hydrology

MDPI

Article

Spatial Downscaling of the CHIRPS Rainfall Product Using Machine Learning Methods: The Catamayo–Chira Transboundary Basin (Ecuador-Peru) Case

Jessica K. Gaona [1,2], Luis-Felipe Duque [1,2,*], Raúl F. Vázquez [3,4] and Candy L. Ocaña [5]

[1] Carrera de Ingeniería Ambiental, Universidad Nacional de Loja (UNL), Avenida Pío Jaramillo Alvarado, La Argelia, Loja 1101608, Ecuador; jessica.gaona@unl.edu.ec
[2] Grupo de Investigación en Hidroinformática, Universidad Nacional de Loja (UNL), Avenida Pío Jaramillo Alvarado, La Argelia, Loja 1101608, Ecuador
[3] Laboratorio de Ecología Acuática (LEA), Facultad de Ciencias Químicas, Universidad de Cuenca, Víctor Manuel Albornoz S/N y Av. de los Cerezos, Cuenca 010215, Ecuador; raulfvazquezz@yahoo.co.uk
[4] Departamento de Ingeniería Civil, Facultad de Ingeniería, Universidad de Cuenca, Av. 12 de Abril S/N, Cuenca 010203, Ecuador
[5] Instituto Binacional de Investigación para el Desarrollo Sostenible de la Cuenca del Marañón de Perú y Ecuador, Universidad Nacional de Jaén (UNJ), Carretera Jaén-San Ignacio KM 21, Jaén 06801, Peru; candy.ocana@unj.edu.pe
* Correspondence: luis.duque@unl.edu.ec

Abstract

Precipitation modeling is vital for water resource management in basins with limited gauged data. In this study, the 5 km Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) product was downscaled to 1 km at the annual and mean monthly scales for the Catamayo–Chira catchment, a key water source for Ecuador and Peru. Single-variable and multivariable machine learning (ML) methods were applied to data from 10 gauged stations from 2001 to 2023. Predictors included longitude (Long), latitude (Lat), altitude, Normalized Difference Vegetation Index, and Land Surface Temperature. Performance metrics were utilized to assess the methods. The results demonstrated a notable improvement after downscaling compared to the original CHIRPS estimates. The most effective single-variable methods were simple linear regression (LR) for Long and Lat, and non-linear ML methods such as support vector machine with linear kernel (SVM-lin) and with radial basis function kernel (SVM-rbf), and artificial neural networks (ANN), employing all predictors. Surprisingly, single-variable linear methods yield better results than multivariable non-linear ones. These models provided acceptable fits to annual and mean monthly observations, and their performance tended to be better during the drier months. Downscaled annual precipitation distributions successfully captured differences between "El Niño" and non-"El Niño" years. The current study could be replicated in basins with limited gauging data, thereby enhancing water resource management.

Keywords: precipitation; CHIRPS; machine learning; linear regression; support vector machine; random forest; artificial neural network; performance metrics; model validation

Check for updates

Academic Editors: Ezio Todini and Tammo Steenhuis

Received: 30 January 2026
Revised: 4 March 2026
Accepted: 6 March 2026
Published: 10 March 2026
Corrected: 9 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Freshwater, circulating through the water cycle, is an essential resource for human survival and maintaining ecosystem processes [1,2]. In this cycle, precipitation is a key meteorological variable and a primary driver for understanding climate variability and

change [3,4]. Accurate knowledge of its spatial and temporal distribution is therefore crucial for hydrometeorological studies, including hydrological modeling, water resource management, and assessing the impacts of climate change on surface water availability, as well as flood and drought risk assessment [5–7].

Traditionally, precipitation has been measured using rain gauges, which are considered the most reliable measurement technique [8]. However, the global distribution of rainfall stations is markedly non-uniform, with many regions, particularly in developing countries, constrained by sparse and limited monitoring networks [9,10]. Such data scarcity represents a significant challenge for effective water resource planning and management [10].

In this context, remote sensing provides a valuable alternative source of precipitation information. In recent years, significant progress has been made in the development and availability of satellite precipitation products (SPP). These products can complement or, when corrected rigorously, replace rain gauge data, thereby providing broader rainfall coverage for hydrometeorological studies. Thus, a wide array of SPPs have been created so far, including Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (IMERG-GPM) mission, the Global Precipitation Measurement (GPM) mission, Precipitation Estimation from Remote Sensing Information using Artificial Neural Networks (PERSIANN-CDR), Global Satellite Mapping of Precipitation (GSMaP), the Multi-Source Weighted-Ensemble Precipitation (MSWEP), and the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) [5]. The main limitation of SPPs is their low spatial resolution, which often fails to meet the accuracy requirements for hydrometeorological applications [11,12]. To address this issue, spatial downscaling techniques have been applied to refine coarse-resolution satellite data, generating high-resolution SPPs that support regional and local-scale studies [13–15].

The literature offers two main approaches for the spatial downscaling of precipitation: dynamical and statistical downscaling [16]. Dynamical downscaling relies on regional climate models that simulate the complex physical interactions among the atmosphere, ocean, and land surface. Although physically robust, this approach requires substantial computational resources and large volumes of data, which limit its practical applicability [16–19]. In contrast, statistical downscaling establishes empirical relationships between the target variable (the coarse-resolution SSP) and predictor variables at a finer spatial scale, offering a highly efficient approach that has been widely used for satellite-based precipitation downscaling in recent studies [8,20–23].

Standard predictor variables in statistical downscaling include the Normalized Difference Vegetation Index (NDVI) [24,25], particularly in arid and semi-arid regions where vegetation growth is mainly dependent on rainfall. In such areas, Chen et al. [26] also introduced the use of Land Surface Temperature (LST) as a predictor variable to enhance the downscaling of the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) from a coarser (approximately 25 km) to a finer (approximately 1 km) spatial resolution. Altitude is another crucial factor, as it improves the representation of precipitation patterns in areas with complex terrain [27]. Finally, Fang et al. [28] emphasized that precipitation is also influenced by latitude and longitude.

A wide range of statistical downscaling models have been applied using these predictors, ranging from linear regressions (LR) to non-linear (NL) algorithms. Early studies mainly employed parametric approaches, such as Univariate Regression (UR) and Multiple Linear Regression (MLR) models, which later evolved into more sophisticated, NL algorithms, including Support Vector Machines (SVM), Random Forests (RF), and Artificial Neural Networks (ANN). These methods are particularly effective at capturing the complex, non-linear interactions between precipitation and its predictors [5,29–33]. Several SPPs have been downscaled worldwide using such approaches within Machine Learning (ML)

frameworks. Examples include Chen et al. [20], who applied UR, MLR, and NL models to refine GPM IMERG data in typical arid to semi-arid areas of China. Sharifi et al. [8] employed MLR and ANN-based models, along with spline interpolation, to downscale the GPM product over an Austrian basin. Retalis et al. [34] used an ANN-based strategy to downscale the CHIRPS product over the island of Cyprus. Kofidou et al. [5] provide an extensive review of studies addressing the downscaling of SPPs.

In general, studies show that NL models often outperform LR methods in downscaling tasks, although there are some exceptions [8]. It is therefore important to emphasize that the performance of downscaling techniques is both region- and data-specific [6,8,20]; in some cases, LR models can outperform NL approaches. Thus, it is very relevant to examine the performance of different data-learning approaches/methods when spatially downscaling SPPs under local conditions.

In particular, the current study site, the transboundary Catamayo–Chira basin (Ecuador–Peru), has not previously been the focus of any precipitation downscaling studies. Applying established downscaling techniques (both linear and non-linear machine learning methods) to this data-scarce Andean catchment, therefore, represents a novel contribution, generating fine-scale (1 km) CHIRPS precipitation estimates where none existed before. These high-resolution estimates are especially valuable given that the basin supplies surface water to two major irrigation projects, the Zapotillo project in Ecuador and the Poechos reservoir in Peru, which are vital to the economies of both countries [7]. The downscaled results, therefore, provide essential inputs for local-scale hydrological assessments and support improved water resource management in this binational context [7]. Therefore, the main research question is: Is it feasible to identify a data-learning method that can generate acceptably precise precipitation datasets for the study site by downscaling an SPP?

This study aimed not only to support better water resource management in the binational study basin but also to contribute to the discussion on the performance of ML methods for downscaling SPPs. These methods (single-variable and multivariable models) and the modeling protocol applied herein could easily be used/replicated in other latitudes for similar assessments.

2. Materials and Methods

2.1. The Study Site

The study site, the Catamayo–Chira binational basin, is a Pacific–Andean system located in southern Ecuador and northern Peru, spanning the coordinates 3°30′ to 5°08′ south and 79°10′ to 81°11′ west, covering an area of 17,199 km^2 (Figure 1). The elevation ranges from 0 to almost 4000 m above sea level (a.s.l.). The orogenesis of the Andes Mountain range in the study area has created a complex topography, generating different altitudinal floors and microclimates [35]. The Catamayo–Chira basin originates in humid, high-elevation areas of the Andean Cordillera. It descends into semi-humid and semi-arid zones in its lower reaches. However, isolated arid regions can be found in the eastern part of the catchment, such as the Catamayo Valley. This spatial variability is shaped by the basin's complex topography, characterized by numerous small ridges and intersecting valleys that originate from the two distinct mountain chains observed in the Ecuadorian and Peruvian Andes.

In the lower part of the transboundary Catamayo–Chira catchment (0–80 m a.s.l.), the wet season is short, occurring from January to April, with annual rainfall ranging from only 10 to 80 mm. As elevation increases toward the middle and upper sectors, the wet season becomes longer and total rainfall increases. During this period, the rainy season runs from October to May, with March having the highest precipitation [36]. Precipitation regimes are strongly affected by the El Niño–Southern Oscillation (ENSO), especially its

negative phase (La Niña). ENSO has a notable influence on rainfall in the lower elevations (<800 m a.s.l.). Under "normal" conditions, annual precipitation averages about 349 mm, whereas during strong La Niña episodes and other neutral ENSO events, precipitation averages 2527 mm and 752 mm, respectively. Above 800 m a.s.l., ENSO effects are less pronounced, with average annual totals rising from about 984 mm in "normal" years to 1646 mm in years with intense adverse ENSO conditions [37]. Temperatures range from 24 °C in the lower reaches to 7 °C in the upper reaches of the basin [38].

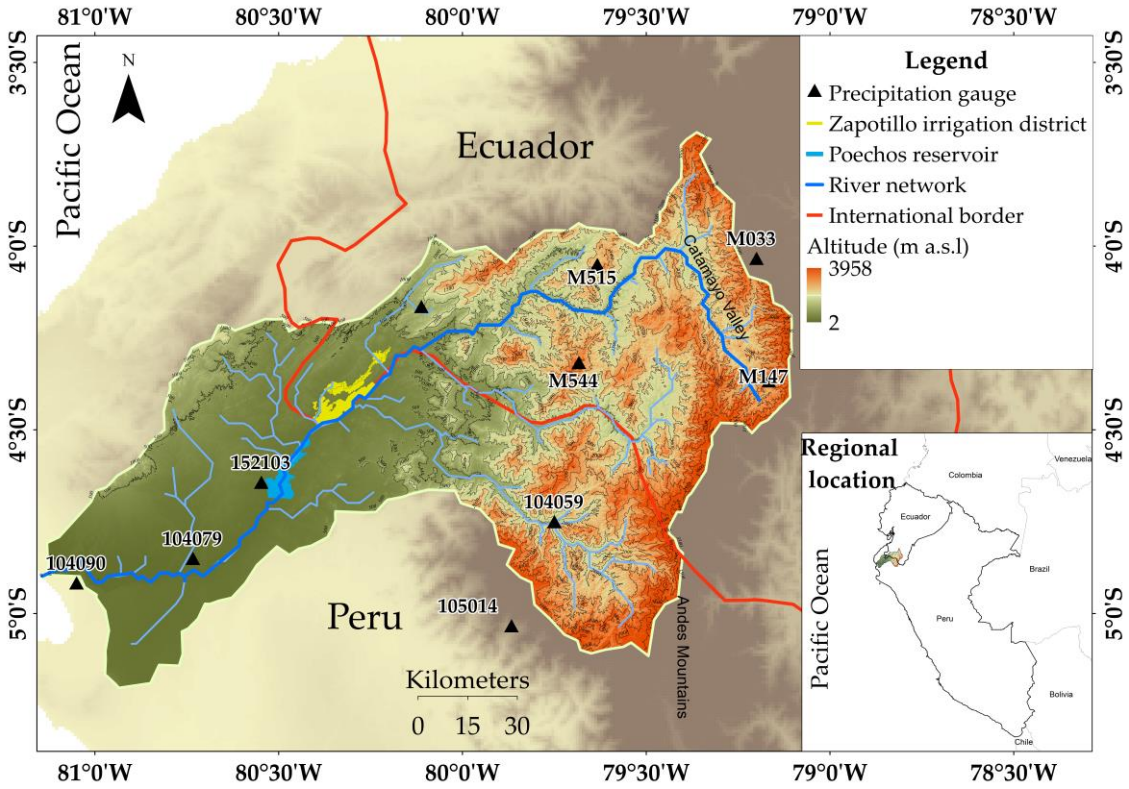

Figure 1. Regional location of the study site, the Catamayo–Chira transboundary (Ecuador–Peru) basin, and distribution of the main water courses, topography, and meteorological stations used in this research.

Within this basin, two major irrigation schemes are particularly noteworthy: Ecuador's Zapotillo irrigation project and Peru's Poechos reservoir (Figure 1). These initiatives rely on the catchment's surface water resources and, together, support the irrigation of approximately 11,500 hectares.

2.2. Available Information

2.2.1. Observed Meteorological Information

For this study, data recorded in 23 meteorological stations managed by the National Institute of Meteorology and Hydrology (INAMHI) of Ecuador and the National Meteorological and Hydrological Service (SENAMHI) of Peru were analyzed. Given the binational nature of the study basin, observations were collected using pluviometry, with different instrumentation and standards, and stored in different formats. Historically, political instability and economic constraints have led to discontinuous data monitoring. Factors such as dispersion in efforts, a lack of coordination, poor follow-up, and inconsistent data storage mean that the meteorological time series in Ecuador is incomplete, often of short duration, and heterogeneous [39].

To address these limitations, quality control measures were implemented. First, missing data was quantified; INAMHI stations were selected based on complete annual records,

while SENAMHI stations were selected using a gap-maximum threshold of 7%. Despite the World Meteorological Organization (WMO) recommendation of a 5% bound [40], a 7% threshold was applied to maintain a larger volume of observed data. Subsequently, an internal consistency analysis was performed to detect outliers (based on the interquartile range—IQR—method) and to evaluate the homogeneity of the series (using the double-mass curve), thereby ensuring that the observed variability reflects genuine climatic signals rather than artifacts caused by external factors, including changes in instrumentation or station location. In addition, local station maintenance reports and relevant technical documentation were consulted to support decisions regarding the treatment of anomalous observations.

Following this control, the observational data was reduced to 10 stations across the basin (Figure 1 and Table 1), covering 23 years from January 2001 to December 2023. To ensure consistency across the dataset, daily records from SENAMHI stations were first aggregated into monthly totals. These were then combined with the monthly total series provided by INAMHI. Subsequently, annual precipitation values were derived by summing the monthly totals for each year within the 2001–2023 period. All quality control and preprocessing procedures were implemented using task-specific routines developed in R® (version 4.5.1).

Table 1. Main properties of the study precipitation stations and their respective data availability periods.

Code	Name	Location		Altitude (m a.s.l.)	Data Availability		
		Latitude	Longitude		Period	Length (Year)	Temporal Scale
M515	Catacocha	4°3′	79°38′	1808	2001–2011	9	Monthly
M544	Colaisaca	4°19′	79°41′	2410	2001–2011	11	
M033	La Argelia	4°2′	79°12′	2160	2001–2021	21	
M0765	Sabanilla	4°10′	80°7′	733	2001–2017	11	
M147	Yangana	4°22′	79°10′	1835	2003–2011	8	
104090	La Esperanza *	4°55′	81°3′	7	2001–2023	23	Daily
152103	Lancones *	4°39′	80°33′	133	2003–2023	18	
104079	Mallares *	4°33′	80°18′	44	2001–2023	18	
105014	Santo Domingo *	5°2′	79°52′	1490	2002–2022	18	
104059	Sausal de Culucan *	4°45′	79°45′	997	2001–2022	19	

Note: * indicates that the gauge stations are in Peru. a.s.l = above the average sea level.

The methodology was validated for the studied range, except for 2020, 2022, and 2023, when the number of available stations was insufficient.

2.2.2. Satellite-Based Precipitation Data

In this study, satellite-based Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) was used. This is a precipitation dataset spanning 35 years, from 1981 to the present, covering part of the Earth's globe (50° S–50° N and all longitudes). Its spatial resolution is 0.05° (about 5 km). It is a reanalysis product, meaning its data is derived by combining satellite images and observed data to create spatially distributed time series of rainfall. It has been widely used for various water resources assessments [41–43], including hydrological modeling [43].

CHIRPS therefore offers advantages over newer datasets like IMERG (Integrated Multi-satellitE Retrievals for Global Precipitation Measurement), providing a longer time series for climate analysis, along with good spatial and temporal (daily) resolutions and global coverage. It has been selected for the current study based on these aspects, and considering

that prior worldwide evaluations obtained acceptable estimates of daily precipitation even for Andean regions [42], which suggests it could be particularly valuable in the study basin, where ground-based precipitation stations are limited, providing continuous, spatially consistent precipitation data.

Herein, daily values were considered for the period 1 January 2001–31 December 2023, downloaded from the Climate Hazards Center (CHC) https://www.chc.ucsb.edu/data/chirps (accessed on 27 April 2024), and subsequently aggregated to monthly averages and annual values, as explained later in the text. For all the predictive variables, ArcMap® and R® were used for data processing.

2.2.3. Predictor Variables

Predictor variables included the Normalized Difference Vegetation Index (NDVI), Land Surface Temperature (LST), longitude (Long), latitude (Lat), and altitude. The sources from which these data were obtained are described below.

Normalized Difference Vegetation Index (NDVI) Data

NDVI data was obtained from the MOD13A3.061 product, derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor onboard the Terra satellite of the United States (US) National Aeronautics and Space Administration (NASA), with a spatial resolution of 1 km and a temporal resolution of 16 days. This data was downloaded from the AppEEARS website https://appeears.earthdatacloud.nasa.gov/ (accessed on 27 April 2024) for the period 2001–2023. Monthly/annual NDVI values were calculated by averaging the 16-day NDVI data for each month/year. Since NDVI values below zero generally indicate water bodies, non-vegetated surfaces, or data affected by atmospheric conditions, they were excluded [21].

Land Surface Temperature (LST) Data

LST data was obtained from the MOD11A2.061 product, which is also part of the MODIS satellite mission. This product provides daytime and nighttime LST data with a temporal and spatial resolution of 8 days and 1 km, respectively [44], and was downloaded from the AppEEARS website https://appeears.earthdatacloud.nasa.gov/ (accessed on 27 April 2024) for the period 2001–2023. These values were first aggregated to the monthly level by averaging the data within each month, and subsequently to the annual and mean (total) monthly levels.

Altitude, Longitude (Long), and Latitude (Lat) Data

The altitude data used in this study was obtained from the global Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) with a spatial resolution of 1 km available at https://catalog.data.gov/dataset/srtm30-global-1-km-digital-elevation-model-dem-version-11-land-surface (accessed on 27 April 2024). This DEM, produced through an international collaboration led by the NASA and the U.S. National Geospatial-Intelligence Agency (NGA), covers latitudes from 56° S to 60° N. For the empirical validation, the vertical accuracy of the SRTM DEM was assessed using geodetic control points, yielding a root mean square error (RMSE) of approximately 3.53 m [45]. Latitude and longitude data were derived from the 1 km DEM and served as a standard reference framework for integrating all spatial variables in subsequent analyses.

2.3. General Spatial Scaling Analysis Framework

Statistical spatial downscaling aims to establish an empirical relationship between the variable of interest and the predictor variables at a coarse spatial resolution [46]. It is assumed that this relationship remains valid at a finer spatial resolution, allowing for the

prediction of high-resolution precipitation (downscaled variable) from predictor variables that are also available at a finer resolution [8,20,47].

In the present study, this concept was applied to spatially downscale the CHIRPS satellite product. Both single- and multivariable models were used to assess the statistical relationship between satellite-based precipitation and geospatial predictor variables. These models included simple linear regression (LR), multiple linear regression (MLR), random forest (RF), support vector machine (SVM), and artificial neural networks (ANN). The model parameters for the non-linear ML algorithms were configured using a grid search algorithm through the scikit-learn library and adjusted based on the methodological recommendations of related studies, such as [20,48].

Before training, the predictor variables used in the three non-linear models were standardized using the Standard Scaler method. For SVM and ANN, scaling is essential because these algorithms are susceptible to the relative scales of the predictors; standardization improves training stability, accelerates convergence, and reduces errors [49,50]. In contrast, RF is primarily invariant to transformations of the predictors, and scaling has little or no effect on its performance [50]. Nonetheless, standard scaling was applied to all models to maintain consistency in preprocessing, which benefits sensitive models without harming RF.

The estimation protocol applied at the annual and average monthly levels (Figure 2) included the following steps:

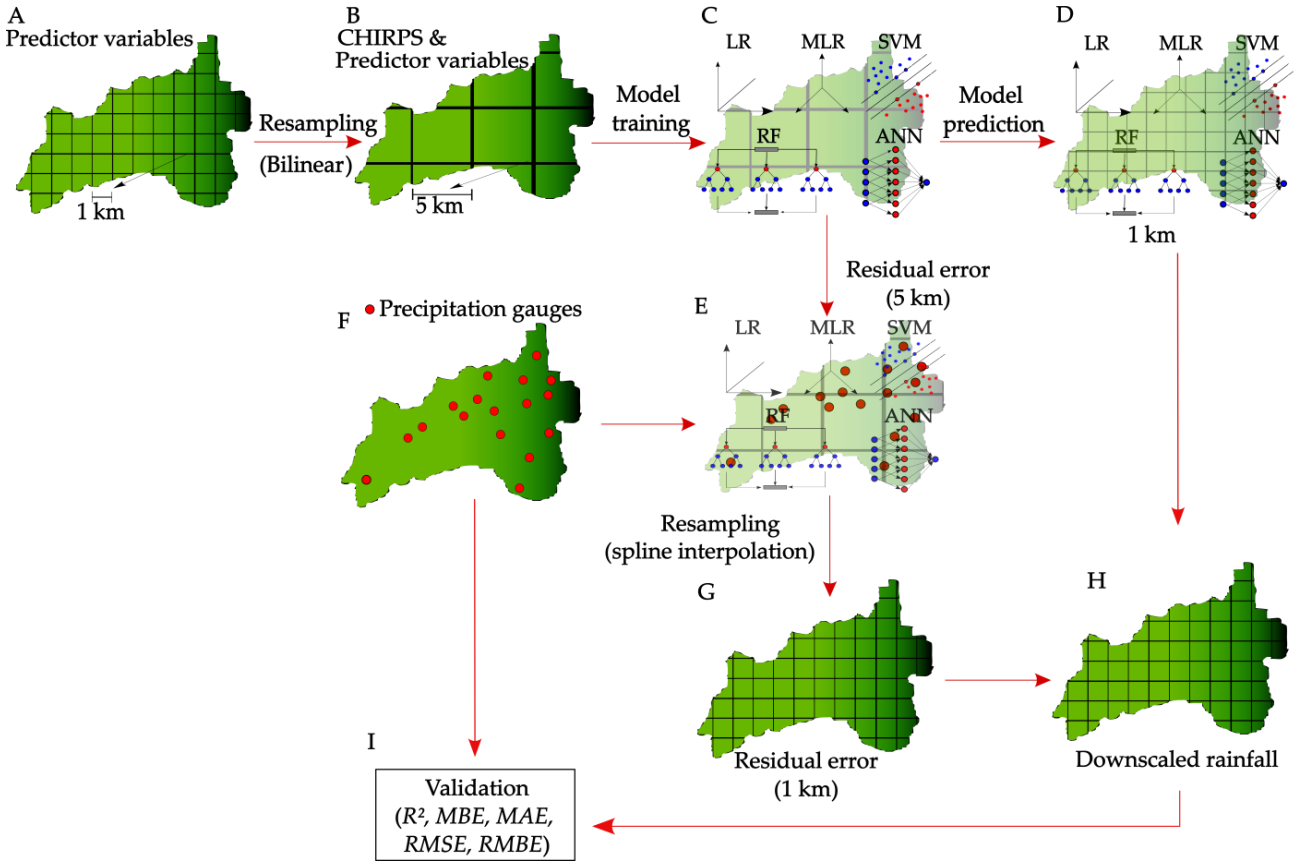

Figure 2. Schematic representation of the protocol used in this study for the spatial scaling of the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) for the Catamayo–Chira transboundary basin (Ecuador–Peru), using machine learning (ML) methods: simple (LR) and multiple linear regression (MLR), random forest (RF), support vector machine (SVM), and artificial neural networks (ANN). R^2 = Pearson's type coefficient of determination; *MBE* = mean bias error; *RMSE* = root mean square error; *RMBE* = relative mean bias error. The protocol is explained later in the text.

1. First, the spatial predictor variables (altitude, longitude, latitude, NDVI, and LST) were upscaled from 0.01° (about 1 km) to 0.05° (about 5 km), which is the native resolution of CHIRPS, using the pixel resampling method and bilinear interpolation [51–53] (Figure 2A,B).
2. Then, the linear and non-linear machine learning models were trained based on the CHIRPS data and the five spatial predictor variables at a 5 km spatial resolution (Figure 2C).
3. Next, the trained ML models were used to predict precipitation at a finer spatial resolution using the estimators at their native 1 km resolution (Figure 2D).
4. Residual errors were calculated as the difference between the predicted precipitation at 5 km (Figure 2E) and the data recorded at weather stations (Figure 2F and Table 1) located within a given pixel. These errors were then downscaled (resampled) to 1 km using spline interpolation, yielding a spatial distribution of residual errors for each pixel of the 1 km map (Figure 2G).
5. The final downscaled product was obtained by summing up the predicted spatial distribution of rainfall at 1 km resolution and the residual errors with the exact resolution (Figure 2H).
6. Finally, the 1 km downscaled precipitation distribution was validated using the performance metrics considered in this study (Figure 2I), based on the error residuals between this 1 km downscaled precipitation and the respective observed precipitation (ground-based rainfall).

This methodology has been applied worldwide using various satellite precipitation products, SPPs [5,8,20,21,25,46,54,55], with the core objective of producing high-resolution rainfall fields that accurately reflect observed conditions in years with available data. To the best of our knowledge, however, it has not previously been implemented in the Catamayo–Chira basin using CHIRPS products. The details of each data learning model used in the present study, along with the validation process, are presented below.

2.4. Linear Machine Learning Methods

2.4.1. Linear Regression (LR)

Simple linear regression [56] LR (Figure 2C) builds the statistical relationship between precipitation and a single geospatial predictor variable (altitude, longitude, latitude, NDVI, or LST). Its expression is

$$P = \beta_0 + \beta_1 X \quad (1)$$

where [57] P is the satellite-derived precipitation [L], β_0 is the intercept of the regression [L], β_1 is the slope of the predictor variable, and X is the predictor variable used.

2.4.2. Multiple Linear Regression (MLR)

Multiple linear regression (MLR, Figure 2C) models the relationship between precipitation and multiple predictor variables. Before model construction, we performed a multicollinearity analysis of all five predictors. Due to the strong negative correlation between altitude and land surface temperature LST, it was excluded from the MLR model to avoid redundancy and ensure statistical interpretability. The expression is as follows:

$$P = \beta_0 + \sum_{i=1}^{q} \beta_i X_i \quad (2)$$

where β_i is the coefficient of the i-th predictor variable X_i, and q is the number of predictor variables [57] (q = 4 in the current case).

The simple linear regression models LR were implemented using one predictor at a time, not for direct performance comparison with multivariate methods, but to assess the individual predictive capacity of each variable. While these models use fewer inputs than multivariate approaches, they offer valuable insights for gauging the utility of specific predictors. These results help interpret the role of specific geographic and environmental factors in explaining spatial precipitation patterns.

2.5. Non-Linear Machine Learning Methods

2.5.1. Random Forest (RF)

Random forest, RF (Figure 2C), is a non-parametric ML algorithm that combines multiple decision trees and is used in both classification and regression problems [58]. It is based on the bagging (bootstrap aggregation) technique, which generates N bootstrap training samples from the values of the variable under study [59]. Each bootstrap sample is used to construct a regression tree, incorporating randomization to reduce overfitting. At each node of the tree, a random subset of predictor variables is selected to prevent a single variable from dominating the decision at each split [59]. The result for the precipitation estimates corresponds to the average of the predictions from all trees, as shown in the following expression:

$$P = \frac{1}{N}\sum_{i=1}^{N} P_i \tag{3}$$

where N is the number of trees (samples) and P_i is the precipitation value predicted by each tree i. A detailed description of this algorithm can be found in [60].

2.5.2. Support Vector Machine (SVM)

Support vector machine (SVM; Figure 2C) is an ML algorithm designed for both classification and regression problems. The SVM is a kernel-based algorithm that uses structural risk minimization and statistical learning methods to achieve good generalization by minimizing generalization errors rather than training errors. SVM uses a transfer function to non-linearly map input vectors into a high-dimensional feature space, which helps reduce optimization complexity. The inspiration behind the SVR technique is the definition of a regression function approximation based on a set of support vectors derived from a training dataset. According to some studies [49], the SVM function is given by

$$f(x) = \sum_{i=1}^{N} (\propto_i - \propto_i^*) K(x, z) + b_i \tag{4}$$

where $(\propto_i - \propto_i^*)$ is the Lagrange multiplier, $K(x, z)$ is the kernel function inside the multiplier, and b_i is the bias. The kernel functions tested in this study were: the radial basis function (SVM-rbf), linear (SVM-lin), polynomial (SVM-poly), and sigmoid (SVM-sig). A more detailed explanation of this method can be found in [61].

2.5.3. Artificial Neural Networks (ANNs)

Artificial neural networks, ANNs (Figure 2C), are inspired by biological neural networks that simulate human learning [62]. ANNs consist of neurons (nodes) and weighted connections, and include input, hidden, and output layers [63]. Training involves this three-layer structure and uses a backpropagation algorithm that adjusts weights until the error is minimized [49,64]. The final model is selected based on the lowest mean squared error (*MSE*), which is the mean of the squared discrepancies between observations and model predictions (in this case, downscaled precipitation). Implementing the ANN model involves randomly splitting the data into training, validation, and test sets while exploring

various architectures. As explained by Zakaria et al. [65] the following equation describes the ANN mathematical model:

$$y_i = \mathrm{f}\sum_{\mathrm{i}=1}^{N} w_{ij}x_i + b_j, \tag{5}$$

where y_i is the output variable, N is the number of neurons, w_{ij} is the weight connecting the j-th neuron and the i-th neuron, x_i is the input vector, b_j is the bias of the j-th neuron, and f is the activation function.

In addition, different epochs, training rates, and batch sizes were tested to identify the optimal ANN architecture for the present study.

2.6. Performance Metrics

The evaluation process of the applied downscaling protocol was based on the point-to-pixel approach [41,66], which involves comparing the observed values (O) and the downscaled CHIRPS values (P) within a given pixel. The metrics used were Pearson's type coefficient of determination (R^2) [-], mean bias error (MBE) [L], relative mean bias error ($RMBE$) [-], mean absolute error (MAE) [L] and root mean square error ($RMSE$) [L], which were used in similar previous assessments [20,22,41,43,67,68] and are calculated as follows:

$$R^2 = \left(\frac{\sum_{i=1}^{n}\left[\left(P_i - \overline{P}\right)\left(O_i - \overline{O}\right)\right]}{\sqrt{\sum_{i=1}^{n}\left(P_i - \overline{P}\right)^2}\sqrt{\sum_{i=1}^{n}\left(O_i - \overline{O}\right)^2}} \right)^2 \tag{6}$$

$$MBE = \frac{\sum_{i=1}^{n}\left(P_i - O_i\right)}{n} \tag{7}$$

$$MAE = \frac{\sum_{i=1}^{n}\left|P_i - O_i\right|}{n} \tag{8}$$

$$RMBE = \frac{MBE}{\overline{O}} \times 100 \tag{9}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(P_i - O_i\right)^2}{n}} \tag{10}$$

where $\overline{P}$ and $\overline{O}$ are respectively the average values of downscaled and observed precipitation [L], and n is the number of pairs of observed and downscaled precipitation values.

These statistics are based on residual calculations. Commonly [69–71], an i-th residual ($\mathrm{res_i}$) is defined as the difference between O_i and P_i. Nevertheless, in this study, following Walther and Moore [72], the opposite-sign relationship, i.e., $\mathrm{res_i} = (P_i - O_i)$, was adopted so that a positive residual indicates overestimation of observed rainfall. Correspondingly, the mathematical descriptions of the remaining performance metrics used in this study were adjusted to reflect this residual form.

Hereafter, the Pearson R^2, a good measure of data correlation, is not appropriate [73] to evaluate the performance and quality of precipitation products relative to the observed precipitation (ground-based precipitation) because it is oversensitive to peak values and insensitive to additive and proportional differences among observed and estimated (i.e., downscaled) data [69,74]. Hence, R^2 was mainly used in this study to characterize the correlation between the observed and downscaled precipitation data and to enable a direct comparison with previously published similar studies [41,43,68,73]. Alternatives to this statistic are [75] the dimensionless efficiency coefficient (EF_2), also known as the Nash–Sutcliffe coefficient [69], or the modified EF_2 statistic [76] or the dimensionless Kling–Gupta

efficiency (KGE), which have been used with success in prior hydrological and ML spatial studies [41,75,77].

The *MBE*, also termed *BIAS* in the literature, is the average residual and is also commonly used in similar assessments [41,43]. However, it has the implicit pitfall that residuals of the opposite sign and similar magnitude can cancel each other out, leading to a low *MBE* value and producing, as such, a false impression of model accuracy [75]. To overcome this deficiency, the mean absolute error (*MAE*), that is, the average of the absolute values of residuals, is used instead [41,73,75], alongside other variants of the *MBE* statistic [69,73]. Nevertheless, *MBE* was also used in this study for a direct comparison with similar previously published studies. *RMBE*, defined as the ratio of *MBE* to the mean of the observations in the analysis period, was used in this study and expressed as a percentage by multiplying the ratio by 100.

The *RMSE* is generally a good statistic regarding performance metrics, as it is directly related to the EF_2 statistic [70], which measures the average combined systematic and random errors between the observed and estimated data. It has also been used in similar studies [43,67]. For a more comprehensive discussion of these performance metrics, see, for instance, the work of Vázquez et al. [75], Moriasi et al. [74] or Walther and Moore [72].

For inspecting the spatial distribution of the residuals for every pixel of interest (i.e., the location of a precipitation-gauging station), the dimensionless relative error ($RE_i = (O_i - P_i)/O_i$) was considered. Observe that both the numerator and denominator of this expression are defined relative to the i-th observation. This has been successfully used in prior studies that focused on pixel-by-pixel analyses, although sometimes the absolute value of the preceding expression was used [75,78] when the mathematical signs of the errors were not relevant.

The performance metrics (8–12) were first applied during the model-building/training stage (Figure 2C), by comparing each model's 5 km predictions with the original CHIRPS values. This step, referred to as "model evaluation", assesses the robustness of each method. The same metrics were then applied after downscaling using the 1 km estimates and the gauge-based observations (Figure 2I). Lastly, to evaluate the effectiveness of the downscaling approaches at 1 km, the metrics were also computed using the raw CHIRPS data and the gauge observations; this step was termed "downscaling validation".

It is important to note that the training and validation results presented in this study are not directly comparable as they are based on distinct methodological stages with different inputs and objectives. The training stage involves predicting CHIRPS precipitation from geospatial predictors and comparing these estimates to the observed station data, without any post-processing or correction. In contrast, the validation stage includes an additional residual-correction step, in which downscaled CHIRPS fields are adjusted using gauge-derived errors interpolated across space. This correction process improves the alignment with observations and is fundamental to the methodology's goal of reconstructing high-resolution rainfall fields for years for which data are available. As a result, differences in performance across the two stages reflect not inconsistency but the effects of the residual adjustment and structural differences in the workflow.

2.7. Assessment of the Spatial Distribution of Precipitation for "El Niño" Years (ENYs)

As mentioned in the introduction, the precipitation regime of the study zone is strongly influenced by "El Niño" events. In this context, the spatial distribution of annual downscaled CHIRPS precipitation for "El Niño" years (ENYs) included in the study period was compared with the corresponding distribution in the original CHIRPS data. This allowed for an assessment of whether the estimated precipitation spatial distributions for ENYs showed notably different patterns to those observed in regular years. Indeed, it is expected

that "El Niño" events significantly disrupt precipitation distributions, often reversing or profoundly altering the established rainfall patterns observed in normal (i.e., regular) years. Specifically, during ENYs, lower elevations, such as coastal regions, tend to receive more precipitation due to anomalous sea surface warming [79]. In comparison, higher elevations or mountainous areas may experience reduced rainfall, which is directly linked to shifts in atmospheric temperature gradients [80,81].

The ENYs of 2017 and 2023 were included in the study. Notably, 2023 was a year marked by disrupted precipitation patterns along the coasts of Ecuador and Peru [79]. Unfortunately, data availability for this year was limited, making it impossible to include an analysis of this dataset. Subsequently, only ENYs from 2017 were included in this brief analysis.

3. Results and Discussion

As stated earlier in Section 2, a validation of the performance of the downscaling methods is relevant and should be based on a combined graphical and statistical approach [82,83]. It is essential to select performance metrics that measure different features of model error [69,75,83]. Statistics measuring different aspects were selected and applied, namely, average systematic errors (*MBE* and *MAE*) and the correlation between observed and downscaled precipitation (R^2). Because *RMSE* is a function [84] of the Nash and Sutcliffe [85] coefficient of efficiency (EF_2), a typical measure of the combined average systematic and random error [69,75], it was also selected in this study to measure the combined features of model error.

Similar statistics, capturing at least the same information on model errors, have been employed in related research [20,22,41–43]. Nevertheless, such statistics are often used because they are "commonly used" [20], without a clear understanding and/or analysis of the error features that they measure. In this regard, it is common to find, in the scientific literature, the use of statistics that measure redundant information regarding model errors [41,42]. Using different performance metrics does not necessarily yield more accurate evaluations of precipitation downscaling models. Other works [46,67,73] do not measure all the error features addressed in the current research, and some even incorrectly report model performance statistics [22,46]. With this clarification in mind, the results for both the training and validation stages are presented below.

3.1. Model Training

3.1.1. Simple Linear Regression Models

As mentioned earlier (Figure 2C), single- and multivariable methods were applied to the CHIRPS precipitation dataset, considering the predictor variables—latitude (Lat), longitude (Long), altitude (Alt), Normalized Difference Vegetation Index (NDVI), and Land Surface Temperature (LST)—from 2001 to 2023 at a 5 km resolution. Table 2 summarizes the linear model fits using the evaluation statistics *MAE*, *RMSE*, and R^2. Range ("interval") and global values are displayed. It is important to clarify that the reported R^2 global values were calculated over the entire evaluation dataset, aggregating across all time steps and spatial points used in the model training. This approach ensures a robust estimate of the model's total explanatory power across the full spatiotemporal domain.

Regarding annual precipitation, all performance metrics indicate that the most simple LR method is the regression for Long and NDVI predictors; this explains, on average, 66% and 65%, respectively, of the variance in annual CHIRPS precipitation. This underscores the critical impact of Long and NDVI on annual rainfall when using simple linear regression. The other variables exerted a comparatively weaker effect on annual rainfall, with

associated R^2 values below 52%, *MAE* values above 192 mm, and *RMSE* values above 265 mm.

Table 2. Values of the training statistics (annual and monthly average) characterizing the performance of the Simple Linear Regression models of Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) as a function of the study predictor variables for the period 2001–2023 (5 km resolution). Lat = Latitude; Long = Longitude; Alt = Altitude; NDVI = Normalized Difference Vegetation Index; LST = Land Surface Temperature. R^2 = Pearson's type Coefficient of Determination; *MAE* = Mean Absolute Error; *RMSE* = Root Mean Square Error.

Temporal Resolution	Statistic	Simple Linear Regression (LR) Model				
		Lat	**Long**	**Alt**	**NDVI**	**LST**
Annual	R^2 interval	[0.12, 0.45]	[0.16, 0.82]	[0.04, 0.64]	[0.36, 0.66]	[0.02, 0.46]
	R^2 (-)	0.50	0.66	0.51	0.65	[0.44]
	MAE (mm) interval	[188.3, 466.4]	[98.2, 474.6]	[150.4, 540.9]	[120.0, 411.7]	[174.4, 526.5]
	MAE (mm)	265.9	177.6	239.3	192.0	254.0
	RMSE (mm) interval	[219.1, 552.9]	[129.1, 591.5]	[189.2, 659.5]	[162.1, 526.8]	[213.0, 651.4]
	RMSE (mm)	317.3	259.4	313.5	264.1	324.4
January	R^2 (-)	0.25	0.52	0.30	0.54	0.19
	MAE (mm)	32.8	22.9	30.6	21.7	31.7
	RMSE (mm)	38.1	30.6	36.9	29.8	37.3
February	R^2 (-)	0.30	0.55	0.38	0.61	0.00
	MAE (mm)	38.4	25.9	33.9	23.4	22.4
	RMSE (mm)	44.3	35.3	41.7	32.4	34.0
March	R^2 (-)	0.36	0.17	0.05	0.59	0.02
	MAE (mm)	86.3	99.6	107.1	68.1	87.8
	RMSE (mm)	105.2	119.3	128.0	84.3	108.8
April	R^2 (-)	0.32	0.59	0.32	0.54	0.37
	MAE (mm)	51.5	35.9	49.9	39.0	48.4
	RMSE (mm)	58.8	45.7	58.7	48.2	58.1
May	R^2 (-)	0.17	0.71	0.52	0.39	0.46
	MAE (mm)	14.2	7.6	9.9	11.5	10.8
	RMSE (mm)	16.7	9.9	12.7	14.3	13.5
June	R^2 (-)	0.13	0.46	0.34	0.17	0.32
	MAE (mm)	5.7	4.7	4.5	5.7	5.3
	RMSE (mm)	8.5	6.7	7.4	8.3	7.6
July	R^2 (-)	0.00	0.25	0.25	0.11	0.29
	MAE (mm)	4.3	4.1	3.7	4.2	3.9
	RMSE (mm)	7.1	6.2	6.2	6.8	6.1
August	R^2 (-)	0.01	0.52	0.54	0.22	0.37
	MAE (mm)	4.6	3.0	2.6	4.0	3.4
	RMSE (mm)	6.0	4.2	4.1	5.3	4.7
September	R^2 (-)	0.09	0.75	0.66	0.34	0.39
	MAE (mm)	6.8	3.3	3.3	5.6	5.3
	RMSE (mm)	8.1	4.2	4.9	6.9	6.6
October	R^2 (-)	0.06	0.75	0.67	0.38	0.46
	MAE (mm)	18.2	8.9	8.5	13.9	13.1
	RMSE (mm)	21.6	11.1	12.7	17.5	16.5
November	R^2 (-)	0.03	0.72	0.64	0.47	0.46
	MAE (mm)	18.5	10.0	9.1	12.9	12.9
	RMSE (mm)	21.7	11.7	13.1	16.1	16.3
December	R^2 (-)	0.11	0.78	0.66	0.64	0.42
	MAE (mm)	24.4	11.5	12.5	14.0	18.4
	RMSE (mm)	28.1	14.2	17.4	17.9	21.5

In all months except March, the LR for Long was the best method. From May to December, the LR for Alt was generally the third best-performing method. During the winter months from January to March, the simple LR for NDVI ranked second-best; in April, it ranked third, and in December, fourth, according to all performance statistics. Only in July did LST become a slightly more significant predictor, ranking third among the best-performing models.

Therefore, Long and, in some months, Alt and NDVI were important predictors. Notably, during the drier months of June through August, lower R^2 values were observed, along with significantly lower *MAE* and *RMSE* values. These aspects appear to contradict each other, since low R^2 values, reflecting very weak data correlations and poor model performance, would usually imply very high MAE and RMSE values, unless under very special data conditions. When examining observations from the drier months (June through November), extremely low precipitation levels were recorded in these months (for certain years and the south-western Peruvian stations), and some data inconsistencies were identified (these data were subsequently discarded from the study).

Furthermore, when examining the respective scatter plots for some of the months and study predictors (Figure 3), the previously mentioned special data conditions were confirmed, namely, (i) nearly horizontal data distributions and (ii) a notable data concentration around zero precipitation (highlighted in the plots with ellipses), with few data pairs being far from zero precipitation. The first issue explains the low R^2 values. The second accounts for why relatively low *MAE* and *RMSE* values were obtained: the critical number of near-zero precipitation data pairs with low residual errors pulled the average errors (represented here by *MAE* and *RMSE*) towards relatively low values, despite the significant errors associated with the rest of the data pairs, as shown in the scatter plots of Figure 3.

The respective *MBE* values are not listed in Table 2 because all are zero (Figure 3), indicating that the linear regression methods used to model the CHIRPS precipitation dataset consistently provided unbiased estimates.

Thus, for the training stage and for annual rainfall, Long was the most effective predictor, followed by Alt and finally NDVI. This indicates a marked longitudinal gradient of precipitation in the basin, probably associated with the transition from wetter conditions on the eastern Andean flank to a drier climate towards the west, near the Pacific coast. Similarly, several comparable studies have identified a strong correlation between geographic coordinates (latitude and longitude) and precipitation [20,46,86].

It was not surprising that altitude (Alt) was a good predictor during certain months (from May to December), as research consistently shows that altitude enhances predictive accuracy when used with various statistical techniques and machine learning (ML) algorithms to process satellite-based precipitation products such as CHIRPS [87,88]. For example, Buttafuoco and Conforti [89] and Li et al. [90] demonstrated that including elevation in ML-based models improved rainfall forecast accuracy in mountainous regions. Moreover, Mahmoud et al. [91] and Daoud et al. [92] reported that the predictive power of remote sensing products was significantly increased when altitude was included, especially in arid and semi-arid areas. Several studies have shown that vegetation indices, such as NDVI, are closely associated with precipitation patterns, particularly in arid and semi-arid regions (such as those in the study basin), where plant growth is constrained by water availability [20,93–95].

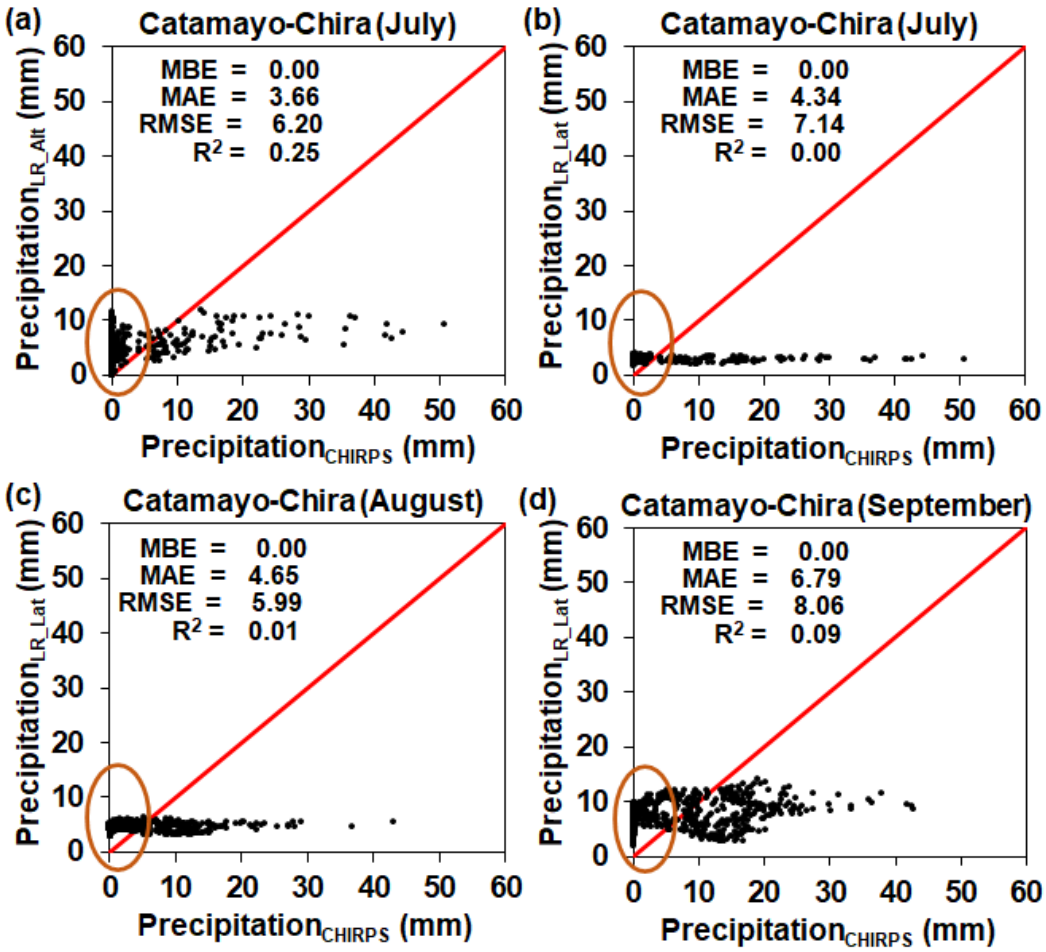

Figure 3. Scatter plots of monthly Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) precipitation and precipitation estimated through the simple linear regression (LR) for the following predictors: (**a**) altitude (Alt) in July; (**b**) latitude (Lat) in July; (**c**) Lat in August; and (**d**) Lat in September. Each plot displays the relevant performance metrics, namely, Pearson's Type II Coefficient of Determination (R^2), Mean Bias Error (*MBE*), Mean Absolute Error (*MAE*), and Root Mean Square Error (*RMSE*). The 1:1 coincidence (red) line is included in every plot. The ellipses highlight the zones with the highest data density.

3.1.2. Multivariable Machine Learning Models

Table 3 summarizes the parameters used to implement the SVM, RF, and ANN methods for predicting annual precipitation and monthly precipitation averages using the set of predictors (longitude, latitude, altitude, and NDVI). Although some earlier hydrology-focused studies reported limited parameter information [20], current best practices emphasize transparency in ML applications. In this study, all model parameters were optimized using a grid search algorithm, selecting the configurations that yielded the best training results; the final parameter settings are presented in Table 3. A comparison with the information listed by Chen et al. [20] shows that the SVM's internal parameters are very similar. However, the Gamma parameter in the current research was set to depend on the variance of the predictable variable ($Var(X)$). For the RF method, Chen et al. [20] report the number of trees and the mtry parameters, but omit the number of nodes, a critical parameter for this method [75]. The number of trees differs between the two studies. Finally, the ANN model's internal parameters differ, except for the number of hidden layers.

Table 3. Parameter configuration of the three nonlinear spatial downscaling models for annual precipitation and monthly average. SVM = Support Vector Machine; RF = Random Forest; ANN = Artificial Neural Network; Nr = number; q = number of predictor variables; X = predictor variable; $Var(X)$ = variance of X.

Downscaling Method	Parameter	Value
SVM	Kernel type	Linear, polynomial, radial basis function (rbf), sigmoid
	Cost	1.0
	Gamma	$1/(q \times Var(X))$
RF	Nr of trees	100
	mtry	4
ANN	Nr of hidden layers	5
	Epochs	300
	Learning rate	0.001
	Batch size	32

Similar to the simple LR methods (Table 2), Table 4 summarizes the fits of the multivariable ML methods using the evaluation statistics *MAE*, *RMSE*, and R^2. Both range ("interval") and global values are displayed. Similarly, *MBE* values are not listed in Table 4 because, although not strictly zero, they remain very close to zero. In contrast to what was observed for the simple LR methods, the top-performing multivariable ML methods—namely, RF, ANN, and SVM-rbf—consistently delivered the most accurate annual and monthly average estimates at the 5 km resolution (training stage). The quality of the precipitation estimates for the remaining ML methods was significantly lower. As one can see, in annual terms, the non-linear methods notably outperformed the linear one (MLR). This superiority was also confirmed on a monthly scale.

Table 4. Values of the training statistics (annual and monthly average) characterizing the performance of the multivariable machine learning models of Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) as a function of the study predictor variables for the period 2001–2023 (5 km resolution). MLR = Multiple Linear Regression; SVM = Support Vector Machine; SVM-lin = SVM with linear kernel; SVM-rbf = SVM with radial basis function kernel; SVM-poly = SVM with polynomial kernel; SVM-sig = SVM with sigmoid kernel; RF = Random Forest; ANN = Artificial Neural Network. R^2 = Pearson's type Coefficient of Determination; *MBE* = Mean Bias Error; *MAE* = Mean Absolute Error; *RMSE* = Root Mean Square Error.

Temporal Resolution	Statistic	Linear Regression Model	Non-Linear Regression Model					
		MLR	SVM-lin	SVM-rbf	SVM-poly	SVM-sig	RF	ANN
Annual	R^2 interval	[0.48, 0.83]	[0.45, 0.82]	[0.54, 0.88]	[0.36, 0.62]	[0.33, 0.74]	[0.99, 1.00]	[0.96, 0.99]
	R^2 (-)	0.78	0.76	0.84	0.65	0.71	1.00	0.98
	MAE (mm) interval	[84.4, 357.9]	[79.8, 354.1]	[61.2, 339.9]	[121.7, 380.2]	[109.9, 433.6]	[10.4, 33.4]	[26.5, 75.7]
	MAE (mm)	136.1	131.3	109.3	187.3	160.5	15.4	40.7
	RMSE (mm) interval	[117.2, 481.0]	[124.3, 496.9]	[91.4, 466.7]	[161.5, 522.5]	[150.5, 560.5]	[16.4, 47.3]	[38.4, 104.2]
	RMSE (mm)	204.2	214.1	175.4	256.0	233.9	23.7	58.7
January	R^2 (-)	0.61	0.61	0.84	0.52	0.58	0.98	0.91
	MAE (mm)	20.3	19.9	11.7	23.2	21.0	3.6	8.8
	RMSE (mm)	27.4	27.7	17.9	30.5	28.4	5.5	13.0
February	R^2 (-)	0.72	0.70	0.90	0.55	0.71	1.00	0.98
	MAE (mm)	18.1	17.0	8.8	26.0	18.6	2.1	4.7
	RMSE (mm)	27.3	28.5	16.0	34.5	27.7	3.3	6.8

Table 4. *Cont.*

Temporal Resolution	Statistic	Linear Regression Model	Non-Linear Regression Model					
		MLR	SVM-lin	SVM-rbf	SVM-poly	SVM-sig	RF	ANN
March	R^2 (-)	0.67	0.65	0.93	0.52	0.44	1.00	0.99
	MAE (mm)	58.9	57.8	22.6	71.6	78.6	4.1	12.0
	RMSE (mm)	75.7	78.0	34.3	91.2	97.9	6.1	15.7
April	R^2 (-)	0.71	0.69	0.95	0.57	0.66	1.00	0.98
	MAE (mm)	30.2	29.4	10.6	37.2	33.3	2.8	7.5
	RMSE (mm)	38.3	39.8	16.4	46.9	41.3	4.1	10.5
May	R^2 (-)	0.72	0.71	0.90	0.65	0.69	0.99	0.96
	MAE (mm)	7.5	7.3	3.6	8.1	7.5	1.1	2.5
	RMSE (mm)	9.7	9.9	5.9	10.8	10.1	1.7	3.7
June	R^2 (-)	0.49	0.45	0.81	0.57	0.42	0.98	0.94
	MAE (mm)	4.3	4.1	2.1	3.9	4.1	0.6	1.2
	RMSE (mm)	6.5	6.8	4.0	6.0	6.9	1.2	2.3
July	R^2 (-)	0.29	0.05	0.75	0.37	0.05	0.98	0.99
	MAE (mm)	3.7	2.8	1.5	2.5	2.8	0.4	0.4
	RMSE (mm)	6.0	7.3	3.6	5.7	7.3	1.0	0.8
August	R^2 (-)	0.60	0.57	0.85	0.59	0.56	0.99	0.96
	MAE (mm)	2.4	2.3	1.1	2.3	2.3	0.4	0.7
	RMSE (mm)	3.8	4.0	2.4	3.9	4.0	0.7	1.2
September	R^2 (-)	0.78	0.76	0.95	0.68	0.75	1.00	0.98
	MAE (mm)	2.9	2.7	0.9	3.4	2.7	0.2	0.7
	RMSE (mm)	3.9	4.2	1.8	4.8	4.2	0.5	1.3
October	R^2 (-)	0.79	0.75	0.97	0.68	0.71	1.00	0.99
	MAE (mm)	8.0	7.3	1.9	8.6	7.5	0.5	1.4
	RMSE (mm)	10.1	11.1	3.7	12.5	11.9	1.0	2.4
November	R^2 (-)	0.77	0.73	0.97	0.63	0.72	1.00	0.99
	MAE (mm)	8.6	8.1	2.0	9.1	8.3	0.5	0.9
	RMSE (mm)	10.5	11.5	3.7	13.4	11.7	0.9	1.6
December	R^2 (-)	0.80	0.79	0.95	0.63	0.78	1.00	0.99
	MAE (mm)	10.8	10.7	3.8	13.7	11.1	0.9	1.8
	RMSE (mm)	13.4	13.6	6.5	18.1	14.0	1.5	2.9

3.2. Downscaling Validation

3.2.1. Annual Precipitation

Table 5 presents the R^2, *MBE*, *MAE*, and *RMSE* values for annual precipitation using single- and multivariable ML models at 1 km resolution for the study period from 2001 to 2023. The table also reports the intervals of variation for these statistics, calculated for every year in the study period, except 2020, 2022 and 2023, as the number of stations with precipitation records did not exceed 4 in those years.

It is worth noticing that presenting the results for both simple linear models based on individual predictors and multivariable models (MLR, SVM, RF, ANN) using this set of predictors (Long, Lat, Alt and NDVI) is consistent with practices in comparable downscaling studies ([8,20,86]) and serves two complementary purposes: (1) to identify which individual variables (e.g., longitude and latitude) exhibit strong predictive potential, and (2) to evaluate the overall performance of more complex multivariable approaches. Although these results are presented together, the methods are not directly ranked against one another; rather, their performance under local conditions is assessed.

Regarding the R^2 statistic, very high values (above 0.96) were obtained for most downscaling methods, except for the simple linear regression (LR) model using the NDVI predictor and the non-linear SVM-poly model. It is also noteworthy that excellent R^2 values were recorded for the simple LR models using the predictor variables Long and Lat (geographical variables). However, in certain years, lower R^2 values of around 0.5, were

obtained, as depicted by the R^2 intervals associated with the NDVI simple LR model and the SVM-poly non-linear model. The R^2 values suggest that the worst downscaling methods were the simple LR model with the NDVI predictor and the non-linear SVM-poly model.

Table 5. Annual validation statistics (point-pixel) based on downscaled Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) values (1 km resolution) and observed data, considering records between 2001 and 2023. The yearly variation in the respective statistics is also shown. R^2 = Pearson's type Coefficient of Determination; *MBE* = Mean Bias Error; *MAE* = Mean Absolute Error; *RMSE* = Root Mean Square Error; Lat = Latitude; Long = Longitude; Alt = Altitude; NDVI = Normalized Difference Vegetation Index; LST = Land Surface Temperature; MLR = Multiple Linear Regression; SVM = Support Vector Machine; SVM-lin = SVM with linear kernel; SVM-rbf = SVM with radial basis function kernel; SVM-poly = SVM with polynomial kernel; SVM-sig = SVM with sigmoid kernel; RF = Random Forest; ANN = Artificial Neural Network.

Statistic	**Simple Linear Regression (LR) Model**					**MLR**
	Lat	**Long**	**Alt**	**NDVI**	**LST**	
R^2 (-) interval R^2 (-)	[1.00, 1.00] 1.00	[1.00, 1.00] 1.00	[0.89, 0.99] 0.97	[0.49, 0.94] 0.89	[0.89, 1.00] 0.97	[0.70, 1.00] 0.98
MBE (mm) interval *MBE* (mm)	[10.7, 43.2] 18.4	[−18.2, −11.6] −16.1	[−97.3, −38.2] −35.0	[−144.8, −46.6] −79.90	[−120.3, −22.6] −39.0	[14.6, 60.7] −32.9
MAE (mm) interval *MAE* (mm)	[10.7, 43.2] 18.4	[11.6, 18.2] 16.1	[41.5, 119.8] 73.3	[108.3, 373.2] 124.2	[24.1, 129.0] 71.5	[23.7, 278.6] 52.1
RMSE (mm) interval *RMSE* (mm)	[11.7, 45.7] 21.1	[12.7, 20.3] 17.7	[60.4, 142.6] 103.0	[148.4, 503.4] 190.6	[31.2, 182.8] 99.4	[31.8, 362.5] 87.8
Statistic	**SVM-lin**	**SVM-rbf**	**SVM-poly**	**SVM-sig**	**RF**	**ANN**
R^2 (-) interval R^2 (-)	[0.73, 1.00] 0.98	[0.97, 1.00] 0.99	[0.57, 1.00] 0.92	[0.91, 0.99] 0.98	[0.96, 1.00] 0.99	[0.96, 1.00] 0.99
MBE (mm) interval *MBE* (mm)	[−58.5, 10.2] −30.3	[−37.7, 25.5] −13.3	[−161.2, 21.1] −56.7	[−51.0, 30.3] −19.5	[−50.2, 4.6] −16.0	[−55.8, 45.2] −15.3
MAE (mm) interval *MAE* (mm)	[37.0, 263.5] 57.6	[21.4, 59.0] 35.3	[21.5, 207.4] 79.4	[50.1, 108.4] 62.4	[14.4, 60.0] 47.7	[9.7, 60.7] 40.4
RMSE (mm) interval *RMSE* (mm)	[44.9, 338.5] 89.6	[31.9, 70.6] 51.8	[25.9, 328.6] 159.6	[66.6, 143.4] 88.4	[16.7, 84.9] 66.7	[11.7, 96.6] 55.4

However, the more suitable *RMSE* statistic indicates that the best downscaling methods for simple LR models are those using the predictor variables Long and Lat (geographic), and, for the multivariable approach, the non-linear models ANN and SVM-rbf, since these methods yielded *RMSE* values below 56 mm. In this context, not only is it notable that the Long and Lat simple LR models outperformed (with notably lower values for the *RMSE*) the more complex non-linear models, but also that, apparently, altitude (Alt) is not as important a predictor variable as Long and Lat (for downscaling purposes), confirming what was previously observed for the original CHIRPS dataset at the annual scale (model training). These latter results were confirmed by the *MAE*, which yielded values below 41 mm for those four methods.

Nevertheless, the *MBE* could not confirm the above because it is limited in its ability to characterize average errors across observations and estimates; it is primarily suitable for characterizing over- or underestimations of precipitation. This weakness is clearly illustrated in Table 2 by the non-zero *MAE* values, in contrast to the null *MBE* value. For the *MBE*, errors of similar magnitude but with the opposite sign tend to cancel each other out, producing a false impression of a lower global error, while the *MAE* can perceive this.

However, *MBE* was used in this research not only to provide the reader with a basis for comparison with similar studies but also to gauge the bias in downscaled precipitation,

enabling the reader to judge whether it over- or underestimates the observed precipitation. Hereafter, Figure 4 shows the time fluctuations in the *RMBE* as a function of the best ML methods, namely, simple linear regression (LR) for the predictors Lat and Long and the SVM-rbf and ANN non-linear methods. In general, the LR for Lat overpredicts precipitation while the opposite was observed for the LR for Long. SVM-rbf tends to underpredict, whilst ANN underpredicts in the first half of the study period and overpredicts in the remaining portion.

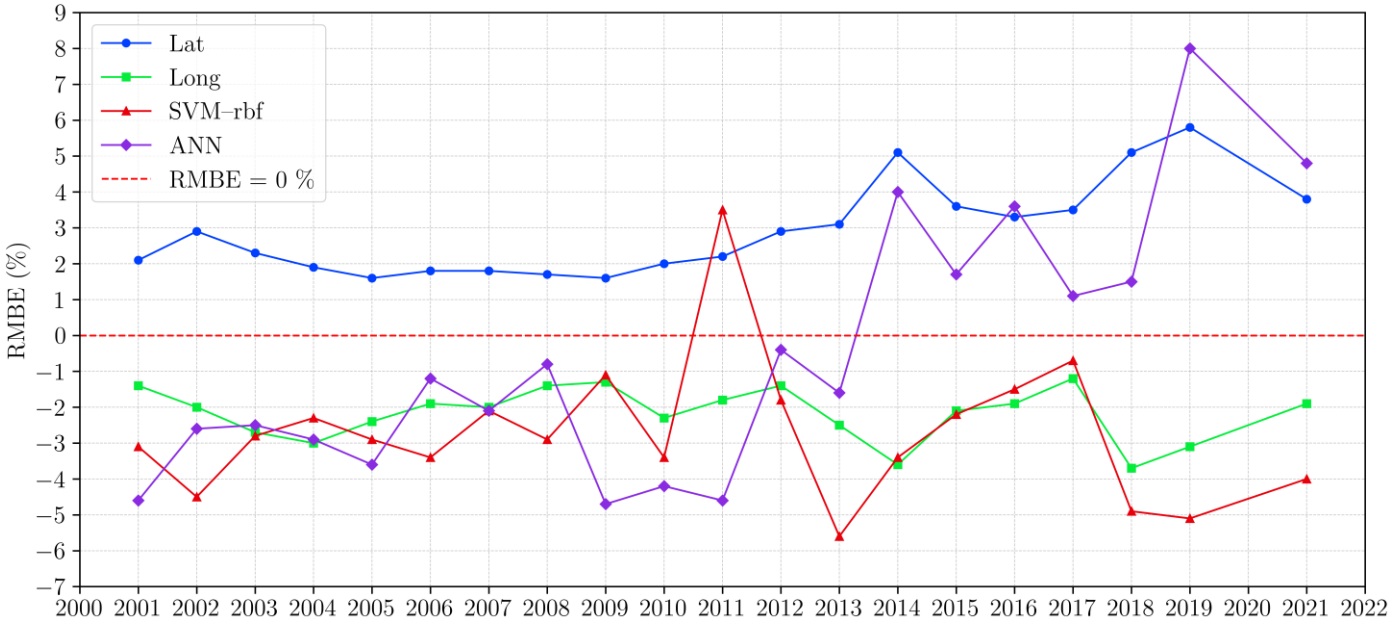

Figure 4. Annual variation in the Relative Mean Bias Error (*RMBE*) in the period 2001–2023 as a function of the best machine learning (ML) methods for downscaling precipitation, namely, simple linear regression (LR) for latitude (Lat) and longitude (Long), Support Vector Machine with the radial basis function kernel (SVM-rbf) and Artificial Neural Network (ANN).

Furthermore, the R^2 failed to differentiate the top four methods, as high values were observed for less suitable downscaling methods, confirming the expected weakness of this statistic in accurately characterizing the error between observations and the estimated data. It is also notable that the SVM method with the radial basis function kernel (SVM-rbf) and the ANN method outperformed all other SVM variants, as well as multivariable MLR and RF methods.

Figure 5 presents a comparison of the *RMSE* annual values using the original CHIRPS data set (5 km resolution) and the 1 km resolution downscaled precipitation from the best simple LR model (Long predictor) and the best multivariable model, the SVM-rbf. The figure also includes, in addition to the bars, the number of data points used to calculate the annual performance metrics. As previously mentioned, annual statistics were only calculated if the number of data points was at least 4 in the year of interest; accordingly, no metrics were calculated for 2020, 2022 and 2023 (Figure 5).

Figure 5 clearly shows that, over the entire study period, the *RMSE* values obtained after downscaling are substantially lower than those for the original CHIRPS dataset, confirming the considerable advantage of spatial downscaling from 5 km (original CHIRPS) to 1 km. Furthermore, the figure highlights that the simple LR method achieves better performance metrics than the more complex best multivariable method, the SVM-rbf.

Regarding the intervals of variation in the annual *RMSE* and *MAE* values reported in Table 5, higher (poorer) values were most often observed in 2017, as well as in 2006, 2007, 2010, 2011, and 2021 (depending on the study's ML method of interest). Inspection of

amount of available data for the annual calculation of the performance metrics (Figure 5) did not reveal any pattern(s) in the years mentioned above that could lead to the poorer *RMSE* and *MAE* values.

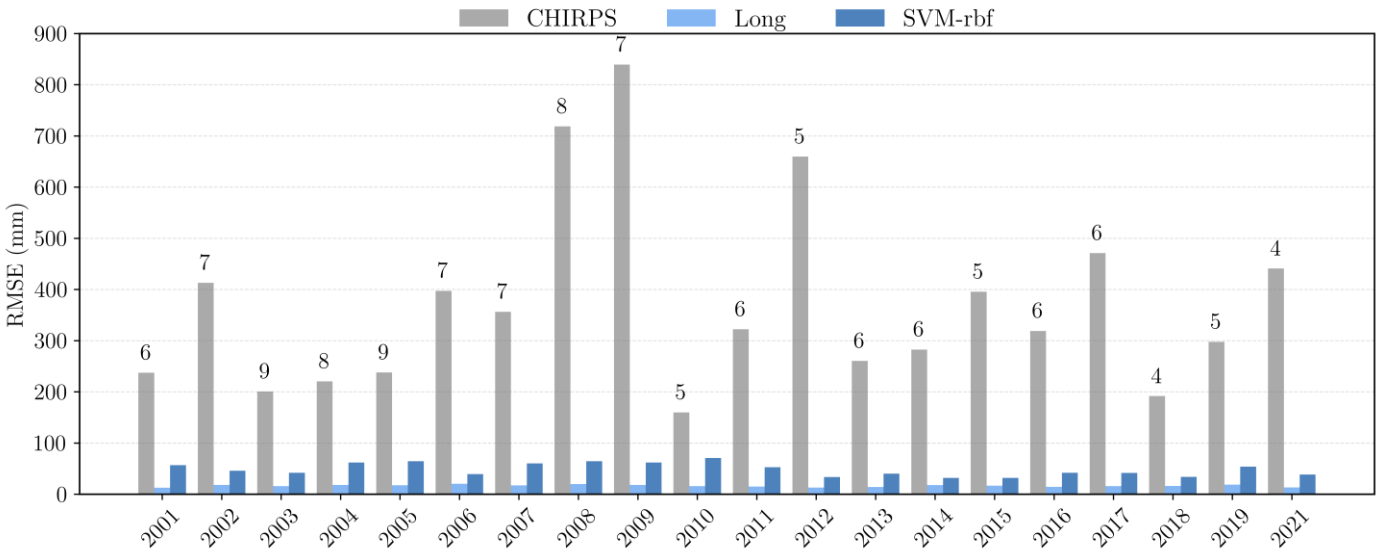

Figure 5. Annual Root Mean Square Error (RMSE) calculated over the period 2001–2021 (point-pixel analysis) obtained by comparing observed data with (i) the original Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) dataset (5 km resolution) and the downscaled precipitation (1 km resolution), using (ii) the best linear model (longitude—Long—predictor) and the best multivariable model (four predictors), Support Vector Machine with the Radial Basis Function kernel (SVM-rbf).

The analysis of performance metrics alone showed that the top-performing ML methods for annual precipitation estimation were the simple LR with Long and Lat predictors (single-variable methods) and the SVM-rbf and ANN (multivariable methods).

Research conducted in semi-arid basins and Andean Mountain regions agrees that algorithms such as SVM and RF excel in capturing the complex nonlinear relationships between precipitation and environmental predictors, including Alt, NDVI, and LST [22,96]. Conversely, the results presented herein indicate that simple linear regression (for the Long and Lat predictors) achieves better performance metrics than multivariable methods (non-linear) such as RF and some SVM variants for finer-scale precipitation prediction. In this context, studies in tropical regions have highlighted the critical importance of selecting appropriate predictors to maximize the performance of machine learning methods for precipitation downscaling [34]. This reinforces the relevance of this study's findings, which show that a simple linear method, when used with appropriate predictors, can yield better results than advanced non-linear multivariable ML methods and demonstrates a remarkable ability to improve both the accuracy and spatial resolution of annual downscaled precipitation. This emphasizes the main idea behind this research: several methods, ranging from straightforward to more complex, should be explored for downscaling precipitation from satellite-based products rather than blindly relying on the few methods reported in the specialized literature.

3.2.2. Average Monthly Precipitation

Table 6 shows the R^2, *MAE*, and *RMSE* values from the spatial downscaling of average monthly precipitation using both single-variable and multivariable methods at a 1 km resolution for the period from 2001 to 2023. For the non-linear models, because land surface temperature (LST) data for February and March were unavailable, average monthly precipitation was downscaled without this variable.

Table 6. Evaluation statistics values for average monthly observed precipitation and the downscaled Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) dataset (1 km resolution) calculated in the period 2001–2023 (point-pixel analysis). Lat = Latitude; Long = Longitude; MLR = Multiple Linear Regression; SVM = Support Vector Machine; SVM-lin = SVM with linear kernel; SVM-rbf = SVM with radial basis function kernel; SVM-poly = SVM with polynomial kernel; SVM-sig = SVM with sigmoid kernel; RF = Random Forest; ANN = Artificial Neural Network. R^2 = Pearson's type Coefficient of Determination; *MBE* = Mean Bias Error; *MAE* = Mean Absolute Error; *RMSE* = Root Mean Square Error.

Month	Statistic	Linear Regression			Non-Linear Method			
		Lat	Long	MLR	SVM-lin	SVM-rbf	RF	ANN
January	R^2 (-)	1.00	1.00	0.93	0.97	0.99	0.98	0.98
	MBE (mm)	2.1	−1.7	−3.2	−2.6	−0.1	1.7	5.7
	MAE (mm)	2.1	1.7	13.3	9.5	4.9	6.5	6.7
	RMSE (mm)	2.3	1.9	15.8	12.2	5.9	9.2	9.9
February	R^2 (-)	1.00	1.00	0.94	0.99	1.00	1.00	1.00
	MBE (mm)	2.7	−2.1	45.8	5.7	−3.2	0.9	1.4
	MAE (mm)	2.7	2.1	45.8	10.2	3.2	1.3	5.3
	RMSE (mm)	3.0	2.3	52.0	12.4	3.9	1.3	6.3
March	R^2 (-)	1.00	1.00	0.64	0.84	0.97	0.93	0.99
	MBE (mm)	7.3	−2.9	10.3	10.4	3.3	21.2	8.2
	MAE (mm)	7.3	2.9	87.3	45.6	15.1	24.7	11.7
	RMSE (mm)	8.0	3.2	110.9	61.4	19.8	39.1	14.6
April	R^2 (-)	1.00	1.00	0.90	0.98	0.98	0.99	0.99
	MBE (mm)	3.8	−3.0	−1.2	−1.4	−3.0	−1.4	4.6
	MAE (mm)	3.8	3.0	13.5	6.0	6.6	5.8	6.1
	RMSE (mm)	4.1	3.3	18.8	9.0	9.1	6.8	7.9
May	R^2 (-)	1.00	1.00	1.00	1.00	0.98	0.99	1.00
	MBE (mm)	0.7	−0.8	−1.1	−1.3	−1.8	−0.8	1.5
	MAE (mm)	0.7	0.8	1.5	1.5	1.9	2.3	1.7
	RMSE (mm)	0.8	0.9	1.9	2.1	3.9	3.2	2.4
June	R^2 (-)	1.00	1.00	1.00	1.00	0.99	0.99	0.97
	MBE (mm)	0.3	−0.3	0.4	0.1	−2.3	−0.6	−3.0
	MAE (mm)	0.3	0.3	0.6	0.7	2.3	1.5	4.9
	RMSE (mm)	0.3	0.3	1.0	0.9	5.3	2.6	7.6
July	R^2 (-)	1.00	1.00	1.00	1.00	1.00	0.99	0.98
	MBE (mm)	0.0	−0.2	−0.4	−0.1	−1.8	−0.5	−0.6
	MAE (mm)	0.0	0.2	0.5	0.1	1.8	0.9	1.7
	RMSE (mm)	0.0	0.2	0.7	0.1	4.0	2.3	3.5
August	R^2 (-)	1.00	1.00	1.00	1.00	1.00	0.98	0.99
	MBE (mm)	0.1	−0.2	−0.5	−0.4	−1.7	−1.4	−2.2
	MAE (mm)	0.1	0.2	0.6	0.6	1.7	1.8	2.5
	RMSE (mm)	0.1	0.3	0.9	0.9	3.5	2.9	4.1
September	R^2 (-)	1.00	1.00	1.00	0.99	0.98	0.98	1.00
	MBE (mm)	0.2	−0.4	−0.6	−0.4	−1.0	−1.8	−2.3
	MAE (mm)	0.2	0.4	0.7	1.0	1.2	2.1	2.5
	RMSE (mm)	0.3	0.4	1.1	1.4	2.4	3.7	4.0
October	R^2 (-)	1.00	1.00	1.00	0.99	0.99	1.00	0.99
	MBE (mm)	0.5	−1.0	−1.7	−2.1	−2.1	−3.1	−3.2
	MAE (mm)	0.5	1.0	1.8	2.7	3.2	3.3	3.2
	RMSE (mm)	0.5	1.1	2.6	3.9	4.9	5.3	5.1

Table 6. *Cont.*

Month	Statistic	Linear Regression			Non-Linear Method			
		Lat	**Long**	**MLR**	**SVM-lin**	**SVM-rbf**	**RF**	**ANN**
November	R^2 (-)	1.00	1.00	1.00	0.99	0.97	1.00	1.00
	MBE (mm)	0.4	−0.9	−2.0	−3.0	−2.5	−2.7	−1.3
	MAE (mm)	0.4	0.9	2.0	3.4	3.3	2.7	1.6
	RMSE (mm)	0.4	1.0	2.7	4.8	6.4	4.3	2.5
December	R^2 (-)	1.00	1.00	1.00	0.98	0.97	0.99	0.99
	MBE (mm)	0.9	−1.3	−1.1	−4.6	−3.4	−3.5	−2.0
	MAE (mm)	0.9	1.3	4.6	5.3	3.6	4.3	2.7
	RMSE (mm)	1.0	1.5	5.4	7.5	7.9	6.6	4.4

Notably, the R^2 values are nearly ideal (i.e., 1.0) across all months and ML methods, indicating strong agreement between the observed average monthly precipitation and the downscaled CHIRPS precipitation. The *RMSE* and *MAE* values confirm that the simple LR with Lat and Long predictors produced downscaled CHIRPS precipitation with relatively small errors in most months, except for March and April, when the *RMSE* reached 8.0 mm (and the *MAE* reached 7.3 mm). Indeed, the table shows that, systematically, the simple LR for the Lat and Long predictors is the top-performing single-variable method across all months. It is also noteworthy that the *MAE* and *RMSE* values for these linear regression models are systematically lower than those for the multivariable models, confirming the annual analysis, which showed that simple linear regression yields better results than the multivariable (linear and non-linear) regression. For the latter methods, SVM-rbf only performs acceptably in January, RF in February and April, and ANN in March, November, and December. In the remaining six months, SVM-lin is the best-performing non-linear multivariable method. This departs from what was observed in the annual analysis—namely, that the best-performing non-linear multivariable methods were SVM-rbf and ANN.

Table 6 also shows that, for the simple linear regressions using predictors Lat and Long, the *MBE* and *MAE* values are similar in magnitude across all months (although not similar in sign, as *MAE* is always positive). This indicates that the downscaled precipitation from these ML methods consistently either overestimated (positive) or underestimated (negative) precipitation at 1 km resolution each month. Conversely, if the over- or underestimations were not systematic, the magnitude of the *MBE* value for a given month would differ from that of the corresponding *MAE* value, since residuals of similar size but with opposite signs would tend to cancel each other out. This is not the case for the non-linear multivariable models. This is highlighted in Figure 6, which also demonstrates (considering *RMBE*) the systematic over- or underestimation of precipitation by the simple LR methods in all months. Once again, this is not observed in the non-linear multivariable methods in all months.

As previously described, the spatial analysis of the agreement between downscaled precipitation and observations was based on Relative Error (RE) at the pixel level, accounting for the locations of the 10 ground gauging stations. Therefore, Figure 7 illustrates the spatial–temporal distribution of the RE, expressed as a percentage, for monthly average precipitation downscaled using the Support Vector Machine with a linear kernel (SVM-lin), which was the best-performing multivariable method according to the evaluation metrics listed in Table 6.

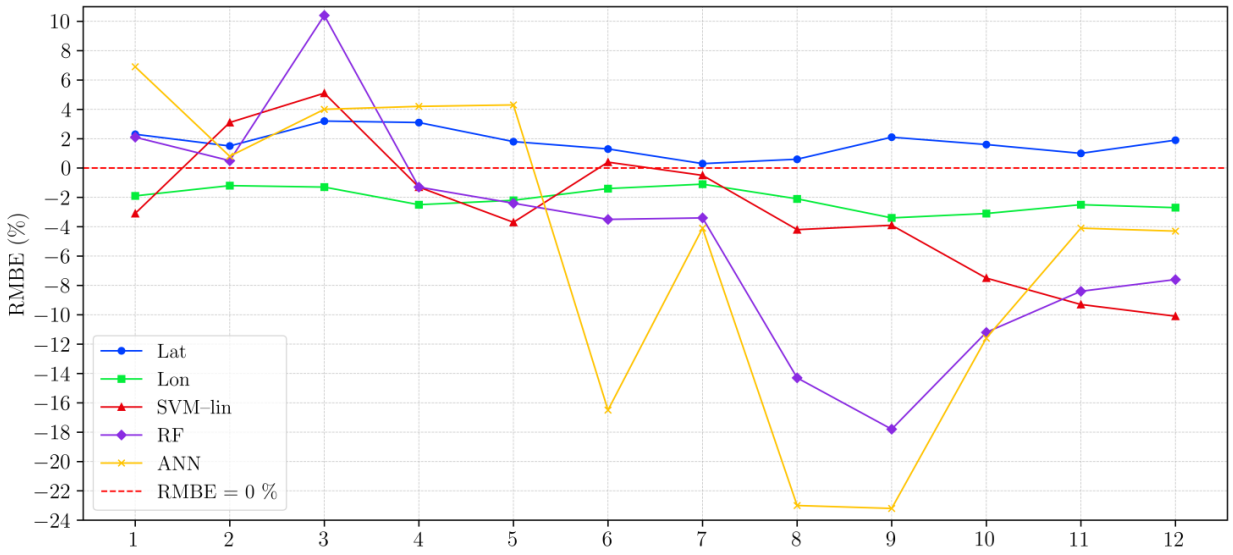

Figure 6. Average monthly variation in the Relative Mean Bias Error (*RMBE*) obtained in the period 2001–2023 as a function of the machine learning (ML) methods for downscaling precipitation, simple linear regression (LR) for latitude (Lat) and longitude (Long), Support Vector Machine with a linear kernel (SVM-lin), Random Forest (RF) and Artificial Neural Network (ANN).

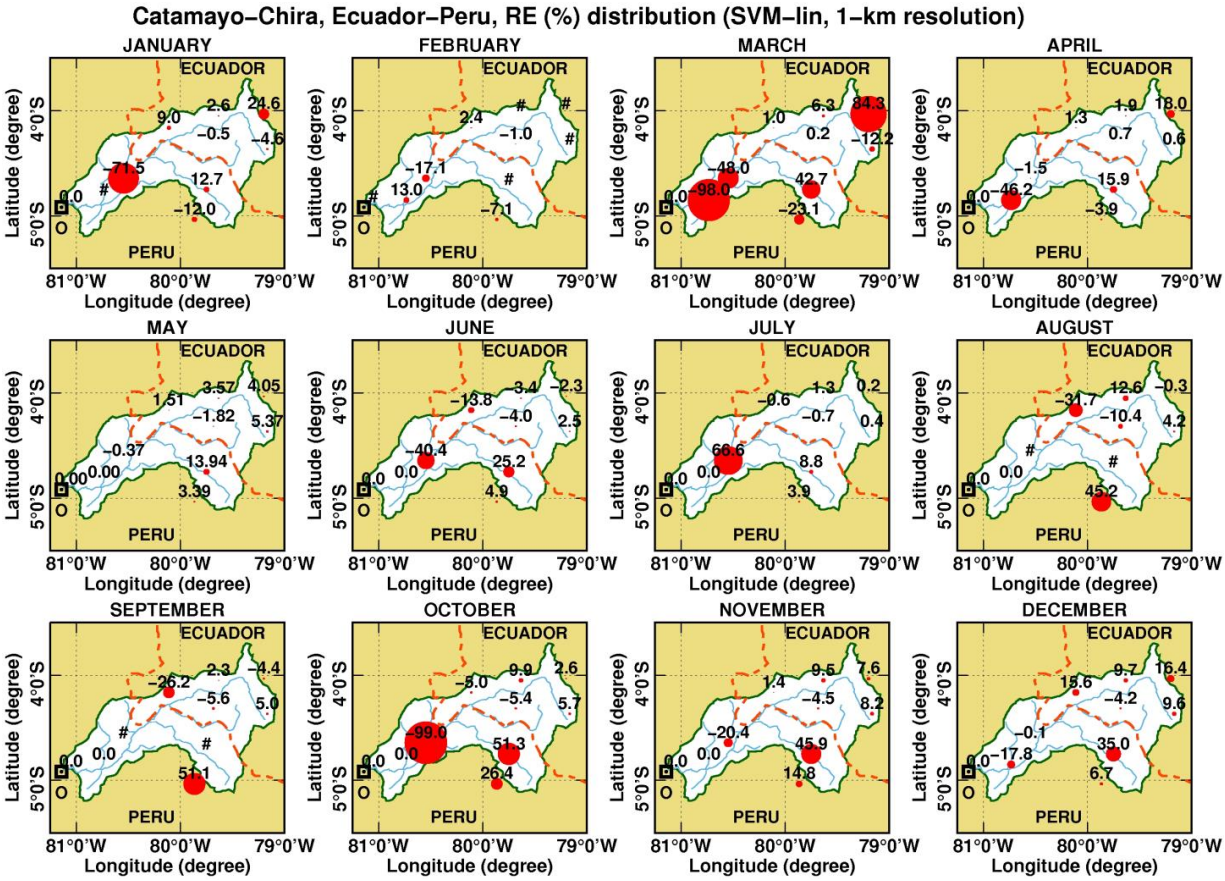

Figure 7. Spatial–temporal distribution of the monthly values of the Relative Error (RE) for the period 2001–2023 (1 km resolution), considering the downscaling method Support Vector Machine with the linear kernel (SVM-lin). The percentual RE values are labelled on the precipitation-gauging spots. Furthermore, the size of the plotted circles is proportional to their respective RE values. The outlet of the study basin is depicted through a square symbol. The symbol "#" is used when observations are not available. The international boundary is depicted as a red dashed line; the main rivers are shown as blue lines. Coordinate system: Geographic.

In general, Figure 7 shows relatively low RE values for most stations and months of the year, except in January, March, and the drier months of July and October. The most errors were observed at some southwestern gauging stations, although a significant error was also noted in March for one of the Ecuadorian stations. In the remaining months, low RE values were observed at most gauging stations. However, in some months, such as March, August, and September, there were stations for which either no observations were available or observations were discarded because they showed inconsistencies, such as zero precipitation in the months of interest throughout the study period (2001–2023). These cases are depicted in Figure 7 using the symbol "#".

Finally, it is important to note that the strong validation performance observed for the best single-variable and multivariable models (as reflected in near-perfect R^2 values and low error metrics at both monthly and annual scales) is not an artefact of model overfitting but a direct outcome of the method's design. The downscaling framework aims to reconstruct high-resolution precipitation fields for years with available observations using residual correction derived from known station-level errors. Because the residuals are computed from observed and predicted values at gauged locations and then spatially interpolated, the corrected downscaled values are expected to closely match the original observations. As such, the high validation metrics confirm the method's effectiveness in reproducing known rainfall fields at station locations, in line with its intended purpose.

3.2.3. Assessment of the Spatial Distribution of Precipitation for "El Niño" Years (ENYs)

Since the coastal "El Niño" year (ENY) 2017 was the only strong ENSO event during the study period, it was selected as the focus of the analysis. Figure 8 shows the spatial distribution of annual downscaled precipitation (obtained by simple linear regression using the longitude predictor at 1 km resolution) and the corresponding CHIRPS original data (at 5 km resolution) for a normal year (2001) and the coastal ENY 2017.

For a normal year, Figure 8a,b show that downscaled rainfall maps successfully replicate key orographic rainfall features observed in the CHIRPS product. Lower rainfall volumes are concentrated in the western, low-altitude parts of the basin, while the highest precipitation is seen along the eastern Andean ridge, where easterly winds transport moisture from the Amazon. In contrast, some interior valleys in the upper eastern ridge, such as the Catamayo Valley (Figure 1), despite their elevation, exhibit comparatively low precipitation due to their enclosed topography, which restricts moisture inflow. These patterns confirm that the model captures the dominant geographic controls on rainfall across the basin, though care is warranted when interpreting results in highly complex terrain.

When comparing the normal year (2001) with the coastal ENY 2017, a clear spatial anomaly becomes evident. In 2017, the coastal plain west of the Andes exhibits strong positive precipitation anomalies, while negative anomalies are observed in the eastern highlands. This pattern is consistent with prior ENSO-related studies (e.g., [73,90,97]), which report similar spatial disruptions associated with strong ENSO phases [80,98].

This suggests that the rainfall pattern in the coastal ENY 2017 was significantly stronger than that observed in 2001. This anomaly pattern aligns with findings by Tote et al. [37] for the strong negative ENSO year (1997/1998). It should also be noticed that the full extent of these anomalies is not captured here, as the analysis is based on calendar years rather than on hydrological years, which would include all El Niño months (December–March). It is hypothesized that the ENY 2023 also presented strong anomalies that notably disturbed the regular precipitation patterns, as reported for the coastal zones of Ecuador and Peru [79]. Nevertheless, as already stated, data availability was minimal for this year and was therefore discarded from this brief analysis.

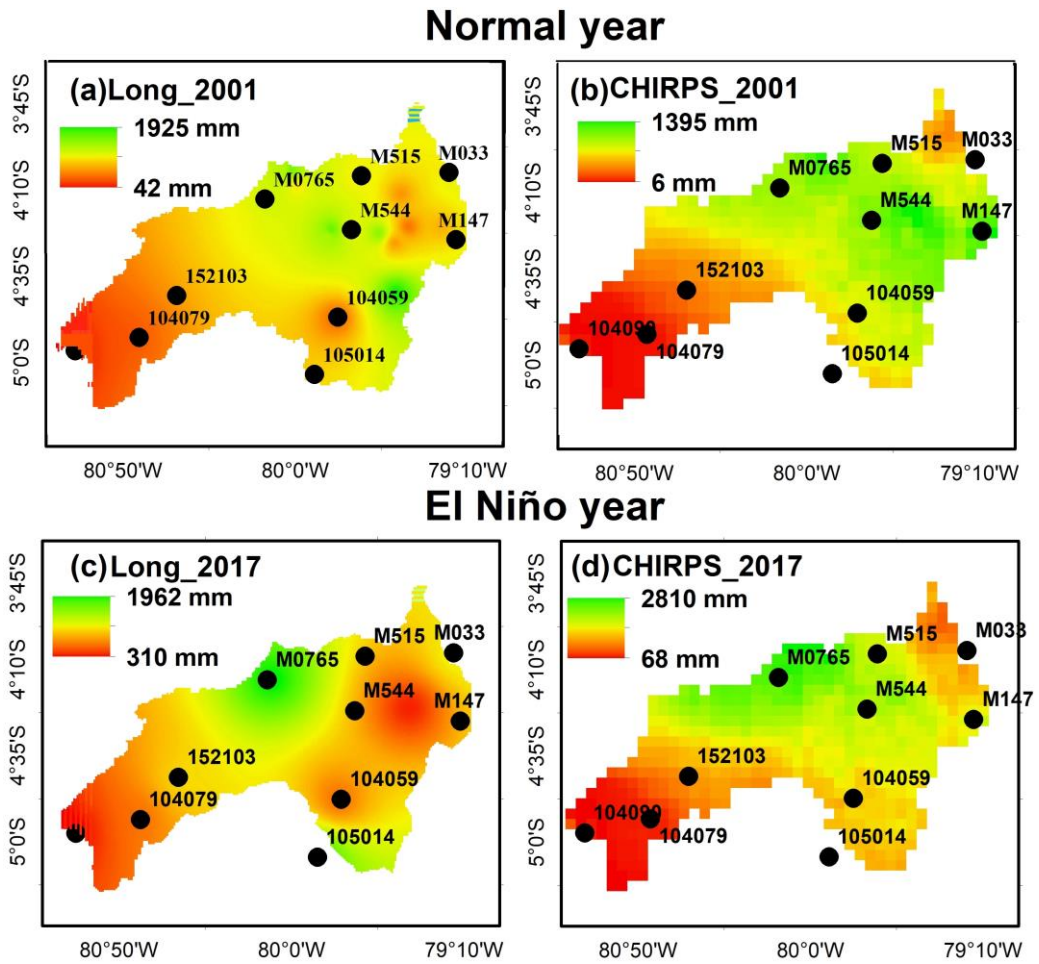

Figure 8. Spatial distribution of annual downscaled precipitation (simple linear regression for the longitude predictor; 1 km resolution) and original CHIRPS product (5 km resolution) for (**a**,**b**) a normal year (2001); (**c**,**d**) the "El Niño" year (2017). The gauging station identification codes are also displayed in the different plots. Coordinate system: Geographic.

4. Limitations and Future Work

There are some limitations to the approach presented in this study. First, the analysis focused exclusively on evaluating and comparing downscaling methods using geospatial predictors (e.g., latitude, elevation, NDVI, LST). As such, no comparisons were made with alternative baseline methods such as kriging of gauge observations, bias-corrected CHIRPS fields, or blended interpolation–satellite approaches. We recognize that such benchmarks could serve as valuable reference points for assessing the practical added value of the downscaling framework. However, incorporating these alternatives was beyond the scope of the current study, which examined the relative internal performance of downscaling algorithms using terrain and land-surface inputs. Future studies could expand the comparison to include kriged fields, ensemble models, or other bias-correction strategies, particularly for operational water resource applications.

Furthermore, this study employed grid search to tune and evaluate the SVM, RF, and ANN downscaling models; future work could also explore additional strategies to further assess model robustness. These may include sensitivity analyses of regularization parameters and the application of model-agnostic feature-importance methods (e.g., SHAP or permutation importance). Such approaches could enhance interpretability and provide deeper insights into model behavior.

Likewise, in this study, the same set of rain gauges was used both to adjust the satellite field estimation (residual correction) and to validate the results through point-to-pixel comparison. While this approach is common in satellite-based precipitation downscaling studies—particularly in regions with sparse ground-based observations—it may lead to overly optimistic performance estimates at gauged locations. It may not fully reflect model accuracy in ungauged areas. Future work could improve robustness by incorporating spatial or temporal cross-validation schemes, experimenting with alternative residual interpolation methods (such as kriging or other geostatistical approaches), or withholding a subset of stations for independent evaluation where gauge density permits.

It is also necessary to clarify another limitation of the study. Following a long-standing practice, the performance statistics were calculated by comparing the individual rain gauge readings used in the study to the 5×5 km^2 pixel non-linear model averages that contain them, as described in Figure 2 and in point 4 of the estimation protocol in Section 2.3. We acknowledge that this practice ignores both the uncertainty arising from individual rain gauge readings (which should not have a significant impact on measurements from monthly or larger sampling periods) and, more importantly, the uncertainty arising from the effect of spatial variability in individual readings on the pixel mean. Therefore, the findings presented in this paper, while significant, must be considered as only the first step toward further refinement.

While the downscaling results closely reflect observed rainfall values in most of the basin, caution is advised in interpreting model performance in topographically complex or isolated semi-arid zones. These areas, located in the upper eastern ridge of the catchment, present unique moisture transport conditions that may not be fully captured by large-scale predictors or interpolated residuals. Moreover, since climatic patterns in the basin are strongly controlled by elevation and terrain configuration, the robustness of results is lower in these zones. Future studies could benefit from enhanced local rainfall data to improve performance in such sensitive subregions.

Finally, the current study used a fixed set of five geospatial predictors; future work could explore feature selection techniques, test additional variables (e.g., cloud cover, surface roughness), or examine temporal transferability across different rainfall regimes.

5. Conclusions

The satellite-based 5 km Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) dataset was downscaled to 1 km estimates for the transnational basin of Catamayo–Chira, located between Ecuador and Peru, which supplies surface water to two crucial irrigation projects in southern Ecuador and northern Peru. These projects are vital for the economies of both countries. This analysis employed single-variable and multivariable machine learning (ML) methods, utilizing data from 10 gauged stations from 2001 to 2023. Predictors included longitude (Long), latitude (Lat), altitude (Alt), Normalized Difference Vegetation Index (NDVI), and Land Surface Temperature (LST). Various performance metrics, including the Coefficient of Determination (R^2), Mean Bias Error (*MBE*), Mean Absolute Error (*MAE*), and Root Mean Square Error (*RMSE*), were used to assess the methods' performance. Notably, the downscaled 1 km CHIRPS precipitation maps presented here are the first of their kind for the Catamayo–Chira basin, filling a critical gap in rainfall data for the region. This first-of-its-kind application in this transboundary basin also enables a comparative evaluation of single-variable versus multivariable downscaling methods under the basin's complex local conditions, highlighting which approaches best capture its spatial rainfall patterns.

The annual and monthly averages demonstrated a significant improvement in precipitation estimates after downscaling compared to the original CHIRPS estimates. Overall, the

best-performing single-predictor models were obtained using the geographic coordinates (longitude and latitude). At the annual scale, the downscaled 1 km product achieved very low errors for the best single-predictor models: LR-Long produced MAE = 16.1 mm and RMSE = 17.7 mm, and LR-Lat produced MAE = 18.4 mm and RMSE = 21.1 mm; at the mean monthly scale, the same pattern holds: LR-Long errors remain small even in the wettest months (e.g., March RMSE = 3.2 mm), while performance improved during the driest months. These results indicate that large-scale spatial gradients account for a substantial fraction of the variability in CHIRPS rainfall across the basin. In physical terms, these coordinate predictors act as effective proxies for the basin's strong topographic and climatic contrasts, including the marked precipitation increase toward the Andean headwaters and the maximum rainfall along the upper eastern ridge (Andean crest), where easterly moisture transport from the Amazon Basin is intercepted by the mountains. This orographic control, together with the progressive transition toward drier conditions in the lower western sector, helps explain why simple coordinate-based models can perform strongly in this Pacific–Andean setting.

In the multivariate approach, the non-linear methods notably outperformed the linear method (MLR), and, depending on the analysis timescale, different non-linear methods ranked among the top performers. At the annual scale, Support Vector Machine (SVM) with a Radial Basis Function kernel (SVM-rbf) and Artificial Neural Network (ANN), both using all predictors, were the best-performing non-linear methods (with RMSE < 56 mm and MAE < 41 mm). At the monthly scale, SVM with a linear kernel (SVM-lin) was the best-performing non-linear method ($0.1 \leq$ MAE ≤ 45.6 mm and $0.1 \leq$ RMSE ≤ 61.4 mm). Furthermore, the performance of these linear and non-linear methods tended to be better during the drier months, as reflected by the reduced monthly error metrics. Downscaled precipitation distributions effectively captured the differences between "El Niño" and non-"El Niño" events, a notable outcome that underscores the utility of the downscaling approaches used.

Although nonlinear models such as RF, SVM, and ANN are often favored in satellite precipitation downscaling due to their ability to capture complex relationships, our results demonstrate that, in certain contexts, simpler models can perform comparably. In the Catamayo–Chira basin, a Pacific–Andean system, rainfall variability is strongly influenced by west-to-east atmospheric transport from the Pacific Ocean. As a result, geographic predictors (longitude and latitude) exhibit high explanatory power. This is reflected in the strong performance of simple linear regression (LR) models using longitude and latitude alone, which achieved evaluation metrics similar to those of more complex multivariate methods. These findings are consistent with previous studies that have documented exceptions to the general performance trend of non-linear multivariable models, especially in regions where dominant geographic gradients govern precipitation patterns. Importantly, the LR models were not designed for direct comparison with multivariate or non-linear approaches, but rather to assess the individual predictive strength of key variables. When viewed in this context, the observed model behavior is not contradictory but instead reinforces the importance of regional characteristics and predictor selection in determining model suitability.

Ultimately, the approaches employed in this study can be replicated in catchments with limited gauging data, thereby enhancing their water resource management.

Author Contributions: Conceptualization, L.-F.D. and J.K.G.; methodology, L.-F.D. and J.K.G.; software, J.K.G. and L.-F.D.; validation, J.K.G.; formal analysis, J.K.G. and R.F.V.; investigation, J.K.G., L.-F.D., R.F.V. and C.L.O.; resources, J.K.G., C.L.O. and L.-F.D.; data curation, J.K.G.; writing—original draft preparation, J.K.G.; writing—review and editing, L.-F.D. and R.F.V.; visualization, J.K.G., L.-

F.D. and R.F.V.; supervision, L.-F.D.; project administration, L.-F.D.; funding acquisition, L.-F.D. All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded by Universidad Nacional de Loja (Loja, Ecuador).

Data Availability Statement: Most of the information used in this research can be downloaded from the internet through the links provided in the text of this manuscript. However, some of the data used in this research are subject to restrictions on availability. This data might be available upon reasonable request to all the co-authors of this research and their respective institutions.

Acknowledgments: The authors would like to thank Universidad Nacional de Loja for financial support for the present research (covering the APC).

Conflicts of Interest: The authors declare no conflicts of interest. Universidad Nacional de Loja had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

MBE	Mean Bias Error
RMBE	Relative Mean Bias Error
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
RE	Relative Error
ENY	"El Niño" year

References

1. Choi, E.; Rigden, A.J.; Tangdamrongsub, N.; Jasinski, M.F.; Mueller, N.D. US Crop Yield Losses from Hydroclimatic Hazards. *Environ. Res. Lett.* **2024**, *19*, 014005. [CrossRef]
2. Higgins, J.; Zablocki, J.; Newsock, A.; Krolopp, A.; Tabas, P.; Salama, M. Durable Freshwater Protection: A Framework for Establishing and Maintaining Long-Term Protection for Freshwater Ecosystems and the Values They Sustain. *Sustainability* **2021**, *13*, 1950. [CrossRef]
3. Bai, L.; Shi, C.; Li, L.; Yang, Y.; Wu, J. Accuracy of CHIRPS Satellite-Rainfall Products over Mainland China. *Remote Sens.* **2018**, *10*, 362. [CrossRef]
4. Ocampo-Marulanda, C.; Fernández-Álvarez, C.; Cerón, W.L.; Canchala, T.; Carvajal-Escobar, Y.; Alfonso-Morales, W. A Spatiotemporal Assessment of the High-Resolution CHIRPS Rainfall Dataset in Southwestern Colombia Using Combined Principal Component Analysis. *Ain Shams Eng. J.* **2022**, *13*, 101739. [CrossRef]
5. Kofidou, M.; Stathopoulos, S.; Gemitzi, A. Review on Spatial Downscaling of Satellite Derived Precipitation Estimates. *Environ. Earth Sci.* **2023**, *82*, 424. [CrossRef]
6. Du, T.L.T.; Lee, H.; Bui, D.D.; Graham, L.P.; Darby, S.D.; Pechlivanidis, I.G.; Leyland, J.; Biswas, N.K.; Choi, G.; Batelaan, O.; et al. Streamflow Prediction in Highly Regulated, Transboundary Watersheds Using Multi-Basin Modeling and Remote Sensing Imagery. *Water Resour. Res.* **2022**, *58*, e2021WR031191. [CrossRef]
7. Duque, L.-F.; O'Donnell, G.; Cordero, J.; Jaramillo, J.; O'Connell, E. Analysis of the Potential Impacts of Climate Change on the Mean Annual Water Balance and Precipitation Deficits for a Catchment in Southern Ecuador. *Hydrology* **2025**, *12*, 177. [CrossRef]
8. Sharifi, E.; Saghafian, B.; Steinacker, R. Downscaling Satellite Precipitation Estimates With Multiple Linear Regression, Artificial Neural Networks, and Spline Interpolation Techniques. *J. Geophys. Res. Atmos.* **2019**, *124*, 789–805. [CrossRef]
9. Ma, K.; Shen, C.; Xu, Z.; He, D. Transfer Learning Framework for Streamflow Prediction in Large-Scale Transboundary Catchments: Sensitivity Analysis and Applicability in Data-Scarce Basins. *J. Geogr. Sci.* **2024**, *34*, 963–984. [CrossRef]
10. Alsilibe, F.; Bene, K.; Bilal, G.; Alghafli, K.; Shi, X. Accuracy Assessment and Validation of Multi-Source CHIRPS Precipitation Estimates for Water Resource Management in the Barada Basin, Syria. *Remote Sens.* **2023**, *15*, 1778. [CrossRef]
11. Cui, R.; Ma, L.; Hu, Y.; Wu, J.; Li, H. Research on High-Resolution Modeling of Satellite-Derived Marine Environmental Parameters Based on Adaptive Global Attention. *Remote Sens.* **2025**, *17*, 709. [CrossRef]
12. Zhu, H.; Zhou, Q.; Krisp, J.M. Exploring Machine Learning Approaches for Precipitation Downscaling. *Geo-Spat. Inf. Sci.* **2025**, *28*, 2673–2689. [CrossRef]

13. Shahid, M.; Rahman, K.U.; Haider, S.; Gabriel, H.F.; Khan, A.J.; Pham, Q.B.; Mohammadi, B.; Linh, N.T.T.; Anh, D.T. Assessing the Potential and Hydrological Usefulness of the CHIRPS Precipitation Dataset over a Complex Topography in Pakistan. *Hydrol. Sci. J.* **2021**, *66*, 1664–1684. [CrossRef]
14. Atkinson, P.M. Downscaling in Remote Sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *22*, 106–114. [CrossRef]
15. Abdollahipour, A.; Ahmadi, H.; Aminnejad, B. A Review of Downscaling Methods of Satellite-Based Precipitation Estimates. *Earth Sci. Inform.* **2022**, *15*, 1–20. [CrossRef]
16. Sachindra, D.A.; Perera, B.J.C. Statistical Downscaling of General Circulation Model Outputs to Precipitation Accounting for Non-Stationarities in Predictor-Predictand Relationships. *PLoS ONE* **2016**, *11*, e0168701. [CrossRef]
17. Wilby, R.L.; Wigley, T.M.L. Precipitation Predictors for Downscaling: Observed and General Circulation Model Relationships. *Int. J. Climatol.* **2000**, *20*, 641–661. [CrossRef]
18. Hellström, C.; Chen, D.; Achberger, C.; Räisänen, J. Comparison of Climate Change Scenarios for Sweden Based on Statistical and Dynamical Downscaling of Monthly Precipitation. *Clim. Res.* **2001**, *19*, 45–55. [CrossRef]
19. Sylla, M.B.; Gaye, A.T.; Pal, J.S.; Jenkins, G.S.; Bi, X.Q. High-Resolution Simulations of West African Climate Using Regional Climate Model (RegCM3) with Different Lateral Boundary Conditions. *Theor. Appl. Climatol.* **2009**, *98*, 293–314. [CrossRef]
20. Chen, C.; Chen, Q.; Qin, B.; Zhao, S.; Duan, Z. Comparison of Different Methods for Spatial Downscaling of GPM IMERG V06B Satellite Precipitation Product Over a Typical Arid to Semi-Arid Area. *Front. Earth Sci.* **2020**, *8*, 536337. [CrossRef]
21. Wang, L.; Chen, R.; Han, C.; Yang, Y.; Liu, J.; Liu, Z.; Wang, X.; Liu, G.; Guo, S. An Improved Spatial-Temporal Downscaling Method for TRMM Precipitation Datasets in Alpine Regions: A Case Study in Northwestern China's Qilian Mountains. *Remote Sens.* **2019**, *11*, 870. [CrossRef]
22. Jing, W.; Yang, Y.; Yue, X.; Zhao, X. A Spatial Downscaling Algorithm for Satellite-Based Precipitation over the Tibetan Plateau Based on NDVI, DEM, and Land Surface Temperature. *Remote Sens.* **2016**, *8*, 655. [CrossRef]
23. Kay, A.L.; Rudd, A.C.; Coulson, J. Spatial Downscaling of Precipitation for Hydrological Modelling: Assessing a Simple Method and Its Application under Climate Change in Britain. *Hydrol. Process.* **2023**, *37*, e14823. [CrossRef]
24. Foody, G.M. Geographical Weighting as a Further Refinement to Regression Modelling: An Example Focused on the NDVI–Rainfall Relationship. *Remote Sens. Environ.* **2003**, *88*, 283–293. [CrossRef]
25. Chen, S.; Zhang, L.; She, D.; Chen, J. Spatial Downscaling of Tropical Rainfall Measuring Mission (TRMM) Annual and Monthly Precipitation Data over the Middle and Lower Reaches of the Yangtze River Basin, China. *Water* **2019**, *11*, 568. [CrossRef]
26. Chen, C.; Zhao, S.; Duan, Z.; Qin, Z. An Improved Spatial Downscaling Procedure for TRMM 3B43 Precipitation Product Using Geographically Weighted Regression. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 4592–4604. [CrossRef]
27. Jia, W. Regional Climate Model Evaluations of Long-Term changes in Total Precipitation and High Precipitation Events. Ph.D. Thesis, University of Illinois at Urbana-Champaign, Champaign, IL, USA, 2012.
28. Fang, J.; Du, J.; Xu, W.; Shi, P.; Li, M.; Ming, X. Spatial Downscaling of TRMM Precipitation Data Based on the Orographical Effect and Meteorological Conditions in a Mountainous Area. *Adv. Water Resour.* **2013**, *61*, 42–50. [CrossRef]
29. Valverde Ramirez, M.C.; Ferreira, N.J.; De Campos Velho, H.F. Linear and Nonlinear Statistical Downscaling for Rainfall Forecasting over Southeastern Brazil. *Weather Forecast.* **2006**, *21*, 969–989. [CrossRef]
30. Guven, A.; Pala, A. Comparison of Different Statistical Downscaling Models and Future Projection of Areal Mean Precipitation of a River Basin under Climate Change Effect. *Water Supply* **2022**, *22*, 2424–2439. [CrossRef]
31. Wilby, R.L.; Hay, L.E.; Leavesley, G.H. A Comparison of Downscaled and Raw GCM Output: Implications for Climate Change Scenarios in the San Juan River Basin, Colorado. *J. Hydrol.* **1999**, *225*, 67–91. [CrossRef]
32. Sachindra, D.A.; Ahmed, K.; Rashid, M.M.; Shahid, S.; Perera, B.J.C. Statistical Downscaling of Precipitation Using Machine Learning Techniques. *Atmos. Res.* **2018**, *212*, 240–258. [CrossRef]
33. Xu, R.; Chen, N.; Chen, Y.; Chen, Z. Downscaling and Projection of Multi-CMIP5 Precipitation Using Machine Learning Methods in the Upper Han River Basin. *Adv. Meteorol.* **2020**, *2020*, 8680436. [CrossRef]
34. Retalis, A.; Tymvios, F.; Katsanos, D.; Michaelides, S. Downscaling CHIRPS Precipitation Data: An Artificial Neural Network Modelling Approach. *Int. J. Remote Sens.* **2017**, *38*, 3943–3959. [CrossRef]
35. Richter, M.; Diertl, K.H.; Emck, P.; Peters, T.; Beck, E. Reasons for an Outstanding Plant Diversity in the Tropical Andes of Southern Ecuador. *Landsc. Online* **2009**, *12*, 1–35. [CrossRef]
36. Pineda, L.; Ntegeka, V.; Willems, P. Rainfall Variability Related to Sea Surface Temperature Anomalies in a Pacific–Andean Basin into Ecuador and Peru. *Adv. Geosci.* **2013**, *33*, 53–62. [CrossRef]
37. Tote, C.; Govers, G.; Van Kerckhoven, S.; Filiberto, I.; Verstraeten, G.; Eerens, H. Effect of ENSO Events on Sediment Production in a Large Coastal Basin in Northern Peru. *Earth Surf. Process. Landf.* **2011**, *36*, 1776–1788. [CrossRef]
38. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 Global Reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [CrossRef]
39. Campozano, L.; Sánchez, E.; Avilés, Á.; Samaniego, E. Evaluation of Infilling Methods for Time Series of Daily Precipitation and Temperature: The Case of the Ecuadorian Andes. *Maskana* **2014**, *5*, 99–115. [CrossRef]

40. World Meteorological Organization. *Climate Observations and Climate Data Management Guidelines, WMO/TD*; WMO: Geneva, Switzerland, 2009.
41. Paredes-Trejo, F.J.; Barbosa, H.A.; Lakshmi Kumar, T.V. Validating CHIRPS-Based Satellite Precipitation Estimates in Northeast Brazil. *J. Arid Environ.* **2017**, *139*, 26–40. [CrossRef]
42. Arechúa-Mazón, P.; Cisneros-Vaca, C.; Calahorrano-González, J.; Manzano-Cepeda, M. Assessment of Satellite Precipitation Products in an Andean Catchment: Ambato River Basin, Ecuador. *Hydrology* **2025**, *12*, 225. [CrossRef]
43. Andrade, J.M.; Ribeiro Neto, A.; Nóbrega, R.L.B.; Rico-Ramirez, M.A.; Montenegro, S.M.G.L. Efficiency of Global Precipitation Datasets in Tropical and Subtropical Catchments Revealed by Large Sampling Hydrological Modelling. *J. Hydrol.* **2024**, *633*, 131016. [CrossRef]
44. Hengl, T.; Heuvelink, G.B.M.; Tadić, M.P.; Pebesma, E.J. Spatio-Temporal Prediction of Daily Temperatures Using Time-Series of MODIS LST Images. *Theor. Appl. Climatol.* **2012**, *107*, 265–277. [CrossRef]
45. U.S. Geological Survey. Accuracy Assessment of Elevation Data. Available online: https://www.usgs.gov/special-topics/significant-topographic-changes-in-the-united-states/science/accuracy-assessment?utm_source (accessed on 29 January 2025).
46. Sathianarayanan, M.; Hsu, P.H. Spatial downscaling of gpm imerg v06 gridded precipitation using machine learning algorithms. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences—ISPRS Archives*; International Society for Photogrammetry and Remote Sensing: Johor Bahru, Malaysia, 2023; Volume 48, pp. 327–332.
47. Zhan, C.; Han, J.; Hu, S.; Liu, L.; Dong, Y. Spatial Downscaling of GPM Annual and Monthly Precipitation Using Regression-Based Algorithms in a Mountainous Area. *Adv. Meteorol.* **2018**, *2018*, 1506017. [CrossRef]
48. Zhang, Z.; Wang, D.; Qiu, J.; Zhu, J.; Wang, T. Machine Learning Approaches for Improving Near-Real-Time IMERG Rainfall Estimates by Integrating Cloud Properties from NOAA CDR PATMOS-x. *J. Hydrometeorol.* **2021**, *22*, 2767–2781. [CrossRef]
49. Essam, Y.; Huang, Y.F.; Ng, J.L.; Birima, A.H.; Ahmed, A.N.; El-Shafie, A. Predicting Streamflow in Peninsular Malaysia Using Support Vector Machine and Deep Learning Algorithms. *Sci. Rep.* **2022**, *12*, 3883. [CrossRef] [PubMed]
50. de Amorim, L.B.V.; Cavalcanti, G.D.C.; Cruz, R.M.O. The Choice of Scaling Technique Matters for Classification Performance. *Appl. Soft Comput.* **2023**, *133*, 109924. [CrossRef]
51. Vázquez, R.F.; Feyen, J. Assessment of the Effects of DEM Gridding on the Predictions of Basin Runoff Using MIKE SHE and a Modelling Resolution of 600m. *J. Hydrol.* **2007**, *334*, 73–87. [CrossRef]
52. Castleman, K.R. Geometric Transformations. In *Microscope Image Processing*; Merchant, F.A., Castleman, K.R., Eds.; Academic Press: Cambridge, MA, USA, 2023; pp. 47–54.
53. Accadia, C.; Mariani, S.; Casaioli, M.; Lavagnini, A.; Speranza, A. Sensitivity of Precipitation Forecast Skill Scores to Bilinear Interpolation and a Simple Nearest-Neighbor Average Method on High-Resolution Verification Grids. *Weather Forecast.* **2003**, *18*, 918–932. [CrossRef]
54. Zeng, Z.; Chen, H.; Shi, Q.; Li, J. Spatial Downscaling of IMERG Considering Vegetation Index Based on Adaptive Lag Phase. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4201415. [CrossRef]
55. Li, Y.; Pang, B.; Zheng, Z.; Chen, H.; Peng, D.; Zhu, Z.; Zuo, D. Evaluation of Four Satellite Precipitation Products over Mainland China Using Spatial Correlation Analysis. *Remote Sens.* **2023**, *15*, 1823. [CrossRef]
56. Seber, G.A.F.; Lee, A.J. *Linear Regression Analysis*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2003.
57. Montgomery, D.C.; Peck, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*, 5th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2012.
58. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
59. Liaw, A.; Wiener, M. Classification and Regression by RandomForest. *R News* **2002**, *2*, 18–22.
60. Tyralis, H.; Papacharalampous, G.; Langousis, A. A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources. *Water* **2019**, *11*, 910. [CrossRef]
61. Vapnik, V. *The Nature of Statistical Learning Theory*, 2nd ed.; Springer: New York, NY, USA, 1995.
62. Zamanisabzi, H.; King, J.P.; Dilekli, N.; Shoghli, B.; Abudu, S. Developing an ANN Based Streamflow Forecast Model Utilizing Data-Mining Techniques to Improve Reservoir Streamflow Prediction Accuracy: A Case Study. *Civ. Eng. J.* **2018**, *4*, 1135–1156. [CrossRef]
63. Abraham, A. Artificial Neural Networks. In *Handbook of Measuring System Design*; John Wiley & Sons: Chichester, UK, 2005.
64. Sen, A.; Gümüsay, M.U.; Kavas, A.; Bulucu, U. Programming an Artificial Neural Network Tool for Spatial Interpolation in GIS—A Case Study for Indoor Radio Wave Propagation of WLAN. *Sensors* **2008**, *8*, 5996–6014. [CrossRef]
65. Nur Adli Zakaria, M.; Abdul Malek, M.; Zolkepli, M.; Najah Ahmed, A. Application of Artificial Intelligence Algorithms for Hourly River Level Forecast: A Case Study of Muda River, Malaysia. *Alex. Eng. J.* **2021**, *60*, 4015–4028. [CrossRef]
66. Fensholt, R.; Sandholt, I. Evaluation of MODIS and NOAA AVHRR Vegetation Indices with in Situ Measurements in a Semi-Arid Environment. *Int. J. Remote Sens.* **2005**, *26*, 2561–2594. [CrossRef]
67. Paca, V.H.d.M.; Espinoza-Dávalos, G.E.; Moreira, D.M.; Comair, G. Variability of Trends in Precipitation across the Amazon River Basin Determined from the CHIRPS Precipitation Product and from Station Records. *Water* **2020**, *12*, 1244. [CrossRef]

68. Hsu, J.; Huang, W.-R.; Liu, P.-Y.; Li, X. Validation of CHIRPS Precipitation Estimates over Taiwan at Multiple Timescales. *Remote Sens.* **2021**, *13*, 254. [CrossRef]
69. Legates, D.R.; McCabe, G.J., Jr. Evaluating the Use of "Goodness-of-Fit" Measures in Hydrologic and Hydroclimatic Model Validation. *Water Resour. Res.* **1999**, *35*, 233–241. [CrossRef]
70. Vázquez, R.F.; Hampel, H. A Simple Approach to Account for Stage–Discharge Uncertainty in Hydrological Modelling. *Water* **2022**, *14*, 1045. [CrossRef]
71. Vázquez, R.F.; Feyen, L.; Feyen, J.; Refsgaard, J.C. Effect of Grid Size on Effective Parameters and Model Performance of the MIKE-SHE Code. *Hydrol. Process.* **2002**, *16*, 355–372. [CrossRef]
72. Walther, B.A.; Moore, J.L. The Concepts of Bias, Precision and Accuracy, and Their Use in Testing the Performance of Species Richness Estimators, with a Literature Review of Estimator Performance. *Ecography* **2005**, *28*, 815–829. [CrossRef]
73. Shi, Y.; Song, L.; Xia, Z.; Lin, Y.; Myneni, R.B.; Choi, S.; Wang, L.; Ni, X.; Lao, C.; Yang, F. Mapping Annual Precipitation across Mainland China in the Period 2001–2010 from TRMM3B43 Product Using Spatial Downscaling Approach. *Remote Sens.* **2015**, *7*, 5849–5878. [CrossRef]
74. Moriasi, D.N.; Arnold, J.G.; Van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Trans. ASABE* **2007**, *50*, 885–900. [CrossRef]
75. Vázquez, R.F.; Mejía, D.; Mosquera, P.V.; Hampel, H. Estimating the Maximum Depth of Andean Lakes: A Comparative Analysis Using Machine Learning. *Water* **2024**, *16*, 3570. [CrossRef]
76. Shin, M.-J.; Jung, Y. Using a Global Sensitivity Analysis to Estimate the Appropriate Length of Calibration Period in the Presence of High Hydrological Model Uncertainty. *J. Hydrol.* **2022**, *607*, 127546. [CrossRef]
77. Wei, C.; Zhao, Q.; Lu, Y.; Fu, D. Assessment of Empirical Algorithms for Shallow Water Bathymetry Using Multi-Spectral Imagery of Pearl River Delta Coast, China. *Remote Sens.* **2021**, *13*, 3123. [CrossRef]
78. Zhao, N.; Yue, T.; Zhou, X.; Zhao, M.; Liu, Y.; Du, Z.; Zhang, L. Statistical Downscaling of Precipitation Using Local Regression and High Accuracy Surface Modeling Method. *Theor. Appl. Climatol.* **2017**, *129*, 281–292. [CrossRef]
79. Peng, Q.; Xie, S.-P.; Passalacqua, G.A.; Miyamoto, A.; Deser, C. The 2023 Extreme Coastal El Niño: Atmospheric and Air-Sea Coupling Mechanisms. *Sci. Adv.* **2024**, *10*, eadk8646. [CrossRef]
80. Huang, P.; Chen, Y.; Li, J.; Yan, H. Redefined Background State in the Tropical Pacific Resolves the Entanglement between the Background State and ENSO. *NPJ Clim. Atmos. Sci.* **2024**, *7*, 147. [CrossRef]
81. Mu, Y.; Jones, C.; Carvalho, L.M.V.; Xue, L.; Liu, C.; Ding, Q. Pacific Decadal Oscillation and ENSO Forcings of Northerly Low-Level Jets in South America. *NPJ Clim. Atmos. Sci.* **2024**, *7*, 297. [CrossRef]
82. Refsgaard, J.C. Parameterisation, Calibration and Validation of Distributed Hydrological Models. *J. Hydrol.* **1997**, *198*, 69–97. [CrossRef]
83. Beven, K.J. *Rainfall-Runoff Modelling: The Primer*, 2nd ed.; John Wiley and Sons: Chichester, UK, 2012.
84. Vázquez, R.F.; Brito, J.E.; Hampel, H.; Birkinshaw, S. Assessing the Performance of SHETRAN Simulating a Geologically Complex Catchment. *Water* **2022**, *14*, 3334. [CrossRef]
85. Nash, J.E.; Sutcliffe, J. V River Flow Forecasting through Conceptual Models Part I—A Discussion of Principles. *J. Hydrol.* **1970**, *10*, 282–290. [CrossRef]
86. Jing, W.; Yang, Y.; Yue, X.; Zhao, X. A Comparison of Different Regression Algorithms for Downscaling Monthly Satellite-Based Precipitation over North China. *Remote Sens.* **2016**, *8*, 835. [CrossRef]
87. Bayissa, Y.; Tadesse, T.; Demisse, G.; Shiferaw, A. Evaluation of Satellite-Based Rainfall Estimates and Application to Monitor Meteorological Drought for the Upper Blue Nile Basin, Ethiopia. *Remote Sens.* **2017**, *9*, 669. [CrossRef]
88. Gupta, S.; Chandrashekhar, U.; Vippagunta, P.; Latha Balakisti, S.; Rao Lakkimsetty, N.; Sree Pokkuluri, K. Predicting Future Rainfall with Various Machine Learning Models. *J. Inf. Syst. Eng. Manag.* **2024**, *2025*, 2468–4376. [CrossRef]
89. Buttafuoco, G.; Conforti, M. Improving Mean Annual Precipitation Prediction Incorporating Elevation and Taking into Account Support Size. *Water* **2021**, *13*, 830. [CrossRef]
90. Li, H.; Zhang, Y.; Lei, H.; Hao, X. Machine Learning-Based Bias Correction of Precipitation Measurements at High Altitude. *Remote Sens.* **2023**, *15*, 2180. [CrossRef]
91. Mahmoud, M.T.; Mohammed, S.A.; Hamouda, M.A.; Mohamed, M.M. Impact of Topography and Rainfall Intensity on the Accuracy of IMERG Precipitation Estimates in an Arid Region. *Remote Sens.* **2021**, *13*, 13. [CrossRef]
92. Ben Daoud, N.; Daoudi, L.; Rachdane, M.; Gourfi, A.; Saidi, M.E. Suitability of Satellite-Based Rainfall Products for Estimating Rainfall Erosivity in Areas with Contrasted Climate and Terrain Properties: Example of West-Central Morocco. *J. Earth Syst. Sci.* **2024**, *133*, 78. [CrossRef]
93. Immerzeel, W.W.; Rutten, M.M.; Droogers, P. Spatial Downscaling of TRMM Precipitation Using Vegetative Response on the Iberian Peninsula. *Remote Sens. Environ.* **2009**, *113*, 362–370. [CrossRef]
94. Chen, F.; Yu, L.; Qiang, L.; Li, X. Spatial Downscaling of TRMM 3B43 Precipitation Considering Spatial Heterogeneity. *Int. J. Remote Sens.* **2014**, *35*, 3074–3093. [CrossRef]

95. Hunink, J.E.; Immerzeel, W.W.; Droogers, P. A High-Resolution Precipitation 2-Step Mapping Procedure (HiP2P): Development and Application to a Tropical Mountainous Area. *Remote Sens. Environ.* **2014**, *140*, 179–188. [CrossRef]
96. Medrano, S.C.; Satgé, F.; Molina-Carpio, J.; Zolá, R.P.; Bonnet, M.-P. Downscaling Daily Satellite-Based Precipitation Estimates Using MODIS Cloud Optical and Microphysical Properties in Machine-Learning Models. *Atmosphere* **2023**, *14*, 1349. [CrossRef]
97. Rollenbeck, R.; Orellana-Alvear, J.; Bendix, J.; Rodriguez, R.; Pucha-Cofrep, F.; Guallpa, M.; Fries, A.; Célleri, R. The Coastal El Niño Event of 2017 in Ecuador and Peru: A Weather Radar Analysis. *Remote Sens.* **2022**, *14*, 824. [CrossRef]
98. Vega-Camarena, J.P.; Brito-Castillo, L. Precipitation Response in Mountainous and Coastal Regions of Northwestern Mexico under ENSO Scenarios during the Landfall of Tropical Cyclones. *Theor. Appl. Climatol.* **2024**, *155*, 8599–8610. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.