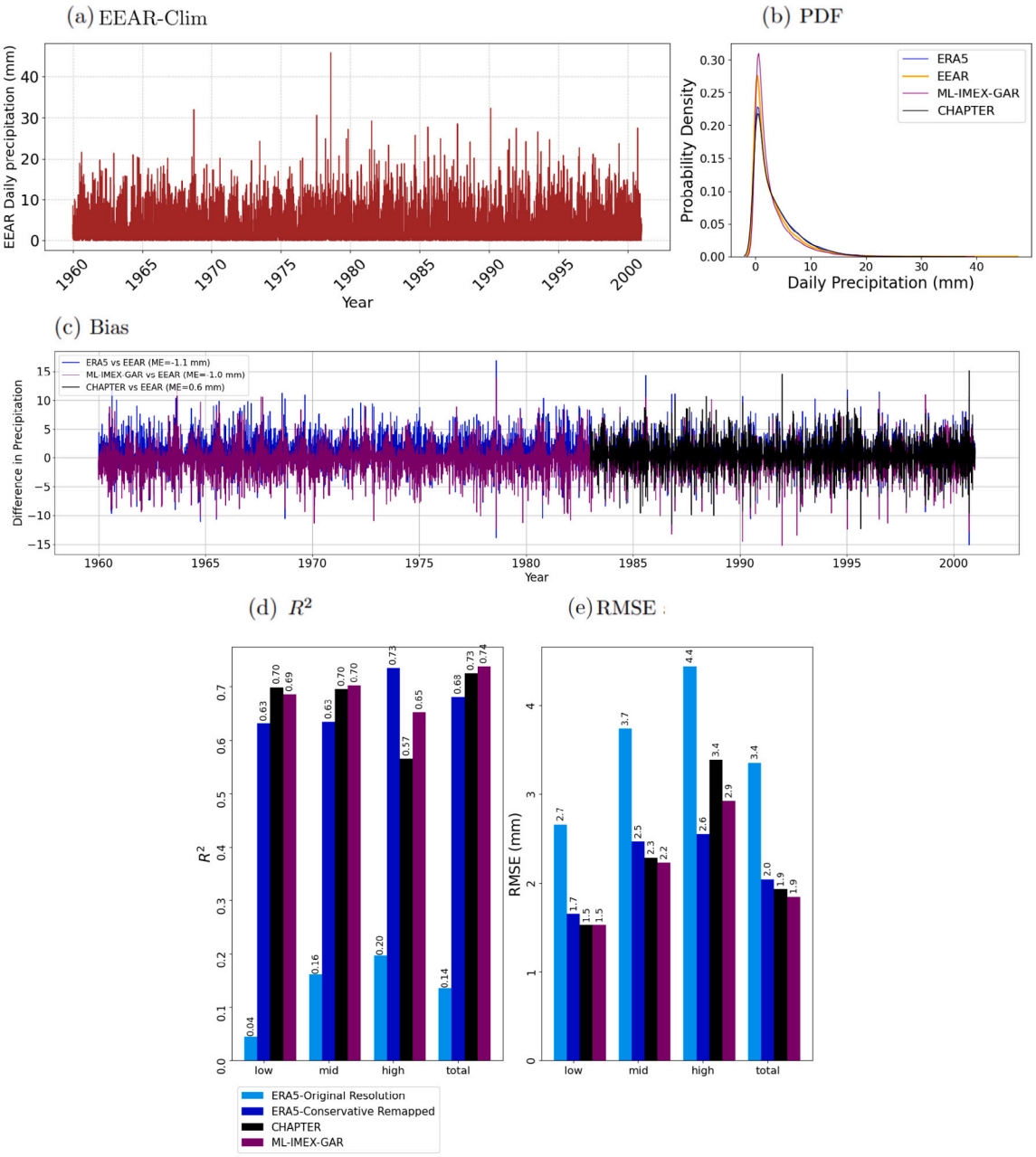


Fig. 8. Overview of (a) the EEAR daily precipitation time series; (b) probability density functions of daily precipitation for datasets; (c) daily precipitation biases relative to EEAR; (d) R^2; and (e) RMSE across the evaluated datasets.

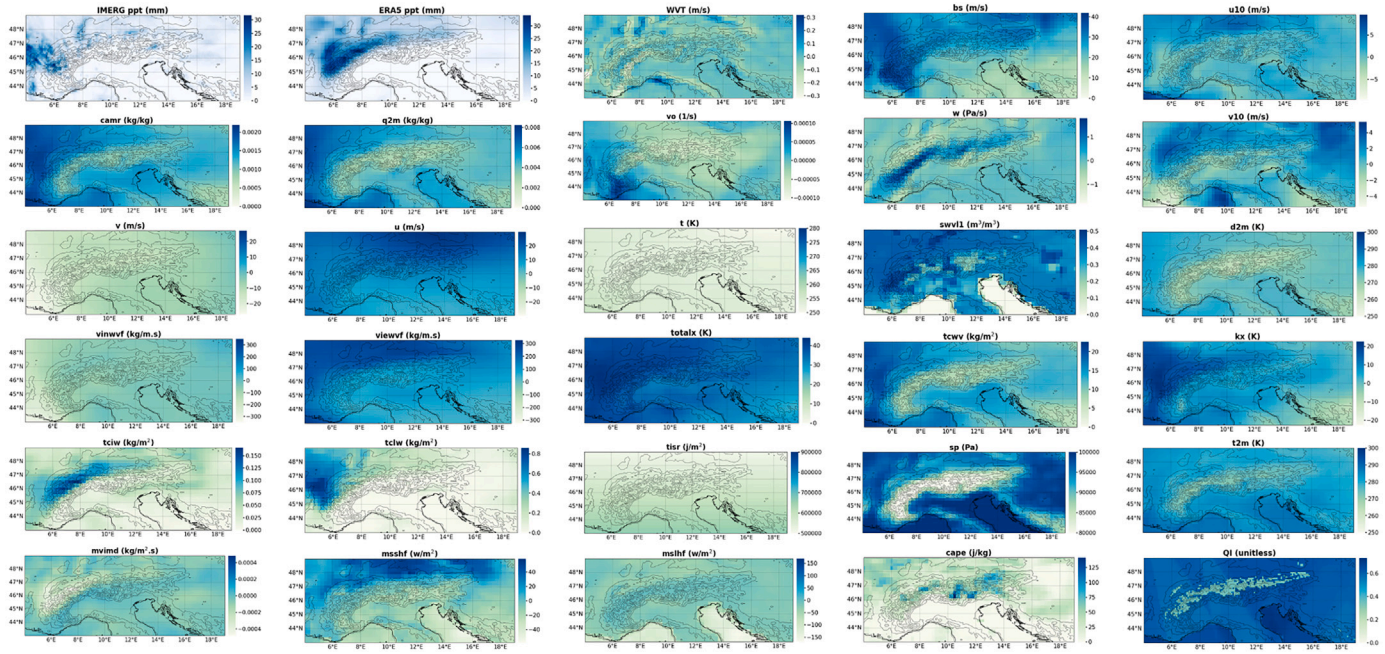


Fig. A.1. Spatial distribution of the considered variables on 9 November 2002 (randomly selected day).

5. Conclusion and future work

This study presents a new daily precipitation dataset (ML-IMEX-GAR) and the ML algorithm we have developed to generate such a dataset. ML-IMEX-GAR is a backward temporal extension of IMERG over the Greater Alpine Region from 1960 to 2020. By incorporating atmospheric variables from ERA5 and applying feature selection methods like SHAP, we have identified three key predictors which are the total column ice water, ERA5 precipitation, and column-average mixing ratio. Using these, we have developed a high-performing model that more accurately reconstructs IMERG-like precipitation. This model shows very good performances with respect to in-situ monthly and daily precipitation data and outperforms a corresponding dynamical downscaling forced by ERA5 (CHAPTER), with a much lower computational burden. ERA5 at its original spatial resolution shows the weakest agreement with HISTALP and EEAR-Clim, CHAPTER improves upon it, and ML-IMEX-GAR performs best, with the lowest bias and RMSE and the highest R^2. This is particularly true at low and mid elevations, while at higher elevations this superiority is confirmed only when compared against monthly data (HISTALP); nevertheless, even in the comparison with daily in-situ observations (EEAR-Clim), ML-IMEX-GAR achieves performances comparable to the downscaled ERA5 which has been found to best represent the reference.

We have delivered a high-resolution daily precipitation dataset with a significantly extended temporal coverage. These characteristics make ML-IMEX-GAR particularly well-suited for climate studies in regions with complex orography, where observational data often suffer from sparse coverage and spatial inhomogeneity. The advantage of the method is that the ML approach is trained with a large amount of data available at lower elevation, and is then applied both at low and at high elevations, showing good skills in producing a reliable dataset in both regions.

We acknowledge that IMERG has certain limitations. In particular, precipitation estimates in regions with complex orography (Pradhan et al., 2022) may be less reliable, as steep terrain can cause signal mixing across neighboring surface pixels. Furthermore, since IMERG applies gauge correction, its accuracy is constrained by the limited availability of rain gauges in mountainous areas. IMERG has also been shown to underestimate heavy precipitation events (*e.g.*, Aksu et al., 2023; Xiong et al., 2025). Therefore, incorporating additional reference datasets could provide a more robust basis for model training.

These findings highlight the effectiveness of ML in bridging gaps in precipitation records, especially in complex terrains where ground measurements are sparse. The proposed method not only enhances the spatial and temporal continuity of precipitation datasets but also offers a valuable tool for hydrological modeling and climate studies in data-scarce regions. Future research may explore the incorporation of spatially-aware models or deep learning techniques to further improve performance in topographically complex areas.

For future development, this study holds strong potential to be expanded in several impactful directions. Although ML-IMEX-GAR demonstrated solid predictive performance, it does not account for spatial dependencies, making the inclusion of spatially-aware machine learning models, such as Convolutional Neural Networks or Conditional Generative Adversarial Networks (CGANs), a promising next step to enhance spatial coherence. Future work could also focus on improving model accuracy during high-precipitation events through advanced ensemble or hybrid modeling approaches. Enhancing the spatial or temporal resolution of the dataset, for instance by moving to sub-daily scales or finer grids, could further refine its applicability. Additionally, applying the model to other mountainous regions would help evaluate its transferability to topographically diverse environments. Integrating the ML-IMEX-GAR dataset into hydrological models could provide valuable insights into its influence on streamflow, water resources, and climate impact assessments.

CRediT authorship contribution statement

Iman Goudarzi: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **Davide Fazzini:** Writing – review & editing, Validation, Supervision, Software, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Claudia Pasquero:** Writing – review & editing, Supervision. **Agostino Niyonkuru Meroni:** Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **Matteo Borgnino:** Writing – original draft, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

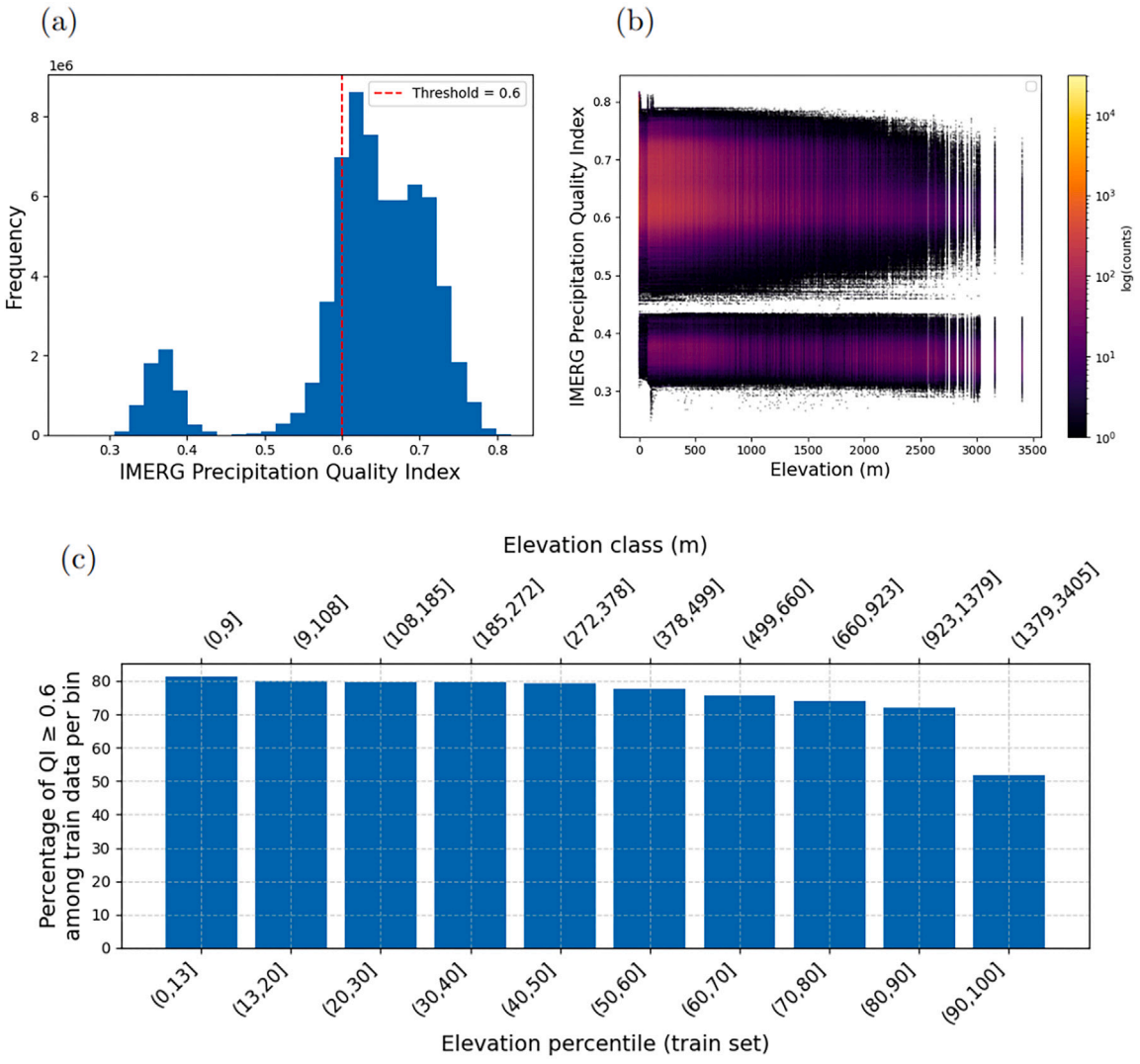


Fig. A.2. IMERG Quality Index analysis.

Funding

This work is supported by the MAXM project funded within the framework of the Bicocca Starting Grant 2023 program of the University of Milano-Bicocca, Italy. Agostino Niyonkuru Meroni was partially supported by OGS and CINECA under HPC-TRES program award number 2023–04, CUP H45E23000410001. Matteo Borgnino has been supported by the European Union - Next Generation EU, Mission 4 Component 1 CUP H53D23011300001, project LocCLIMA.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

Surface specific humidity is obtained from the surface dew-point temperature using the Clausius-Clapeyron equation, namely

$$q_{2m} = \frac{0.622 \times \left(611.12 \times \exp\left[5420 \times \left((1/273.15) - (1/d_{2m})\right)\right]\right)}{sp - 0.378 \times \left(611.12 \times \exp\left[5420 \times \left((1/273.15) - (1/d_{2m})\right)\right]\right)} \quad \text{(A.1)}$$

q_{2m}: Specific humidity at 2 m above ground, representing the water vapor content in the air.

611.12: Reference value of vapor pressure over a flat water surface at the freezing point of water (in Pascals).

5420: Approximation of the latent heat of vaporization divided by the gas constant for water vapor, commonly used in the Clausius–Clapeyron equation.

d_{2m}: Dew point temperature at 2 m above ground (in Kelvin).

sp: Surface pressure (in Pascals) (see Figs. A.1 and A.3).

$$\text{bs} = \sqrt{\left(u_{925} - u_{500}\right)^2 + \left(v_{925} - v_{500}\right)^2} \quad \text{(A.2)}$$

Bulk Shear: The difference in wind velocity between two pressure levels.

u_{500}, u_{925}: west to east wind component at pressure levels of 500 hPa and 925 hPa, respectively.

v_{500}, v_{925}: south to north wind component at 500 hPa and 925 hPa.

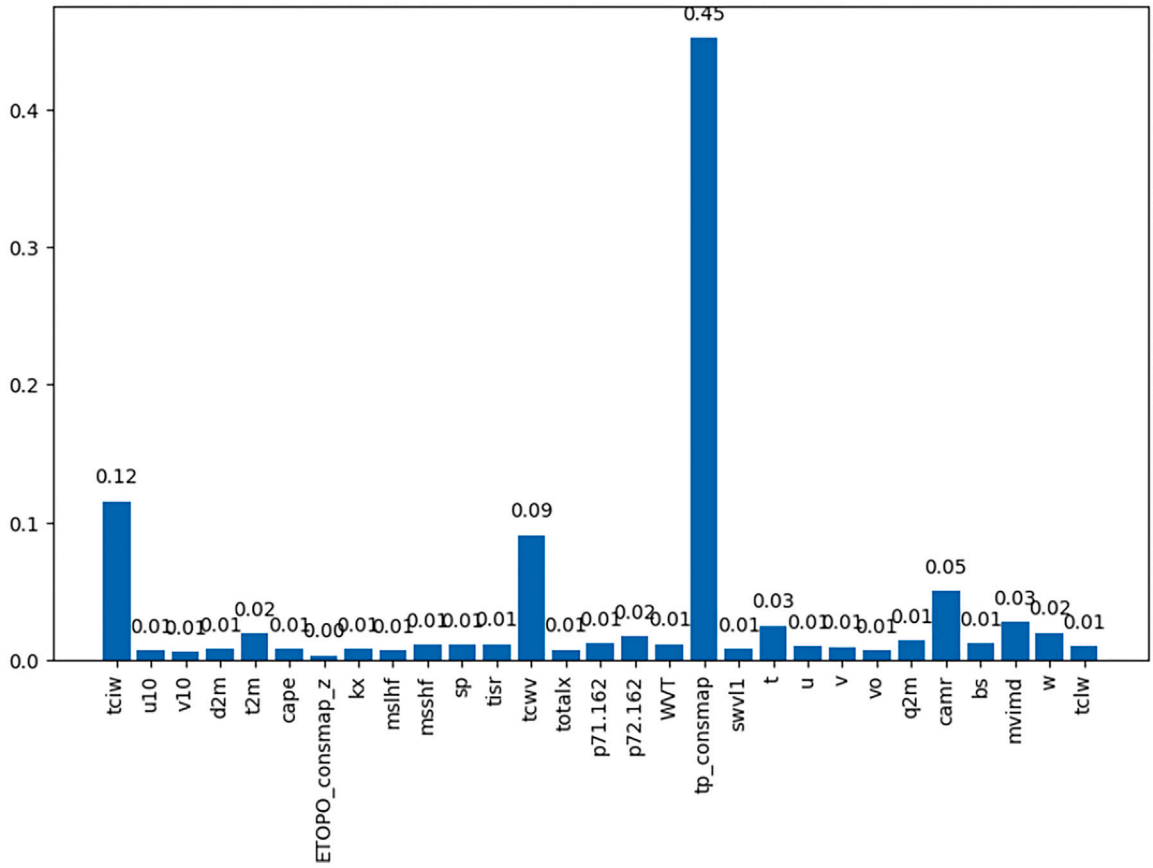


Fig. A.3. Feature Importance Analysis.

$$\text{camr} = \frac{\text{IntegratedWaterVapor}}{SP/9.81} \tag{A.3}$$

camr: Column-Average Mixing Ratio

sp: Surface pressure.

9.81: Gravitational constant (m/s^2)

$$\text{Up slope Wind} = U_{10}\,(\partial h/\partial x) + V_{10}\,(\partial h/\partial y) \tag{A.4}$$

$$\text{Water Vapor Transport (wvt)} = \text{upslope wind} \times q_{2m} \tag{A.5}$$

u_{10} and v_{10}: Wind vector at 10 m.

$\frac{\partial h}{\partial x}$ and $\frac{\partial h}{\partial y}$: Gradients of terrain elevation in the x- and y-directions, respectively.

Data availability

The ML-IMEX-GAR dataset generated in this study is publicly available on Zenodo at doi:10.5281/zenodo.16631516. The analysis code also can be accessed on GitHub at github.com/imanam1371/maxm-20240516.

References

Aksu, H., Taflan, G.Y., Yaldiz, S.G., Akgül, M.A., 2023. Evaluation of IMERG for GPM satellite-based precipitation products for extreme precipitation indices over Turkiye. Atmos. Res. 291, 106826.

Amjad, M., Yilmaz, M.T., Yucel, I., Yilmaz, K.K., 2020. Performance evaluation of satellite-and model-based precipitation products over varying climate and complex topography. J. Hydrol. 584, 124707.

Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D.K., Knapp, K.R., Cecil, L.D., Nelson, B.R., Prat, O.P., 2015. PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. Bull. Am. Meteorol. Soc. 96 (1), 69–83.

Auer, I., Böhm, R., Jurkovic, A., Lipa, W., Orlik, A., Potzmann, R., Schöner, W., Ungersböck, M., Matulla, C., Briffa, K., et al., 2007. HISTALP—historical instrumental climatological surface time series of the Greater Alpine Region. Int. J. Clim.: A J. R. Meteorol. Soc. 27 (1), 17–46.

Beck, H.E., Van Dijk, A.I., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B., De Roo, A., 2017. MSWEP: 3-hourly 0.25 global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. Hydrol. Earth Syst. Sci. 21 (1), 589–615.

Beck, H.E., Wood, E.F., Pan, M., Fisher, C.K., Miralles, D.G., Van Dijk, A.I., McVicar, T.R., Adler, R.F., 2019. MSWEP V2 global 3-hourly 0.1 precipitation: methodology and quantitative assessment. Bull. Am. Meteorol. Soc. 100 (3), 473–500.

Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., Ziese, M., 2013. A description of the global land-surface precipitation data products of the global precipitation climatology centre with sample applications including centennial (trend) analysis from 1901–present. Earth Syst. Sci. Data 5 (1), 71–99.

Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. J. Mach. Learn. Res. 13 (1), 281–305.

Bernini, L., Lagasio, M., Milelli, M., Oberto, E., Parodi, A., Hachinger, S., Kranzlmüller, D., Tartaglione, N., 2025. Convection-permitting dynamical downscaling of ERA5 for Europe and the Mediterranean basin. Q. J. R. Meteorol. Soc. e5014.

Bhuiyan, M.A.E., Yang, F., Biswas, N.K., Rahat, S.H., Neelam, T.J., 2020. Machine learning-based error modeling to improve GPM IMERG precipitation product over the Brahmaputra river basin. Forecasting 2 (3), 248–266.

Bongiovanni, G., Matiu, M., Crespi, A., Napoli, A., Majone, B., Zardi, D., 2024a. EEAR-Clim: A High-Density Observational Dataset of Daily Precipitation and Air Temperature for the Extended European Alpine Region. Zenodo, http://dx.doi.org/10.5281/zenodo.10951610.

Bongiovanni, G., Matiu, M., Crespi, A., Napoli, A., Majone, B., Zardi, D., 2024b. EEAR-clim: A high density observational dataset of daily precipitation and air temperature for the extended European Alpine region. Earth Syst. Sci. Data Discuss. 2024, 1–32.

Breugem, A., Wesseling, J., Oostindie, K., Ritsema, C., 2020. Meteorological aspects of heavy precipitation in relation to floods–an overview. Earth-Sci. Rev. 204, 103171.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., Yuan, J., Cortes, D., 2025. xgboost: Extreme gradient boosting. URL: https://github.com/dmlc/xgboost. R package version 3.1.0.0.

Chen, C., He, M., Chen, Q., Zhang, J., Li, Z., Wang, Z., Duan, Z., 2022. Triple collocation-based error estimation and data fusion of global gridded precipitation products over the Yangtze River basin. J. Hydrol. 605, 127307.

Dhinakaran, S., Crespi, A., Jacob, A., Pebesma, E., 2024. Enhancing seasonal climate forecasting for the Alpine region through machine learning statistical downscaling. In: IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 1683–1688.

Du, Y., Xie, Z., Miao, Q., 2020. Spatial scales of heavy Meiyu precipitation events in eastern China and associated atmospheric processes. Geophys. Res. Lett. 47 (11), e2020GL087086.

Flaounas, E., Drobinski, P., Vrac, M., Bastin, S., Lebeaupin-Brossier, C., Stéfanon, M., Borga, M., Calvet, J.-C., 2013. Precipitation and temperature space–time variability and extremes in the Mediterranean region: evaluation of dynamical and statistical downscaling methods. Clim. Dyn. 40, 2687–2705.

Frei, C., Christensen, J.H., Déqué, M., Jacob, D., Jones, R.G., Vidale, P.L., 2003. Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps. J. Geophys. Res.: Atmospheres 108 (D3).

Freitas, E.d.S., Coelho, V.H.R., Bertrand, G.F., Lemos, F.C., Almeida, C.d.N., 2024. IMERG BraMaL: An improved gridded monthly rainfall product for Brazil based on satellite-based IMERG estimates and machine learning techniques. Int. J. Climatol. 44 (11), 3976–3997.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., et al., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Sci. Data 2 (1), 1–21.

Gensini, V.A., Converse, C., Ashley, W.S., Taszarek, M., Machine learning classification of significant tornadoes and hail in the US.

Giorgi, F., Torma, C., Coppola, E., Ban, N., Schär, C., Somot, S., 2016. Enhanced summer convective rainfall at Alpine high elevations in response to climate warming. Nat. Geosci. 9 (8), 584–589.

González-Jardines, P.M., Sierra-Lorenzo, M., Ferrer-Hernández, A.L., Bezanilla-Morlot, A., 2023. Evaluation of candidate predictors for seasonal precipitation forecasting. Atmospheric Clim. Sci. 13 (4), 539–564.

Grazzini, F., Fragkoulidis, G., Teubler, F., Wirth, V., Craig, G.C., 2021. Extreme precipitation events over northern Italy. Part II: Dynamical precursors. Q. J. R. Meteorol. Soc. 147 (735), 1237–1257.

Guo, D., Wang, H., Zhang, X., Liu, G., 2019. Evaluation and analysis of grid precipitation fusion products in Jinsha river basin based on China meteorological assimilation datasets for the SWAT model. Water 11 (2), 253.

Harris, I., Jones, P.D., Osborn, T.J., Lister, D.H., 2014. Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 dataset. Int. J. Climatol. 34 (3), 623–642.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J.-N., 2023. ERA5 hourly data on single levels from 1940 to present. http://dx.doi.org/10.24381/cds.adbb2d47, Copernicus Climate Change Service (C3S) Climate Data Store (CDS). URL: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels.

Huffman, G., 2019. IMERG V06 quality index. NASA Global Precipitation Measurement (GPM) Integrated Multi-Satellite Retrievals For GPM (IMERG) Version 5.

Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K.-L., Joyce, R.J., Kidd, C., Nelkin, E.J., Sorooshian, S., Stocker, E.F., Tan, J., et al., 2020. Integrated multi-satellite retrievals for the global precipitation measurement (GPM) mission (IMERG). Satell. Precip. Meas.: Vol. 1 343–353.

Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Tan, J., 2015. Integrated multi-satellite retrievals for GPM (IMERG) technical documentation. Nasa/Gsfc Code 612 (47), 2019.

Isotta, F.A., Frei, C., Weilguni, V., Percec Tadic, M., Lassegues, P., Rudolf, B., Pavan, V., Cacciamani, C., Antolini, G., Ratto, S.M., et al., 2014. The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. Int. J. Climatol. 34 (5), 1657–1675.

Khan, R.S., Bhuiyan, M.A.E., 2021. Artificial intelligence-based techniques for rainfall estimation integrating multisource precipitation datasets. Atmosphere 12 (10), 1239.

Lanza, L.G., Stagi, L., 2009. High resolution performance of catching type rain gauges from the laboratory phase of the WMO field intercomparison of rain intensity gauges. Atmos. Res. 94 (4), 555–563.

Lanza, L.G., Vuerich, E., 2009. The WMO field intercomparison of rain intensity gauges. Atmos. Res. 94 (4), 534–543.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30.

Lyu, Y., Yong, B., 2024. A novel double machine learning strategy for producing high-precision multi-source merging precipitation estimates over the Tibetan Plateau. Water Resour. Res. 60 (4), e2023WR035643.

MacFerrin, M., Amante, C., Carignan, K., Love, M., Lim, E., 2024. The earth topography 2022 (ETOPO 2022) global dem dataset. Earth Syst. Sci. Data Discuss. 2024, 1–24.

Markonis, Y., Papalexiou, S., Martinkova, M., Hanel, M., 2019. Assessment of water cycle intensification over land using a multisource global gridded precipitation dataset. J. Geophys. Res.: Atmospheres 124 (21), 11175–11187.

Morgenstern, D., Stucke, I., Simon, T., Mayr, G.J., Zeileis, A., 2022. Differentiating lightning in winter and summer with characteristics of the wind field and mass field. Weather. Clim. Dyn. 3 (1), 361–375.

Napoli, A., Crespi, A., Ragone, F., Maugeri, M., Pasquero, C., 2019. Variability of orographic enhancement of precipitation in the Alpine region. Sci. Rep. 9 (1), 13352.

NASA Global Precipitation Measurement Mission, 2024. IMERG V07 release notes. URL: https://gpm.nasa.gov/resources/documents/imerg-v07-release-notes. (Last accessed on October 28, 2024).

NASA Goddard Earth Sciences Data and Information Services Center (GES DISC), 2023. GPM IMERG final precipitation L3 half hourly 0.1° × 0.1°V07.

NOAA National Centers for Environmental Information, 2022. ETOPO 2022 15 arc-second global relief model. http://dx.doi.org/10.25921/fd45-gt74, (Accessed [date]).

Olson, R.S., Moore, J.H., 2016. TPOT: A tree-based pipeline optimization tool for automating machine learning. In: Workshop on Automatic Machine Learning. PMLR, pp. 66–74.

Optuna Contributors, 2024. Optuna: A hyperparameter optimization framework. URL: https://optuna.org/. (Last accessed on October 28, 2024).

Pradhan, R.K., Markonis, Y., Godoy, M.R.V., Villalba-Pradas, A., Andreadis, K.M., Nikolopoulos, E.I., Papalexiou, S.M., Rahim, A., Tapiador, F.J., Hanel, M., 2022. Review of GPM IMERG performance: A global perspective. Remote Sens. Environ. 268, 112754.

Pruppacher, H.R., Klett, J.D., 2010. Microphysics of clouds and precipitation. Atmospheric and Oceanographic Sciences Library, vol. 18, Springer, Dordrecht, http://dx.doi.org/10.1007/978-0-306-48100-0, URL: https://link.springer.com/book/10.1007/978-0-306-48100-0.

Rozante, J.R., Vila, D.A., Barboza Chiquetto, J., Fernandes, A.d.A., Souza Alvim, D., 2018. Evaluation of TRMM/GPM blended daily products over Brazil. Remote. Sens. 10 (6), 882.

Sachindra, D., Ahmed, K., Rashid, M.M., Shahid, S., Perera, B., 2018. Statistical downscaling of precipitation using machine learning techniques. Atmos. Res. 212, 240–258.

Sahoo, A.K., Sheffield, J., Pan, M., Wood, E.F., 2015. Evaluation of the tropical rainfall measuring mission multi-satellite precipitation analysis (TMPA) for assessment of large-scale meteorological drought. Remote Sens. Environ. 159, 181–193.

Schamm, K., Ziese, M., Becker, A., Finger, P., Meyer-Christoffer, A., Schneider, U., Schröder, M., Stender, P., 2014. Global gridded precipitation over land: A description of the new GPCC first guess daily product. Earth Syst. Sci. Data 6 (1), 49–60.

Schneider, U., Ziese, M., Meyer-Christoffer, A., Finger, P., Rustemeier, E., Becker, A., 2016. The new portfolio of global precipitation data products of the global precipitation climatology centre suitable to assess and quantify the global water cycle and resources. Proc. Int. Assoc. Hydrol. Sci. 374, 29–34.

Sharifi, E., Eitzinger, J., Dorigo, W., 2019. Performance of the state-of-the-art gridded precipitation products over mountainous terrain: A regional study over Austria. Remote. Sens. 11 (17), 2018.

Tang, G., Clark, M.P., Papalexiou, S.M., Ma, Z., Hong, Y., 2020. Have satellite precipitation products improved over last two decades? A comprehensive comparison of GPM IMERG with nine satellite and reanalysis datasets. Remote Sens. Environ. 240, 111697.

Tartaglione, N., Parodi, A., Bernini, L., Hachinger, S., Kranzlmüller, D., 2024a. CHAPTER: 3 × 3 km meteorological data 1981–2022 for Europe: 2D extracted fields. http://dx.doi.org/10.25927/0ppk7-znk14, Data set. Leibniz Supercomputing Centre (LRZ), Garching b.M., Germany.

Tartaglione, N., Parodi, A., Bernini, L., Hachinger, S., Kranzlmüller, D., 2024b. Chapter: 3 × 3 km meteorological data 1981–2022 for Europe: 2D extracted fields. http://dx.doi.org/10.25927/0ppk7-znk14, URL: https://rdm.lab.lrz.de/records/0ppk7-znk14.

Toride, K., Iseri, Y., Duren, A.M., England, J.F., Kavvas, M.L., 2019. Evaluation of physical parameterizations for atmospheric river induced precipitation and application to long-term reconstruction based on three reanalysis datasets in Western Oregon. Sci. Total Environ. 658, 570–581.

Van Dijk, A.I., Beck, H.E., Crosbie, R.S., De Jeu, R.A., Liu, Y.Y., Podger, G.M., Timbal, B., Viney, N.R., 2013. The millennium drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. Water Resour. Res. 49 (2), 1040–1057.

Viviroli, D., Archer, D.R., Buytaert, W., Fowler, H.J., Greenwood, G.B., Hamlet, A.F., Huang, Y., Koboltschnig, G., Litaor, M., López-Moreno, J.I., et al., 2011. Climate change and mountain water resources: overview and recommendations for research, management and policy. Hydrol. Earth Syst. Sci. 15 (2), 471–504.

Viviroli, D., Kummu, M., Meybeck, M., Kallio, M., Wada, Y., 2020. Increasing dependence of lowland populations on mountain water resources. Nat. Sustain. 3 (11), 917–928.

Wang, H., Liang, Q., Hancock, J.T., Khoshgoftaar, T.M., 2024. Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods. J. Big Data 11 (1), 44.

Wehbe, Y., Temimi, M., Adler, R.F., 2020. Enhancing precipitation estimates through the fusion of weather radar, satellite retrievals, and surface parameters. Remote. Sens. 12 (8), 1342.

Xin, Y., Yang, Y., Chen, X., Yue, X., Liu, Y., Yin, C., 2022. Evaluation of IMERG and ERA5 precipitation products over the Mongolian Plateau. Sci. Rep. 12 (1), 21776.

Xiong, J., Tang, G., Yang, Y., 2025. Continental evaluation of GPM IMERG V07B precipitation on a sub-daily scale. Remote Sens. Environ. 321, 114690.

Xu, L., Chen, N., Moradkhani, H., Zhang, X., Hu, C., 2020. Improving global monthly and daily precipitation estimation by fusing gauge observations, remote sensing, and reanalysis data sets. Water Resour. Res. 56 (3), e2019WR026444.

Zentralanstalt für Meteorologie und Geodynamik (ZAMG), n.d. HISTALP station-mode data (monthly/daily homogenised CSV export). ZAMG HISTALP database. URL: https://www.zamg.ac.at/histalp/dataset/station/csv.php.

Zhang, L., Li, X., Zheng, D., Zhang, K., Ma, Q., Zhao, Y., Ge, Y., 2021a. Merging multiple satellite-based precipitation products and gauge observations using a novel double machine learning approach. J. Hydrol. 594, 125969.

Zhang, Z., Wang, D., Qiu, J., Zhu, J., Wang, T., 2021b. Machine learning approaches for improving near-real-time IMERG rainfall estimates by integrating cloud properties from NOAA CDR PATMOS-x. J. Hydrometeorol. 22 (10), 2767–2781.

Zhao, B., Hudak, D., Rodriguez, P., Mekis, E., Brunet, D., Eckert, E., Melo, S., 2023. Assessment of IMERG V06 satellite precipitation products in the Canadian great lakes region. J. Hydrometeorol. 24 (6), 1017–1037.