A Nonparametric Statistical Technique for Spatial Downscaling of Precipitation over High Mountain Asia

Yiwen Mei[1], Viviana Maggioni[1], Paul Houser[2], Yuan Xue[2] and Tasnuva Rouf[1]

[1] Sid and Reva Dewberry Department of Civil, Environmental & Infrastructure Engineering, George Mason University

[2] Department of Geography and Geoinformation Science, George Mason University

Submitted to: Water Resources Research

Submission date: March 2020

Corresponding Author: Yiwen Mei, CEIE, George Mason University, Fairfax, VA 22030

Tel: (860) 465-7444, Email: ymei2@gmu.edu

Abstract

The accurate representation of the local-scale variability of precipitation plays an important role in understanding the hydrological cycle and land-atmosphere interactions in the High Mountain Asia region. Therefore, the development of hyper-resolution precipitation data is of urgent need. In this study, we propose a statistical framework to downscale the Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2) precipitation product using the random forest classification and regression algorithm. A set of variables representing atmospheric, geographic, and vegetation cover information are selected as model predictors, based on a recursive feature elimination method. The downscaled precipitation product is validated in terms of magnitude and variability against a set of ground- and satellite-based observations. Results suggest improvements with respect to the original resolution MERRA-2 precipitation product and comparable performance with gauge-adjusted satellite precipitation products.

Keyword: precipitation downscaling, random forest, predictor selection, High Mountain Asia, topographic correction

1. Introduction

The resolution of the input precipitation dataset is one of the most critical elements in hyper-resolution hydrologic modeling (~1-km or finer), as its accuracy greatly dictates the space-time representativeness of the output of land surface fluxes and states. This is particularly critical for complex terrain regions, such as High Mountain Asia (HMA), due to the highly localized precipitation gradients induced by topography. However, the availability of in-situ surface measurements for hydrologic, weather, and climate studies is scarce and their accuracy is influenced by the region terrain complex. Specifically, rain gauge observations are subject to wind-induced under-catch biases; and ground-based weather radar networks are rare and suffer from significant beam blockage and ground return problems in mountainous regions [Gou et al., 2018; Chen et al., 2015; Tong et al., 2014b]. These factors undermine the use of ground-based networks in hydrologic modeling in such basins. Precipitation information can also be obtained from satellite observations and numerical weather prediction models, which provide continuous and consistent estimates. However, their resolutions are too coarse to catch the fine scale variability of precipitation systems over mountainous areas. This is a challenge especially in the context of hyper-resolution, given the prominent heterogeneity of mountainous hydrologic processes [Ma et al., 2018a, 2018b; Tong et al., 2014a]. Thus, downscaling techniques are required to develop hyper-resolution precipitation datasets to be used in land surface modeling [Zorzetto and Marani, 2019].

In broad terms, downscaling techniques can be classified into dynamical and statistical methods [Maraun et al., 2010; Haylock et al., 2006]. Statistical downscaling of precipitation is relatively simple and computationally efficient if compared to dynamical downscaling, which requires the use of either local-scale models or regional climate models. Machine learning is a statistical technique that maps the predictor(s) with a predictand without constructing an explicit function and relying on (if any) existing physical or statistical relationships between the two. Among the plethora of machine learning techniques, the random forest (RF) algorithm stands out for its ability to deal with complex nonlinear relationships and to minimize the overfitting problem [Breiman, 2001]. RF has been applied in a range of hydrologic-related studies, such as streamflow prediction, estimation of soil moisture, groundwater potential mapping, digital soil mapping, susceptibility assessment of natural hazards [Shortridge et al., 2016; Ali et al., 2015; Goetz et al., 2015; Heung et al., 2015; Naghibi and Pourghasemi, 2015].

The use of RF in precipitation downscaling is a relatively new topic. Ibarra-Berastegi et al. [2011] applied an analogues method, based on RF and multilinear regression, to downscale reanalysis precipitation products over two basins in the Ebro Valley in Spain. Their results indicate that the analogues-RF combined method outperformed the analogues-regression combined one. He et al. [2016] proposed a machine-learning algorithm called Prec-DWARF (Precipitation Downscaling With Adaptable RFs) for spatial precipitation downscaling. Prec-DWARF is shown to successfully reproduce the space-time and statistical characteristics of the original rainfall field, but with an overestimation of light rain rates and an underestimation of extreme rainfall. They further discovered that by separately building RFs for low-to-moderate and extreme rainfall rates, the skewed precipitation distribution could be resolved. Bhuiyan et al. [2018] developed a downscaling framework to generate an improved ensemble precipitation product based on quantile-based RF via blending four different precipitation products (three satellite-based products and one reanalysis product), air temperature, near-surface soil moisture, and terrain elevation information, over the Iberian Peninsula. Their results indicate higher consistency of the downscaled precipitation products with ground-based observations compared to any of the single product.

Selecting the predictor variables is an important step, yet receiving little attention in RF-based precipitation downscaling. In the general formulation of RF regression, the importance of a predictor on a predictand is measured by the mean decrease in accuracy, MDA, defined as the change in mean square error (MSE) of the out-of-bag sample to the original model when the predictor is randomly permuted and used in a new prediction [Breiman, 2001]. Randomly permuting an irrelevant predictor should make no difference in the new prediction and thus results in negligible increments in MSE [Genuer et al., 2010; Grömping, 2009]. This concept has been adopted to quantify predictor importance in precipitation downscaling studies [Ma et al., 2018c; Bhuiyan et al., 2018; He et al., 2016]. For example, Bhuiyan et al. [2018] showed that soil moisture and other precipitation variables were ranked as the most important predictors for their study conducted over the Iberian Peninsula. Topographic (elevation, aspect, and slope) and geographic (latitude and longitude) variables are characterized by lower importance in daily precipitation downscaling [Bhuiyan et al., 2018; He et al., 2016], although their predictive value increases when downscaling longer-term (yearly to monthly) cumulative precipitation [Ma et al., 2018c; Xu et al., 2015].

Although the concept of MDA provides a relative ranking of predictor importance, it does not distinguish relevant from irrelevant predictors. To build a parsimonious prediction model for precipitation, additional procedures are required. One popular approach is the recursive feature elimination (RFE) that incorporates a predictor importance index to select a minimal set of variables [Degenhardt et al., 2017; Díaz-Uriarte and Alvarez de Andrés 2006]. RFE starts with the full list of variables by fitting RF model to the set. A portion of variables with the lowest predictor importance index are disregarded and a new RF model is fitted to the rest. The process is repeated until a single variable is left as input. Model performance of every iteration is measured by the MSE of out-of-bag sample and relevant variables are those that make up the model with the minimum out-of-bag MSE. This process provides an intuitive view of the evolution of model performance with the number of predictor variables. While this is a popular approach in the field of bioinformatics, we believe that this is the first time it has been applied to precipitation downscaling.

To explore the potential of RF in precipitation downscaling, this work presents a scheme to produce precipitation at 1km spatial resolution over HMA. A predictor selection method is introduced to reduce the model space, while retaining high model accuracy. The downscaled precipitation dataset is evaluated in terms of rain magnitude and pattern against ground-based observations and high-resolution satellite precipitation products. This work seeks to investigate i) the usefulness of variables representing near-surface atmospheric conditions, geospatial information, and seasonality in precipitation spatial downscaling over a complex terrain region; and ii) the application of an RFE-based procedure for predictor selection. The study is organized as follows. Section 2 describes the HMA region and all data used for downscaling and validation. Section 3 introduces the downscaling framework and the method to validate the downscaled precipitation products. Results are shown in Section 4 and discussed in Section 5. Conclusions and recommendations are presented in Section 6.

2. Study Area and Datasets

2.1. High Mountain Asia

The HMA region is one of the most extensive mountain systems in the world and contains the largest concentration of glacier ice outside the Polar Regions (Figure 1). It is the source of many major Asian river systems, such as Indus, Brahmaputra, Salween, Mekong, Yellow, and Yangtze rivers, which support the ecosystem services, agriculture, energy and livelihood of over one billion people. The region features a complex precipitation climatology under the combined and competitive influences of the Indian and East Asian monsoon systems and of the westerlies disturbances originated from the Caspian and Mediterranean Seas, modulated by the highly elevated terrain [Cannon et al., 2017; Wei et al., 2016; Maussion et al., 2014]. This study focuses on the region that extends from 61°E to 90°E and from 20°N to 41°N, including the central and western Tibetan Plateau (TP) and several major mountain ranges like the Hindu Kush, the Pamir, the Karakoram, the Kunlun, and the Himalaya. These mountain ranges serve as sources of the, from west to east, Amu Darya River, Indus River, Tarim River, Ganges River, and Brahmaputra River.

2.2. Dataset Used in the Downscaling Algorithm

The precipitation data to be downscaled are the uncorrected total precipitation (i.e., without corrections from ground-based stations) from the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). MERRA-2 provides hourly cumulative precipitation at the land surface with a horizontal resolution of 0.5° × 0.625° [Gelaro et al., 2017]. Other MERRA-2 variables used in the precipitation downscaling are surface air temperature, 2m dew point temperature, surface pressure, surface specific humidity, surface absorbed longwave radiation, surface incoming shortwave radiation, top-of-atmosphere incoming shortwave radiation, surface albedo, surface wind speed, surface roughness, zero-plane displacement height, measurement height of variables, and geopotential height.

Vegetation, surface albedo, and land cover information are obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) products. The MODIS normalized difference vegetation indices (NDVI) products, MOD13Q1 and MYD13Q1 version 6, are 250m/16-daily resolution [Didan et al., 2015a; 2015b]. The MODIS surface albedo products, MCD43A3 version 6, is a 500m/daily product [Schaaf and Wang, 2015]. The MODIS land cover, MCD12Q1, is a 500m/yearly product [Friedl and Sulla-Menashe, 2019]. Lastly, the global 90m Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) dataset are also used [Farr et al., 2007].

2.3. Validation Dataset

The ground-based precipitation measurements are collected from three networks providing daily cumulative precipitation (Figure *1*). The first is the Chinese Surface Stations for Global Exchange Version 3.0 product collected by the Chinese Meteorology Administrative (CMA). There are 19 stations in the study area, 9 of those, labeled purple, located within the Tarim basin, an endorheic basin with extremely scant precipitation due to the rain shadows of TP and the Tien Mountain. The other 10 rain gauges (in green) are scattered over the relatively high elevation area of the Ganges river basin, the Brahmaputra river basin, the Inner TP, and the Indus basin. Secondly, the Nepalese Department of Hydrology and Meteorology (DHM) has 7 stations located in the eastern Narayani basin and 4 in the southern Koshi basin, featuring heated tipping buckets (blue dots in Figure *1*). This area is influenced by the Indian monsoon rainfall and highly elevated topography. Thirdly, 7 stations from the Pakistan Meteorology Department (PMD) reside in the Karakoram area are also used (red dots in Figure *1*). 5 out of 7 are located on the Gilgit-Upper Indus river valley and 2 are located on the Central Karakorum National Park on Baltoro Glacier. This region is also characterized by a low amount of precipitation as the mountain ranges block the penetration of moist air from the Mediterranean and Caspian Seas during winter and spring and from the Indian Ocean during summer.

Two satellite-based precipitation datasets are also used for comparison with the downscaled precipitation product. The Climate Hazards Group InfraRed Precipitation (CHIRP) is a satellite-reanalysis product with 0.05°/daily resolution [Funk et al., 2015]. CHIRP uses a monthly precipitation climatology to adjust the global Thermal Infrared Cold Cloud Duration (CCD) rainfall estimates calibrated by the Tropical Rainfall Measuring Mission Multi-satellite Precipitation Analysis 3B42 version 7 (over 2000 to 2013) to produce pentadal precipitation estimates. CHIRP is further bias-corrected with rain gauge observations using a modified inverse distance weighting algorithm to produce the Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) product. As the last step, the daily CCD data and the Version 2 atmospheric model rainfall field from the National Oceanic and Atmospheric Administration (NOAA) Climate Forecast System (CFS) are used to disaggregate CHIRP and CHIRPS to daily resolution.

The NOAA Climate Prediction Center (CPC) morphing technique (CMORPH) product and its gauge-adjusted version are available at 0.072°/half-hourly resolution [Joyce et al., 2004]. CMORPH is produced by propagating passive microwave precipitation estimates backward and forward with the cloud system advection vectors generated based on the geostationary satellites infrared imagery at a half-hourly interval. The gauge-adjusted CMORPH is produced using the probability density function matching technique with the NOAA CPC Unified daily gauge analysis (CPCU) over land and the Global Precipitation Climatology Project over ocean [Xie et al., 2017].

The gauge-adjusted MERRA-2 precipitation product (0.5° × 0.625°/hourly), which is corrected by the CPCU gauge-based product, is also included for inter-comparison purposes [Reichle et al., 2017].

3. Methodology

The kernel of the proposed precipitation downscaling framework is a RF classification and a RF regression. The framework contains several components with different functionalities (Figure *2*). First, an RFE-based procedure is used to select relevant predictors. The selected predictors are either upscaled or topographically corrected to 1km and temporally aggregated/disaggregated to daily. A binary precipitation mask is produced based on daily cumulative precipitation rate greater than 0mm and the RF classification model is trained to the precipitation mask. The 1km/daily precipitation mask is estimated using the 1km predictors. Then, the RF regression model is trained and the 1km precipitation is estimated using the downscaled predictors over rainy pixels only. This framework is demonstrated using 3 years (2006–2008) of data over the study domain.

3.1. Potential Predictors

We consider a total of 13 predictors for precipitation, which include 8 atmospheric variables and 5 auxiliary variables. The atmospheric variables are air temperature, dew point temperature, air pressure, specific humidity, relative humidity, incident longwave radiation, incident shortwave radiation, and wind speed. The 7 atmospheric variables except for the relative humidity are collected from MERRA-2 and regridded to 50km. Relative humidity is calculated based on specific humidity, air temperature, and pressure. These 8 atmospheric variables are then adjusted for discrepancies from the 50km MERRA-2 geopotential height to a 1km terrain grids derived from SRTM to account for topographic effects from the 1km terrain. The topographic corrections contain two parts: i) a bilinear interpolation to match resolution of the 50km MERRA-2 variables to 1km, and ii) a deterministic downscaling to adjust for topographic effects [Rouf et al., 2019; Tao and Barros, 2018; Ruiz-Arias et al., 2010a; 2010b; Cosgrove et al., 2003]. Details on the topographic correction processes are provided in Appendix A.1, accompaned by a brief evaluation of the downscaled variables with another high-resolution reanalysis dataset in Appendix A.2. All 8 atmospheric variables, both their 50km and 1km version, are then averaged from hourly to daily.

We also include 5 auxiliary variables as potential predictors: 1- and 2-month-lagged vegetation, latitude, longitude, and date of year. Xu et al. [2015] showed that vegetation greenness could be used as a proxy for cumulative precipitation. We derive vegetation information from two MODIS NDVI products that have the same resolution, but different record date (MOD13Q1 is 8 days earlier than MYD13Q1). By combining the two, we obtain NDVI records every 8 days. The 8-daily NDVI is spatially aggregated to 50km and 1km resolution. The 1-(2-)month-lagged vegetation of a target day is the NDVI 30 (60) days after the day. Latitude/longitude and date of year are also used, as they contain geographical and seasonality information.

3.2. Selection of Predictors

To determine an optimal number of predictors that balances the model performance and computational costs, an RFE-based procedure is adopted and the concept of MDA is used to quantify the predictor importance. To investigate the consistency of relative predictor importance among different years, we perform the selection separately for the three study years. A 30% of the data points (sampled evenly from rainy and non-rainy pixels and every time step) is withheld for validation. A maximum number of 60 trees is grown and the performance of the model based on one to the maximum number of trees is also investigated. Other meta-parameters of RF that may impact the model performance and training costs, e.g., minimal size of leaf nodes, size of the random subset of variables for each decision split, and size of the bootstrap sample, are adopted from values suggested in the literature (Table 1; He et al., 2016; Genuer et al., 2010; Breiman, 2001).

First, the RF regression model is trained with the 13 variables presented in Section 3.1 and the logarithmic precipitation (P^*) as the predictand. Model performance is measured by the MSE of the out-of-bag sample, η:

$$\eta = \frac{1}{N_{oob}} \sum_{i=1}^{N_{oob}} \left(P_i^* - \frac{1}{N_T I_{i,j}} \sum_{j=1}^{N_T} \widehat{P_{i,j}^*}\, I_{i,j} \right)^2 \tag{1}$$

where $\widehat{P_{i,j}^*}$ denotes an estimate of P_i^* based on the out-of-bag sample using the j^{th} tree. $I_{i,j}$ denotes whether the i^{th} observation of the j^{th} tree is out-of-bag. N_{oob} is the union set of out-of-bag sample size of all trees. The predictor importance is measured by the MDA defined as [Breiman, 2001]:

$$\Delta_n = \frac{1}{N_T} \sum_{j=1}^{N_T} \left(\dot{\eta}_j | v_n - \dot{\eta}_j \right) \tag{2}$$

where Δ_n is the changes in out-of-bag MSE with the n-th variable perturbed $\dot{\eta}_j$ and $\dot{\eta}_j | v_n$ represent MSE of the original and perturbed out-of-bag sample of the n-th variable, respectively. $\dot{\eta}_j$ and $\dot{\eta}_j | v_n$ are defined similarly as Eq.(1):

$$\dot{\eta}_j = \frac{1}{N_{oob,j}} \sum_{i=1}^{N_{oob,j}} \left(P_i^* - \widehat{P_{i,j}^*} \right)^2 \tag{3}$$

$$\dot{\eta}_j | v_n = \frac{1}{N_{oob,j}} \sum_{i=1}^{N_{oob,j}} \left(P_i^* - \widehat{P_{i,j}^*} | v_n \right)^2 \tag{4}$$

where $N_{oob,j}$ is the out-of-bag sample size for the *j-th* tree. Then, all variables are ranked based on their Δ_n and the variable associated with the smallest Δ_n is removed. It is important to note that the variable ranking may change slightly as one repeats the training. To ensure a robust ranking, we follow the method proposed by Gregorutti et al., [2017] to repeat the training multiple times until the same variable appears three times with the smallest Δ_n. With one variable removed, this filtering process is repeated until there is only one variable left. η and Δ_n are recorded for different models and predictors.

The relative model performance is quantified by a normalized MSE (η in Eq.(1) divided by the variance of P^*) shown in Figure *3*. The left three panels show performance of models for the three years as a function of the number of predictors and trees. The performance stays relatively constant for models with 7 predictors or more. To have a better visualization of the effects of the tree number, the right panel shows that for the 7-predictor models the normalized MSEs are less than 0.12 with 50 trees. To evaluate the consistency of the 7 most relevant predictors among different years, Table *2* lists the variable removal order through the iterations. The variable in row 1 is the first variable to be removed, i.e., the least important predictor to precipitation. We observe consistency of the ranking – dew point temperature, specific humidity, and air temperature are always removed at the first three iterations, while relative humidity, pressure, day of year, and wind speed are always kept till the last four iterations. The 7 most important predictors for the three years are relative humidity, pressure, day of year, wind speed, shortwave radiation, longitude, and 2-month-lagged NDVI. Note that although longwave radiation ranks the 7^{th} instead of 2-month-lagged NDVI, which ranks the 8^{th}, in the case of 2007, the latter one is still considered for the sake of having a unified set of predictors. The cross-correlation of the 7 selected predictors and precipitation are assessed, showing correlation coefficients within ±0.5

for 27 out of the 28 cases (only pressure shows a mild correlation of 0.53 with 2-month-lagged NDVI).

3.3. Model Training and Precipitation Downscaling

Precipitation is highly intermittent in nature but this is often concealed when coarse resolutions are considered. Therefore, simple statistical interpolations can result in artificial boundaries that distort the rain shadow effects due to complex terrain. To resolve this issue, our framework creates 1km binary precipitation masks and then applies the mask to the 1km precipitation fields. The general idea is that, given a set of atmospheric and geospatial conditions, it is possible to infer whether the pixels are rainy or not.

First, the RF classification model for the binary precipitation mask, P_m, is trained separately for the 3 study years, using the 7 identified predictors. P_m is defined by P and it is 0 for pixels with no rain and 1 otherwise:

$$P_m = RFC(\Omega) \tag{5}$$

where $RFC(*)$ denotes the RF classification model and Ω denotes the selected predictor set. The next step estimates the 1km mask, $\tilde{P}_m$, quantifying whether the pixels are rainy or not, using the trained $RFC(*)$ with $\tilde{\Omega}$, the predictor set formed by the 1km variables.

Then, an RF regression is performed between the 7 selected variables and P^* of rainy pixel ($P_m=1$) for the training sample:

$$P^*|_{P_m=1} = RFR\left(\Omega|_{P_m=1}\right) \tag{6}$$

where $RFR(*)$ denotes the RF regression model. The downscaled precipitation ($\tilde{P}^*$) is estimated with $\tilde{\Omega}$ for rainy pixels indicated by $\tilde{P}_m$ equals 1 and a final step is to convert the logarithmic $\tilde{P}^*$ back to the actual precipitation rate (precipitation rate is set to 0 for pixels with $\tilde{P}_m$ equals 0).

Figure 4 shows snapshots of the 50km and 1km resolution precipitation mask for a single day. The 1km precipitation mask reveals similar locations of the non-rainy pixels with finer details, a rainy belt located on the south slope of HMA, which is not detectable in the 50km mask. Table 3 lists the size of the training (out-of-bag) and validation population with the model performance metrics. Values of the misclassification rate and normalized MSE for the out-of-bag and validation samples are similar to each other, suggesting stable performance of the models when a different dataset is considered.

3.4. Evaluation of the Downscaled Precipitation Product

The downscaled precipitation product is evaluated against ground- and satellite-based observations. For the comparison with the ground-based networks, the uncorrected and corrected MERRA-2, CMORPH, and CHIRPS are interpolated using the nearest neighbor to 1km and aggregated to daily, matching the downscaled precipitation. The use of nearest neighbor interpolation ensures the original precipitation magnitude of products. Time series of pixels collocated with rain gauges are extracted and error metrics are computed for the common time period. The Taylor's and performance diagrams are used to collectively assess the performance of the precipitation datasets [Roebber, 2009; Taylor, 2001]. The Taylor's diagram shows the normalized standard deviation (σ^*), the normalized centered root mean squared error (E^*), and the correlation coefficient (ρ) collectively in a single panel, allowing to investigate the dynamics among different error components. The differences in E^* and ρ derived by any combination of two precipitation datasets (with respect to the reference) are tested for statistical significance using the F-test and Z-test, respectively, with a significant level of 0.05 (Meng et al., 1992; Snedecor and Cochran, 1989). The performance diagram works complementarily to the Taylor's as it focuses on the detection-based error metrics namely, the probability of detection (*PoD*), false alarm ratio (*FAR*), bias ratio (*BR*), and critical success index (*CSI*) conditioned on an ad-hoc threshold. Rather than null, we set such threshold to 0.01mm/d to discern rain and no-rain. All statistics and error metrics used to produce the two diagrams are defined in Appendix B.

For the comparison with the satellite-based products, the downscaled precipitation is aggregated to 5km and 8km to match the spatial resolutions of CHIRPS and CMORPH, respectively. The CMORPH products are aggregated to daily. Because of the inherent bias between different precipitation products, we use ρ to quantify the similarity. Specifically, ρ for all locations in the 5km/8km terrain is comnputed between the aggregated downscaled precipitation and CHIRPS/CMORPH. Results are tested for statistical significance (to test whether the ρ values are greater than 0) using a Z-test with a significant level of 0.05 (Snedecor and Cochran, 1989).

4. Results

A qualitative assessment of the downscaled product with respect to its original resolution counterpart and to the reference products is presented in Figure 5. All products reveal a similar mean annual precipitation spatial pattern, but different magnitudes. This is particularly evident on the southern slope of the Himalayan Range, where MERRA-2 reaches values higher than 4,000mm/year, while the other products do not exceed 3,800mm/year. The gauge adjustments work similarly in reducing the overall precipitation magnitudes for the MERRA-2 family and CHIRP/CHIRPS, while the corrected CMORPH shows the opposite trend to its uncorrected counterpart. Downscaled MERRA-2 also reveals lower magnitudes than MERRA-2. The spatial distribution of downscaled MERRA-2 reflects the topographic and vegetation features, which are blended in the RF downscaling framework.

4.1. Validation Against Ground-Based Observations

Different precipitation datasets are evaluated against the ground observations using Taylor's diagrams (Figure 6). Results of the statistical significance tests for both normalized centered root mean square error and correlation coefficient are listed in Table *4*. Overall, values of E^* and ρ are generally lower and higher, respectively, in the evaluation against the DHM gauges in comparison to the others. This may point to an issue with light precipitation detection for the satellite and the reanalysis products, as the mean annual precipitation for the CMA-1, CMA-2, and PMD gauges are 41, 336, and 187mm/year, respectively, while that of the DHM gauges is 1526mm/year. Product-wise speaking, the MERRA-2 family is characterized by higher ρ than the satellite-based products in 3 out of the 4 networks (except for CMA-2), with the downscaled one showing the highest ρ at 0.52 and 0.23 over DHM and PMD, respectively. Yet, no single product can be claimed as the best in terms of E^*. For instance, CHIRP shows the lowest E^*s at 1.03 and 1.08 for CMA-1 and PMD, respectively; whereas the downscaled and corrected MERRA-2 take the lead in the DHM and CMA-2 cases, with E^*s at 0.89 and 0.98, respectively. Another consistent observation among the network is that the downscaling framework improves the accuracy of MERRA-2, dragging it to the corrected MERRA-2.

Precipitation detection is then evaluated through the performance diagrams presented in Figure 7. Overall, all products perform better in the DHM network and worse in the PMD one, which agrees with the Taylor's diagrams. All products, except for CHIRPS, show *BRs* greater than 1.38, indicating that false alarms are more severe than missed events. The MERRA-2 family is always characterized by higher *CSIs* compared to the satellite-based products. By comparing the gauge-corrected products to their uncorrected versions (filled dots vs. unfilled dots in the same colors), one can see improvements in their *BRs*, getting closer to 1. But this is not always the case for their *CSIs* as we only observe consistent increases for the MERRA-2 products. Results also suggest that the downscaling framework consistently improves both the bias ratio and the critical success index, making them closer to 1. This could be attributed to the extra 0-precipitation days introduced by the 1km precipitation mask, which simultaneously decreases *PoD* and *FAR*, but leads to an overall better detection. A detailed investigation on each panel shows that the downscaled MERRA-2 has the highest *CSI* at 0.24 over the CMA-1 networks and the second highest *CSIs* for the other three cases (0.35, 0.55, and 0.16 for CMA-2, DHM, and PMD).

Overall, the MERRA-2 family reveals higher correlation and detection consistency with the ground gauge networks, with the downscaled product showing improvements to the original MERRA-2, as it appears as the best product in the DHM (CMA-1) network comparison by means of correlation and random error (detection-based error). All products are generally better in capturing precipitation depicted by the DHM network, but their performance degrades when detecting light precipitation.

4.2. Comparison with Remote Sensing Products

Figure 8 displays the spatial distribution of correlation coefficient between the downscaled precipitation product and each satellite product. First, ρ between downscaled MERRA-2 and CHIRP/CHIRPS is higher than that of the downscaled MERRA-2 and CMORPH/corrected CMORPH. The downscaled product is more similar to CHIRP than its gauge-corrected version (Figure 8a and c); low values of ρ are generally clustered over most of the Tarim and areas around the borders of Indus and Luni basin. Several gray patches (negative ρ or value not significantly larger than 0) also appear within the Tarim basin in the CHIRPS case. For the two CMORPHs (Figure 8b and d), few regions present high ρ values; this includes portions of the western Helmand, the southwestern Luni, most of the coastal basins, and some patches in the Ganges basin. More gray patches are identified in these comparisons, especially over the borders of the Indus with the Amu Darya and Tarim, a mountainous region which consists of the Hindu Kush, Pamir, and Karakoram Mountain. Some gray patches also appear within the Inner TP and on the borders of Ganges and Brahmaputra (the eastern Himalaya range). The correlation maps show results that are in line with the ones in Figure 6 and Figure 7, i.e., the two CMORPHs are always closer to each other than CHIRP to CHIRPS.

To summarize, it is clear that downscaled MERRA-2 has a higher degree of similarity in terms of correlations with the CHIRP and CHIRPS products, especially the near-real-time one, in comparisons to the two CMORPHs. Low correlations are consistently found over the Tarim, southern Indus, and northern Luni basins in all cases. Negative correlation appears over the Hindu Kush-Pamir-Karakoram Mountain region, the eastern Himalaya range, and some patches within the Inner TP in the comparisons to CMORPHs.

5. Discussions

5.1. On the Selection of Predictors

Topography is an important factor that modulates the precipitation distribution in such a mountainous terrain. Our results indicate that the downscaled precipitation preserves the mesoscale features of the MERRA-2 precipitation and inherits the fine-scale features from topography. This is because the proposed RF-based downscaling framework considers 4 near-surface meteorological variables (i.e., air pressure, relative humidity, shortwave radiation, and wind speed) that are corrected for slope, aspect, curvature, shadowing effects, sky obstruction, and reflection to account for the topographic effects. To avoid repeating information, those topographic variables are not explicitly considered as predictors to train the RF models. Moreover, this study investigates the relative predictor importance of 13 variables and selects 7 of those to train the RF models for precipitation downscaling. These variables include surface meteorology, geospatial information, vegetation cover, and seasonality of the region. In addition, physical properties of cloud, soil moisture, and other land surface parameters may also be included in the precipitation downscaling. For example, the cloud optical thickness effective radius and cloud water path are found to be related to precipitation rate [Sharifi et al., 2019] and soil moisture can be used to infer precipitation through an inversed water balance equation [Brocca et al., 2014]. Future studies should focus on the potential of those variables in the downscaling framework.

5.2. On the Similarity to Satellite Products

Our results suggest that the downscaled precipitation product has a high degree of linear agreement with CHIRP. Downscaled MERRA-2 inherits the temporal variability of MERRA-2, which is produced by the Goddard Earth Observing System version 5.12.4 (GEOS v5.12.4) and CHIRP utilizes the CFS v2 precipitation to disaggregate the pentadal 3B42-calibrated CCD precipitation estimates [Gelaro et al., 2017; Funk et al., 2015]. GEOS v5.12.4 and CFS v2 are similar in two ways. They both use a three-dimensional variational data assimilation analysis algorithm based on the Gridpoint Statistical Interpolation scheme and they assimilate observations from common sources as conventional ground-based observations of standard atmospheric variables, radiosondes and pibals, aircraft data, satellite-derived wind, radio occultation data and satellite radiance [Gelaro et al., 2017; Reichle et al., 2017; Saha et al., 2010]. Therefore, our results highlight these underlying similarities between the two reanalysis systems used to produce MERRA-2 and CHIRP.

The near-real-time CMORPH is a satellite-only product whose temporal variability is based on the motion vectors of cloud systems derived from the consecutive geosynchronous earth orbit infrared images. This explains in part the lower agreement with the downscaled product. Additionally, the low correlation can be attributed to the proportion of missing values in CMORPH. The current version of CMORPH has no procedures to gap-fill the snow-covered surface, leading to incomplete spatial coverage during cold seasons [Xie et al., 2017]. This is substantiated by Figure 9, which shows the fraction of missing values during the winter season (December to February) for the study domain. The fraction of missing values reaches almost 50% over the Karakoram Mountain region coincident with the area with negative correlation coefficients revealed by Figure 8b and d.

6. Conclusions

In this study, we developed a nonparametric precipitation downscaling framework based on the RF algorithm for HMA. The proposed framework includes a unique recursive feature elimination procedure for predictor selection. It utilizes i) an RF classification to develop a high-resolution precipitation mask and ii) an RF regression to spatially downscale the precipitation rate over rainy pixels. The RF models are separately built for 3 years, 2006, 2007, and 2008. The use of the 7 selected variables as predictors is demonstrated to be sufficient to provide stable and accurate performance for the RF regression models. The downscaled precipitation product is validated against four ground-based rain gauge networks and is compared to four widely used satellite precipitation products.

Results suggest that the downscaled precipitation preserves the mesoscale features of the MERRA-2 precipitation, while also inheriting the topographic features of the downscaled atmospheric variables and vegetation indices. The ground-based validation results suggest consistent improvements, regardless of the precipitation magnitude or detection, after downscaling the original MERRA-2. The downscaled product outperforms others over the DHM network, while reaches a similar level of performance in the CMA-1 and PMD ones in terms of root mean square error and correlation coefficient. In terms of detection, results suggest that false alarms are more severe than missed events; the MERRA-2 family precipitation always show better critical success indices than the satellite-based products.

The downscaled precipitation product reveals higher similarities in terms of correlation with CHIRP and CHIPRS than the two CMORPHs, especially with CHIRP. Correlations are generally lower over the Tarim basin and parts of the Indus and Luni basins for the CHIRP and CHIRPS cases. For CMORPH, most of the study area are characterized by low correlation, particularly the region of the Hindu Kush, Pamir, and Karakoram Mountains.

In conclusion, the developed precipitation downscaling framework may alleviate the urgent need of high-resolution surface meteorological and climatological data for environmental modeling over the HMA area. Given the low density of the ground-based meteorological network, future studies should focus on indirect validation methods. For instance, the potential of using the downscaled products as input to a hydrological model over basins in HMA could be assessed if streamflow gauges were available in the region. In addition, the application of the framework to other precipitation products might be tested.

Appendix A. Topographic correction of atmospheric variables

A.1. Topographic correction procedures

The topographic correction processes are designed to spatially downscale the 8 hourly atmospheric variables used as potential predictors for precipitation from 50km to 1km. In this appendix, we show only the necessary processes and evaluation results for the 4 selected variables (pressure, relative humidity, shortwave radiation, and wind speed).

The topographic correction method for air pressure (p, Pa) accounts for the pressure difference between the 50km and the 1km terrain elevation. It is based on the hydrostatic equation and the Ideal Gas Law [Cosgrove et al., 2003]:

$$\tilde{p} = pe^{-\frac{g(\tilde{Z}-Z)}{RT_m}} \tag{A.1}$$

where variable with/without "~" indicates variable for the 50km/1km terrain. Z is the terrain elevation (m above sea level). R is the ideal gas constant (287J/kg·K) and g is the gravitational acceleration (9.81m/s^2). T_m is the mean air temperature between the 50km and the 1km terrain elevation, i.e., $\frac{\tilde{T}+T}{2}$, with T representing the air temperature (K). $\tilde{T}$ is adjusted from T based on the lapse rate correction:

$$\tilde{T} = T + \Gamma\left(\tilde{Z} - Z\right) \tag{A.2}$$

where Γ is a dynamic temperature lapse rate estimated as the slope of regressing temperature and elevation difference of a location to its eight neighbors in space for a time step [Rouf et al., 2019].

Relative humidity is defined as the ratio between actual and saturated mixing ratio of water vapor; by recognizing that mixing ratio of water vapor may be further expressed by vapor pressure and air pressure, one arrives at the following equation for relative humidity ($\tilde{r}$, %):

$$\tilde{r} = \frac{\tilde{E}/(\tilde{p} - \tilde{E})}{\widetilde{E_s}/(\tilde{p} - \widetilde{E_s})} \times 100 \quad \text{(A.3)}$$

where $\tilde{E}$ and $\tilde{E}_s$ are actual and saturated vapor pressure (Pa), which, in terms, are related to dew point and air temperature by the Magnus formula (see, for example, Lawrence [2005]):

$$\tilde{E} = Ce^{\left(\frac{A\widetilde{T_d}}{\widetilde{T_d}+B}\right)} \quad \text{(A.4)}$$

where constant A, B, and C are 17.368/22.452, 238.88°C/272.55°C, and 611.21Pa/611.15Pa for water/ice surface. $\tilde{T}_d$ is the dew point temperature and, by replacing which with $\tilde{T}$, one arrives with E_s. $\tilde{T}_d$ is adjusted from T_d using Eq.(A.2) by replacing Γ with Γ_d, a dynamic lapse rate for dew point temperature similarly determined as the air temperature lapse rate [Rouf et al., 2019].

Wind speed (W, m/s) is adjusted for friction velocity under the logarithmic wind profile assumption:

$$W = \frac{U_*}{\kappa} \ln\left(\frac{H - h_0}{z_0}\right) \quad \text{(A.5)}$$

where κ is the Von Kármán constant (~0.41). U_* is friction velocity (m/s); z_0, h_0, and H are surface roughness, zero-plane displacement height, and measurement height (m above ground). Note that wind speed for the 1km terrain, $\tilde{W}$, may be expressed by Eq.(A.5) with the corresponding variables for the 1km terrain. By taking the ratio between $\tilde{W}$ and W, one arrives at the following:

$$\widetilde{W} = W \frac{\widetilde{U_*}}{U_*} \quad \text{(A.6)}$$

Note that the differences between the 1km and 50km $\ln\left(\frac{H-h_0}{z_0}\right)$ term may be neglected as we consider the same measurement height for the terrains. To find $\tilde{W}$, one would need $\tilde{U}_*$. Tao and Barros [2018] reveals that $\frac{\widetilde{U_*}}{U_*}$ may be approximated by $\left(\frac{\widetilde{z_0}}{z_0}\right)^{0.09}$ as one considers the dependence of the geostrophic drag coefficient on surface roughness and the geostrophic wind remains unchanged with the two terrain scales. The variable z_0 is adopted from MERRA-2 while $\tilde{z}_0$ may be determined if one considers its dependency on land cover [Bohn and Vivoni, 2019]. Bohn and Vivoni [2019] provides a surface roughness look-up-table conditional on land cover classes and months. We adopt the table and implement it over the 500m/yearly MODIS land cover product to construct a 1km/monthly surface roughness, named as $\tilde{z}_{LC}$. A temporal disaggregation factor is derived from the hourly MERRA-2 z_0 and then multiple to $\tilde{z}_{LC}$ for $\tilde{z}_0$:

$$\widetilde{z_0} = \widetilde{z_{LC}} \frac{z_0}{\overline{z_0}} \tag{A.7}$$

where $\bar{z}_0$ is the monthly mean of z_0.

The topographic correction of incident shortwave radiation (S, W/m^2) is separated for different components, considering the differences in optical path length of sunlight, the terrain shadowing effects, the openness of terrain, and the reflecting from ambient terrain [Ruiz-Arias et al., 2010a; 2010b]. It comprises four steps. At first, S is partitioned into beam (S_b) and diffuse radiation (S_d) based on the Ruiz-Arias et al. [2010a] regression model. Then, S_b and S_d are adjusted by the following:

$$\widetilde{S_b} = S_b e^{k(\tilde{p}-p)} cos(\theta)\, \delta \tag{A.8}$$

$$\widetilde{S_d} = S_d F_v \tag{A.9}$$

where k (Pa^{-1}) is the broadband attenuation coefficient defined as the top-of-atmosphere and surface radiation difference over pressure difference [Rouf et al., 2019]; the term $e^{k(\tilde{p}-p)}$ is used to account for the differences in optical path length due to the pressure difference. The cosine of solar illumination angle, $cos(\theta)$, ranging between -1 to 1, indicates if the sun is below or above the local horizon (note that values lower than 0 are set to 0); and δ is a binary shadow mask indicating whether the location is blocked by the surrounding terrain. These two factors account for the self- and cast-shadowing caused by the local slope and surrounding terrain. Term F_v in Eq.(A.9) is the sky-view factor accounting for the sky obstruction. In step 3, a reflected radiation component $\tilde{S}_r$ is estimated:

$$\widetilde{S_r} = \tilde{A} F_t \left[\widetilde{S_b} + (1 - F_v) \widetilde{S_d} \right] \tag{A.10}$$

where A is the surface albedo adopted from MODIS; F_t is the terrain configuration factor. Lastly, the sum of these three components gives the 1km incident shortwave radiation [Ruiz-Arias et al., 2010b].

A.2. Pattern-based Comparisons

The topographically-corrected atmospheric variables are compared to the High Asia Refined-Analysis (HAR) product, which is developed by dynamical-downscaling of global analysis data using the Weather Research and Forecasting model [Maussion et al., 2014]. HAR is an hourly atmospheric dataset generated primarily for TP with a high spatial resolution at 10km. The aim here is to provide a qualitative assessment on whether the topographic correction procedures introduce reasonable spatial features to the coarse resolution variables by comparing to the dynamical downscaling applied by HAR. The 1km topographically-corrected variables are spatially aggregated to 10km to match with the HAR ones. Both datasets are temporally aggregated to daily for the comparison as daily is the time scale the 1km variables were used to produce the downscaled precipitation. Three error metrics – mean relative error (ε), normalized centered root mean square error (E^*), and correlation coefficient (ρ) – are produced to assess the consistency between the datasets. E^* and ρ are used earlier in the Taylor's diagrams; ε is used to quantify the systematic difference (see Appendix B for definitions).

Figure A 1 shows the error metrics for the four selected atmospheric variables based on the three study years. Note that since the relative humidity is not directly available from HAR, the specific humidity is compared instead. Looking at the three panels from the first row, we can see a high consistency between air pressure. Majority of the locations have positive ε, ranging from 0.02 to 0.03 (the greenish color), indicating a very slight overestimation to the HAR air pressure by the topographic corrected one. Locations characterized by larger discrepancies are those located on the southern and northern slope of TP. For specific humidity, high ρs (larger than 0.95 for most of the locations) can still be observed while the magnitudes of ε and E^* increase compared to the air pressure case. Patterns of ρ and E^* generally agree with each other, revealing lower E^*s and higher ρs along the southern slope of TP; yet, magnitudes of ε over the southern slope are larger. Moving on to the shortwave radiation, its ε map suggests a slight underestimation as most of the pixels are in blueish color. The overall magnitude of ε is within ±0.1 for the majority locations. The ρ and E^* maps of this case suggest that a large portion of the southern part of the domain (areas in gray) are characterized by E^*s higher than 0.6 and ρs lower than 0.8. The case of wind speed shows the highest differences compared to the other three variables (note the enlarged color bar scales for the bottom three panels). The two wind speed datasets are more similar over TP but reveal relatively large differences over the northern and southern slope of TP.

To sum up, patterns of the topographically-corrected variables are comparable to the HAR ones, especially for the air pressure and specific humidity cases. The overall magnitudes of shortwave radiation are similar as revealed by the low systematic differences, but their distributions can be quite different on the southern part of the domain. Wind speed shows some major differences over the domain as marked by the gray areas.

Appendix B. Definitions of Statistics and Error Metrics

The Taylor's and performance diagrams are used in this study to quantify the discrepancies between different datasets. Both diagrams contain multiple statistics and error metrics; their definitions are listed in Table A 1. Note that x_i and y_i represent the estimation and the reference of a variable, respectively; K is the total number of data pairs. Term H, M, and F represent the number of hit, missing, and false alarm of precipitation conditioned on the 0.01mm/d threshold. Bolded numbers shown in the value range column indicate the idle performance.

Acknowledgments and Data Availability

The work was supported by NASA Grant NNX16AQ89G. The authors would like to thank the entire HiAMT team for the useful discussion and in particular the two teams led by Sujay Kumar and Kyle McDonald. The authors are also grateful to Yagmur Derin for sharing the ground-based data. The experiments of this study were run on ARGO, a research computing cluster provided by the Office of Research Computing at George Mason University, VA (http://orc.gmu.edu). The downscaling framework is implemented by MatLab functions available via Mei's GitHub profile (https://github.com/YiwenMei/AtmDS and https://github.com/YiwenMei/PrecipDS).

The MERRA-2 reanalysis product may be downloaded from https://disc.gsfc.nasa.gov/datasets. The MOD13Q1 and MYD13Q1 products are available from https://lpdaac.usgs.gov/products/mod13q1v006/ and https://lpdaac.usgs.gov/products/myd13q1v006/, respectively. The MCD12Q1 product may be downloaded from https://lpdaac.usgs.gov/products/mcd12q1v006/. The MCD43A3 product is available from https://lpdaac.usgs.gov/products/mcd43a3v006/. The SRTM elevation data may be downloaded from http://srtm.csi.cgiar.org/srtmdata/. The Chinese Surface Stations for Global Exchange Version 3.0 product may be downloaded from https://data.cma.cn/data/cdcdetail/dataCode/SURF_CLI_CHN_MUL_DAY_CES_V3.0.html. Other ground-based precipitation datasets may be available upon request to the corresponding author. The CHIPRS precipitation data are available from ftp://ftp.chg.ucsb.edu/pub/org/chg/products/. The CMORPH precipitation products may be downloaded from ftp://ftp.cpc.ncep.noaa.gov/precip/.

References

Ali, I., Greifeneder, F., Stamenkovic, J., Neumann, M., & Notarnicola, C. (2015). Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data. *Remote. Sens., 7*(12), 16398-16421.

Bhuiyan, M. A. E., Nikolopoulos, E. I., Anagnostou, E. N., Quintana-Seguí, P., & Barella-Ortiz A. (2018). A nonparametric statistical technique for combining global precipitation datasets: development and hydrological evaluation over the Iberian Peninsula. *Hydrol. Earth Syst. Sci., 22*(2), 1371-1389.

Bohn T. J. & Vivoni, E. R. (2019). MOD-LSP, MODIS-based parameters for hydrologic modeling of North American land cover change. *Sci. Data, 6*(1), 144.

Breiman, L. (2001). Random Forests. *Mach. Learn., 45*(1), 5-32.

Brocca, L., Ciabatta, L., Massari, C., Moramarco, T., Hahn, S., Hasenauer, S. et al. (2014). Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data. *J. Geophys. Res. Atmos., 119*(9), 5128-5141.

Cannon, F., Carvalho, L. M. V., Jones, C., Norris, J., Bookhagen, B., & Kiladis, G. N. (2017). Effects of topographic smoothing on the simulation of winter precipitation in High Mountain Asia. *J. Geophys. Res. Atmos., 122*(3), 1456-1474.

Chen, R., Liu, J., Kang, E., Yang, Y., Han, C., Liu, Z. et al. (2015). Precipitation measurement intercomparison in the Qilian Mountains, north-eastern Tibetan Plateau. *The Cryosphere, 9*(5), 1995-2008.

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C. et al. (2003). Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *J. Geophys. Res., 108*(D22), 8842.

Degenhardt, F., Seifert, S. & Szymczak, S. (2017). Evaluation of variable selection methods for random forests and omics data sets. *Brief. Bioinform.,* 1-12.

Didan, K. (2015a). MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. doi: 10.5067/MODIS/MOD13Q1.006

Didan, K. (2015b). MYD13Q1 MODIS/Aqua Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. doi: 10.5067/MODIS/MYD13Q1.006

Díaz-Uriarte, R. & Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics, 7*(1), 3.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S. et al., (2007). The Shuttle Radar Topography Mission. *Rev. Geophys., 45*(2), RG2004.

Friedl, M., Sulla-Menashe, D. (2019). MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. doi: 10.5067/MODIS/MCD12Q1.006

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S. et al. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci. Data, 2*, 150066.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L. et al. (2017). Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Climate, 30*(14), 5419-5454.

Genuer, R., Poggi, J.-M. & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recogn. Lett., 31*(14), 2225-2236.

Goetz, J. N., Brenning, A., Petschko, H. & Leopold, P. (2015). Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput Geosci, 81*, 1-11.

Gou, Y., Ma, Y., Chen, H. & Wen, Y. (2018). Radar-derived quantitative precipitation estimation in complex terrain over the eastern Tibetan Plateau. *Atmos. Res., 203*, 286-297.

Gregorutti, B., Michel, B. & Saint-Pierre, P. (2017). Correlation and variable importance in random forests. *Stat. Comput., 27*(3), 659-678.

Grömping, U. (2009). Variable Importance Assessment in Regression: Linear Regression versus Random Forest. *Am. Stat., 63*(4), 308-319.

Haylock, M. R., Cawley G. C., Harpham, C., Wilby R. L., Goodess, C. M. (2006). Downscaling heavy precipitation over the United Kingdom: a comparison of dynamical and statistical methods and their future scenarios. *Int. J. Climatol., 26*(10), 1397-1415.

Heung, B. Ho, H. C., Zhang, J., Knudby, A., Bulmer, C. E., Schmidt, M. G. (2015). An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma, 265*, 62-77.

He, X., Chaney, N. W., Schleiss, M. & Sheffield, J. (2016). Spatial downscaling of precipitation using adaptable random forests. *Water Resour. Res., 52*(10), 8217-8237.

Ibarra-Berastegi, G., Saénz, J., Ezcurra, A., Elías, A., Diaz Argandoña, J., & Errasti1 I. (2011). Downscaling of surface moisture flux and precipitation in the Ebro Valley (Spain) using analogues and analogues followed by random forests and multiple linear regression. *Hydrol. Earth Syst. Sci., 15*(6), 1895-1907.

Joyce, R. J., Janowiak, J. E., Arkin, P. A. & Xie, P. (2004). CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. *J. Hydrometeor., 5*(3), 487-503.

Lawrence, M. G. (2005). The Relationship between Relative Humidity and the Dewpoint Temperature in Moist Air: A Simple Conversion and Applications. *Bull. Amer. Meteor. Soc., 86*(2), 225-234.

Maraun, D. Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann M. et al. (2010). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys., 48*(3), RG3003.

Maussion, F., Scherer, D., Mölg, T., Collier, E., Curio, J., Finkelnburg, R. (2013). Precipitation Seasonality and Variability over the Tibetan Plateau as Resolved by the High Asia Reanalysis. *J. Climate*, *27*(5), 1910-1927.

Ma, Y., Hong, Y., Chen, Y., Yang, Y., Tang, G., Yao, Y. et al. (2018a). Performance of Optimally Merged Multisatellite Precipitation Products Using the Dynamic Bayesian Model Averaging Scheme Over the Tibetan Plateau. *J. Geophys. Res. Atmos., 123*(2), 814-834.

Ma, Y., Yang, Y., Han, Z., Tang, G., Maguire, L., Chu, Z. et al. (2018b). Comprehensive evaluation of Ensemble Multi-Satellite Precipitation Dataset using the Dynamic Bayesian Model Averaging scheme over the Tibetan plateau. *J. Hydol., 556,* 634-644.

Ma, Z. He, K., Tan, X., Xu, J., Fang, W., He, Yu. et al. (2018c). Comparisons of Spatially Downscaling TMPA and IMERG over the Tibetan Plateau. *Remote Sens., 10*(12), 1883.

Naghibi, S. A. & Pourghasemi, H. R. (2015). A Comparative Assessment Between Three Machine Learning Models and Their Performance Comparison by Bivariate and Multivariate Statistical Methods in Groundwater Potential Mapping. *Water Resour. Manag., 29*(14), 5217-5236.

Reichle, R. H., Liu, Q., Koster, R. D., Draper, C. S., Mahanama, S. P. P., Partyka, G. S. (2017). Land Surface Precipitation in MERRA-2. *J. Climate, 30*(5), 1643-1664.

Roebber, P. J. (2009). Visualizing Multiple Measures of Forecast Quality. *Wea. Forecasting, 24*(2), 601-608.

Rouf, T., Mei, Y., Maggioni, V., Houser, P. & Noonan, M. (2019). A Physically-based Downscaling Technique for a Set of Atmospheric Variables. *J. Hydrometeor., 21*(1), 93-108.

Ruiz-Arias, J. A., Alsamanra, H., Tovar-Pescador, J. & Pozo-Vázquez, D. (2010a). Proposal of a regressive model for the hourly diffuse solar radiation under all sky conditions. *Energ. Convers. Manage., 51*(5), 881-893.

Ruiz-Arias, J. A., Cebecauer, T., Tovar-Pescador, J. & Šúri, M. (2010b). Spatial disaggregation of satellite-derived irradiance using a high-resolution digital elevation model. *Sol. Energy, 84*(9), 1644-1657.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S. et al. (2010). The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc., 91*(8), 1015-1058.

Schaaf, C., Wang, Z. (2015). MCD43A3 MODIS/Terra+Aqua BRDF/Albedo Daily L3 Global - 500m V006 [Data set]. NASA EOSDIS Land Processes DAAC. doi: 10.5067/MODIS/MCD43A3.006

Sharifi, E., Saghafian, B. & Steinacker, R. (2019). Downscaling Satellite Precipitation Estimates With Multiple Linear Regression, Artificial Neural Networks, and Spline Interpolation Techniques. *J. Geophys. Res. Atmos.*, *124*(4), 789-805.

Shortridge, J. E., Guikema, S. D. & Zaitchik, B. F. (2016). Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrol. Earth Syst. Sci.*, *20*(7), 2611-2628.

Tao, J. & Barros, A. P. (2018). Multi-year atmospheric forcing datasets for hydrologic modeling in regions of complex terrain – Methodology and evaluation over the Integrated Precipitation and Hydrology Experiment 2014 domain. *J. Hydrol.*, *567*, 824-842.

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, *106*(D7), 7183-7192.

Tong, K., Su, F., Yang, D. & Hao, Z. (2014a). Evaluation of satellite precipitation retrievals and their potential utilities in hydrologic modeling over the Tibetan Plateau. *J. Hydrol.*, *519*, Part A, 423-437.

Tong, K., Su, F., Yang, D., Zhang, L., & Hao, Z. (2014b). Tibetan Plateau precipitation as depicted by gauge observations, reanalyses and satellite retrievals. *Int. J. Climatol.*, *34*(2), 265-285.

Wei, W., Zhang, R., Wen, M. & Yang, S. (2016). Relationship between the Asian Westerly Jet Stream and Summer Rainfall over Central Asia and North China: Roles of the Indian Monsoon and the South Asian High. *J. Climate*, *30*(2), 537-552.

Xie, P., Joyce, R., Wu, S., Yoo, S.-H., Yarosh, Y., Sun, F. et al. (2017). Reprocessed, Bias-Corrected CMORPH Global High-Resolution Precipitation Estimates from 1998. *J. Hydrometeor.*, *18*(6), 1617-1641.

Xu, S., Wu C., Wang, L., Gonsamo, A., Shen, Y., & Niu, Z. (2015). A new satellite-based monthly precipitation downscaling algorithm with non-stationary relationship between precipitation and land surface characteristics. *Remote Sens. Environ.*, *162*, 119-140.

Zorzetto, E., & Marani M. (2019). Downscaling of Rainfall Extremes FromSatellite Observations. *Water Resour. Res.*, *55*(1), 156-174.

List of Tables

Table 1. RF meta-parameter values for model training.

Table 2. Order of removal of potential predictor variables. Bold text shows variables identified as important predictors for precipitation downscaling.

Table 3. Verification of the RF classification and regression models.

Table 4. Results of the significance tests for E^* and ρ. The upper/lower portions of the table are for the E^*/ρ tests. A value of 1 indicates statistically significant differences are found between the two specified precipitation data in terms of E^* or ρ at a significance level of 0.05.

Table A 1. Definitions of statistics and error metrics used in the study.

Table 1. RF meta-parameter values for model training.

Parameter	Value
Number of predictors	[1 13]
Number of trees	[1 60]
Size of minimum leaf node	10
Size of the random subset for each decision split	One third (square root) of number of predictors for regression (classification)
Size of the bootstrap sample	36.8% of the training sample

Table 2. Order of removal of potential predictor variables. Bold text shows variables identified as important predictors for precipitation downscaling.

		2006	2007	2008
Order of parameter removal	1	Dew point temperature	Dew point temperature	Dew point temperature
	2	Specific humidity	Specific humidity	Specific humidity
	3	Air temperature	Air temperature	Air temperature
	4	1-month-lagged NDVI	1-month-lagged NDVI	Latitude
	5	Latitude	Latitude	Longwave radiation
	6	Longwave radiation	**2-month-lagged NDVI**	1-month-lagged NDVI
	7	**2-month-lagged NDVI**	Longwave radiation	**Longitude**
	8	**Longitude**	**Shortwave radiation**	**2-month-lagged NDVI**
	9	**Shortwave radiation**	**Longitude**	**Shortwave radiation**
	10	**Wind speed**	**Wind speed**	**Wind speed**
	11	**Day of year**	**Day of year**	**Day of year**
	12	**Air pressure**	**Air pressure**	**Air pressure**
	13	**Relative humidity**	**Relative humidity**	**Relative humidity**

Table 3. Verification of the RF classification and regression models using the misclassification rate (MCR) and normzlized MSE.

Sample	Year	RF classification		RF regression	
		Size	MCR (%)	Size	Normalized MSE ($\times10^{-2}$)
Out-of-bag	2006	141,850	3.94	130,788	11.96
	2007	142,379	4.52	124,800	11.99
	2008	142,714	4.25	128,184	11.94
Validation	2006	165,759	3.89	152,566	11.56
	2007	166,374	4.38	145,660	11.60
	2008	166,836	4.22	149,586	11.81

Table 4. Results of the significance tests for the normalized centered root mean square error (E^*) and correlation coefficient (ρ) in the Taylor's diagrams of Figure 6. The upper/lower portions of the table are for the E^*/ρ tests. A value of 1 indicates statistically significant differences are found between the two specified precipitation data in terms of E^* or ρ at a significance level of 0.05.

CMA-1 / CMA-2 DHM / PMD	MERRA-2	Corrected MERRA-2	Downscaled MERRA-2	CMORPH	Corrected CMORPH	CHIRP	CHIRPS
MERRA-2		1 / 1 1 / 1	1 / 1 1 / 1	1 / 1 1 / 1	0 / 1 1 / 1	1 / 1 1 / 1	1 / 1 1 / 1
Corrected MERRA-2	1 / 1 1 / 0		0 / 1 1 / 1	1 / 1 1 / 1	1 / 1 1 / 1	1 / 1 1 / 1	0 / 1 1 / 1
Downscaled MERRA-2	1 / 1 1 / 0	1 / 1 1 / 0		1 / 1 1 / 1	1 / 1 1 / 1	1 / 1 1 / 1	0 / 1 1 / 1
CMORPH	1 / 1 1 / 1	1 / 1 1 / 1	1 / 1 1 / 1		0 / 1 1 / 1	1 / 1 0 / 1	1 / 1 1 / 1
Corrected CMORPH	1 / 0 1 / 1	1 / 1 1 / 1	1 / 1 1 / 1	0 / 1 1 / 0		1 / 1 1 / 1	1 / 1 1 / 1
CHIRP	1 / 1 0 / 0	1 / 1 1 / 0	1 / 1 1 / 1	1 / 1 1 / 1	1 / 1 1 / 1		1 / 1 1 / 1
CHIRPS	1 / 1 1 / 1	1 / 1 1 / 1	1 / 1 1 / 1	0 / 1 0 / 1	0 / 1 1 / 1	1 / 1 1 / 1	

Table A 1. Definitions of statistics and error metrics used in the study.

Name	Definition	Value range
Mean	$\mu_X = \frac{1}{K}\sum_{i=1}^{K} x_i$	\
Variance	$\sigma_Y^2 = \frac{1}{K}\sum_{i=1}^{K} (y_i - \mu_Y)^2$	\
Covariance	$c_{XY} = \frac{1}{K}\sum_{i=1}^{K} (x_i - \mu_X)(y_i - \mu_Y)$	\
Mean relative error	$\epsilon = \frac{\mu_X - \mu_Y}{\mu_Y}$	(-∞, **0**, +∞)
Normalized standard deviation	$\sigma^* = \frac{\sigma_X}{\sigma_Y}$	(0, **1**, +∞)
Normalized centered root mean square error	$E^* = \frac{\sqrt{\frac{1}{K}\sum_{i=1}^{K}(x_i - y_i - \mu_X - \mu_Y)^2}}{\sigma_Y}$	[**0**, +∞)
Correlation coefficient	$\rho = \frac{c_{XY}}{\sigma_X \sigma_Y}$	[-1, **1**]
Probability of detection	$PoD = \frac{H}{H + M}$	[0, **1**]
False alarm ratio	$FAR = \frac{F}{H + F}$	[**0**, 1]
Bias ratio	$BR = \frac{H + F}{H + M}$	(0, **1**, +∞)
Critical success index	$CSI = \frac{H}{H + M + F}$	[0, **1**]

List of Figures

Figure 1. The High Mountain Asia region and locations of precipitation stations.

Figure 2. Flow chart of the precipitation downscaling framework.

Figure 3. Model performance in terms of the normalized mean square error with number of predictors and trees for the three study years.

Figure 4. Comparison between 50km and 1km precipitation masks of January 1st 2007.

Figure 5. Mean annual precipitation during 2006 to 2008 for different precipitation datasets.

Figure 6. Taylor's diagrams for different precipitation products vs. a) the CMA-1, b) the CMA-2, c) the DHM, and d) the PMD gauge networks. The number of gauges used to plot the diagrams are shown in parenthesis in the panel titles.

Figure 7. Same as Figure 6, but for performance diagrams. Solid curves represent CSI while the dash lines represent BR. A precipitation rate of 0.01mm/d is used as the rain/no-rain threshold.

Figure 8. Spatial distributions of ρ derived for the downscaled MERRA-2 with each of the satellite precipitation product for 2006 to 2008 (locations with value that is not significantly larger than 0 are in gray). The major river basins boundaries are superimposed.

Figure 9. Fraction of missing values in CMORPH products during the December to February period. Area in white indicates no missing values. The major river basins boundaries are superimposed.

Figure A 1. Spatial distributions of error metrics derived for the topographically-corrected atmospheric variables with the corresponding HAR ones. Areas in gray indicate that the ε, E*, or ρ are not significantly within ±0.2, lower than 0.6, or larger than 0.8 (those thresholds

change to ±0.5, 1, or 0.6 for the case of wind speed). Note that the comparison domain is smaller than the precipitation downscaling domain given the smaller coverage of HAR.

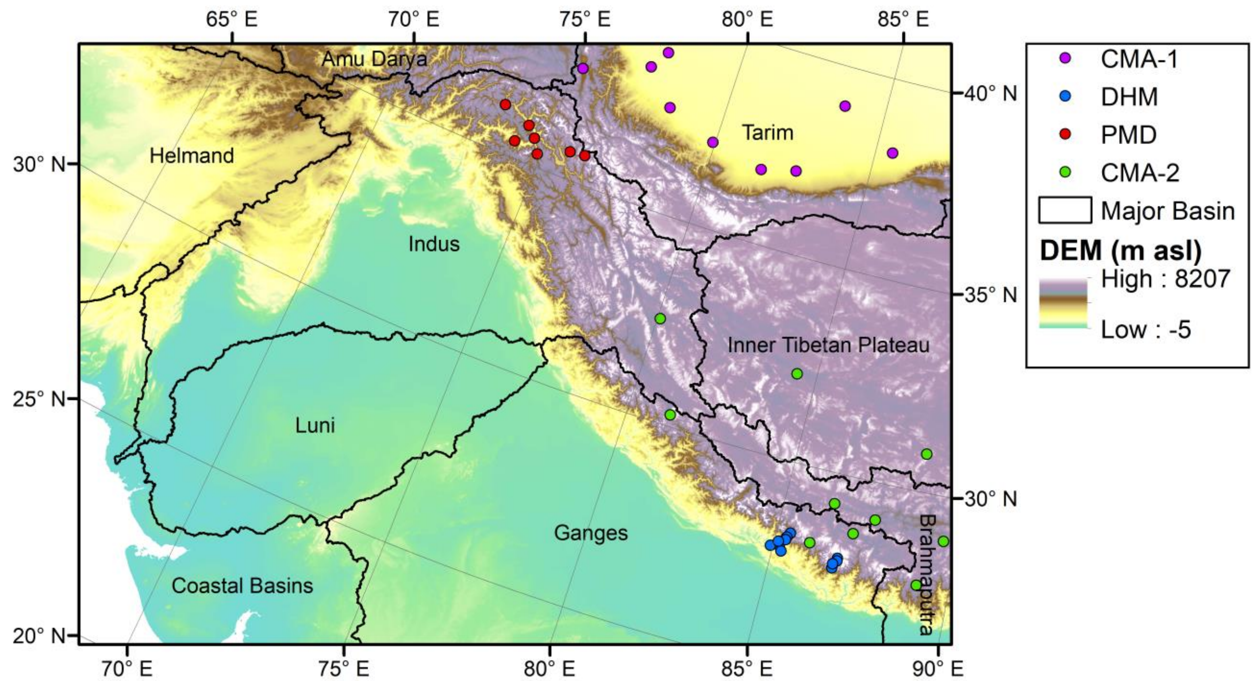


Figure 1. The High Mountain Asia region and locations of precipitation stations.

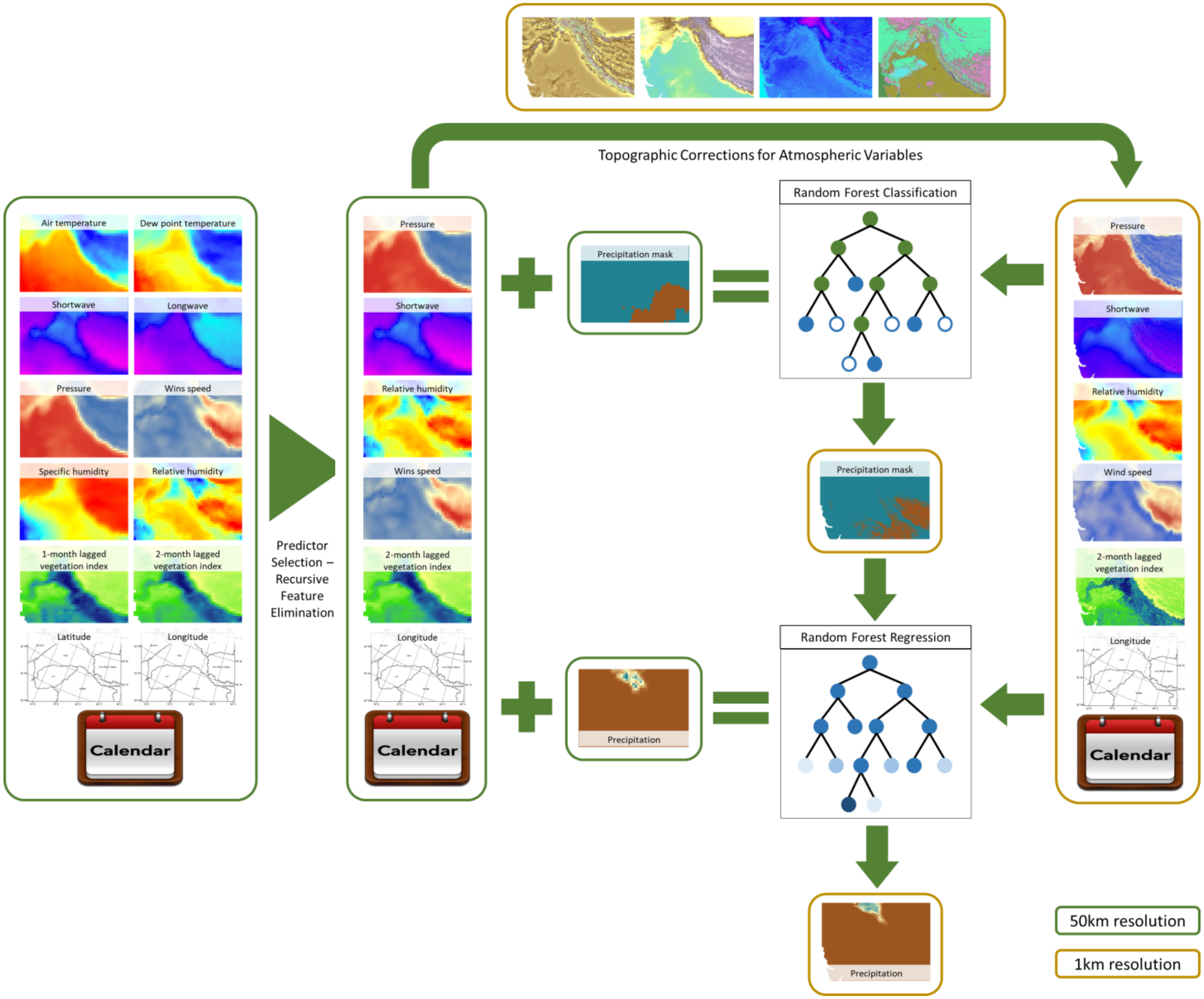


Figure 2. Flow chart of the precipitation downscaling framework.

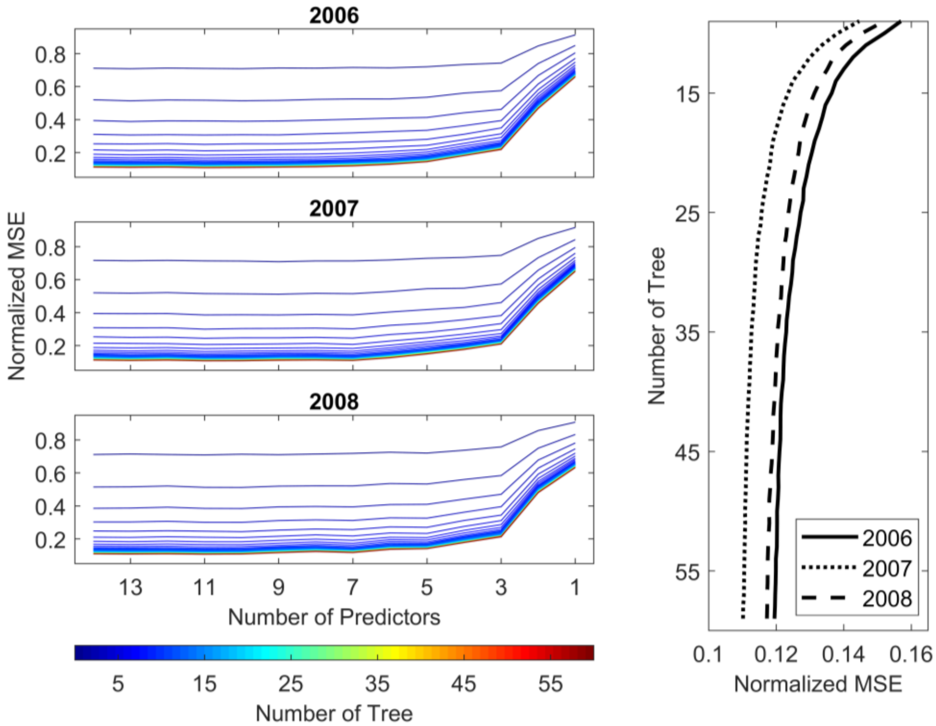


Figure 3. Model performance in terms of the normalized mean square error with number of predictors and trees for the three study years.

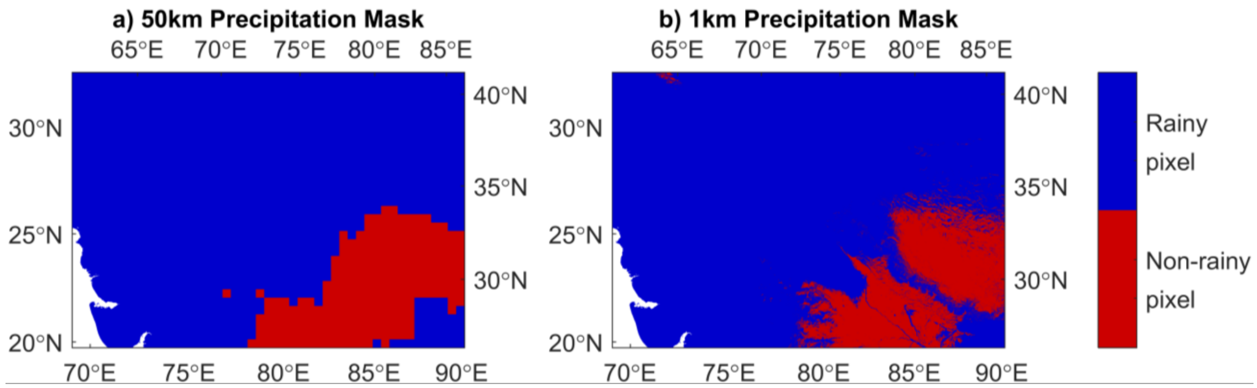


Figure 4. Comparison between 50km and 1km precipitation masks of January 1st 2007.

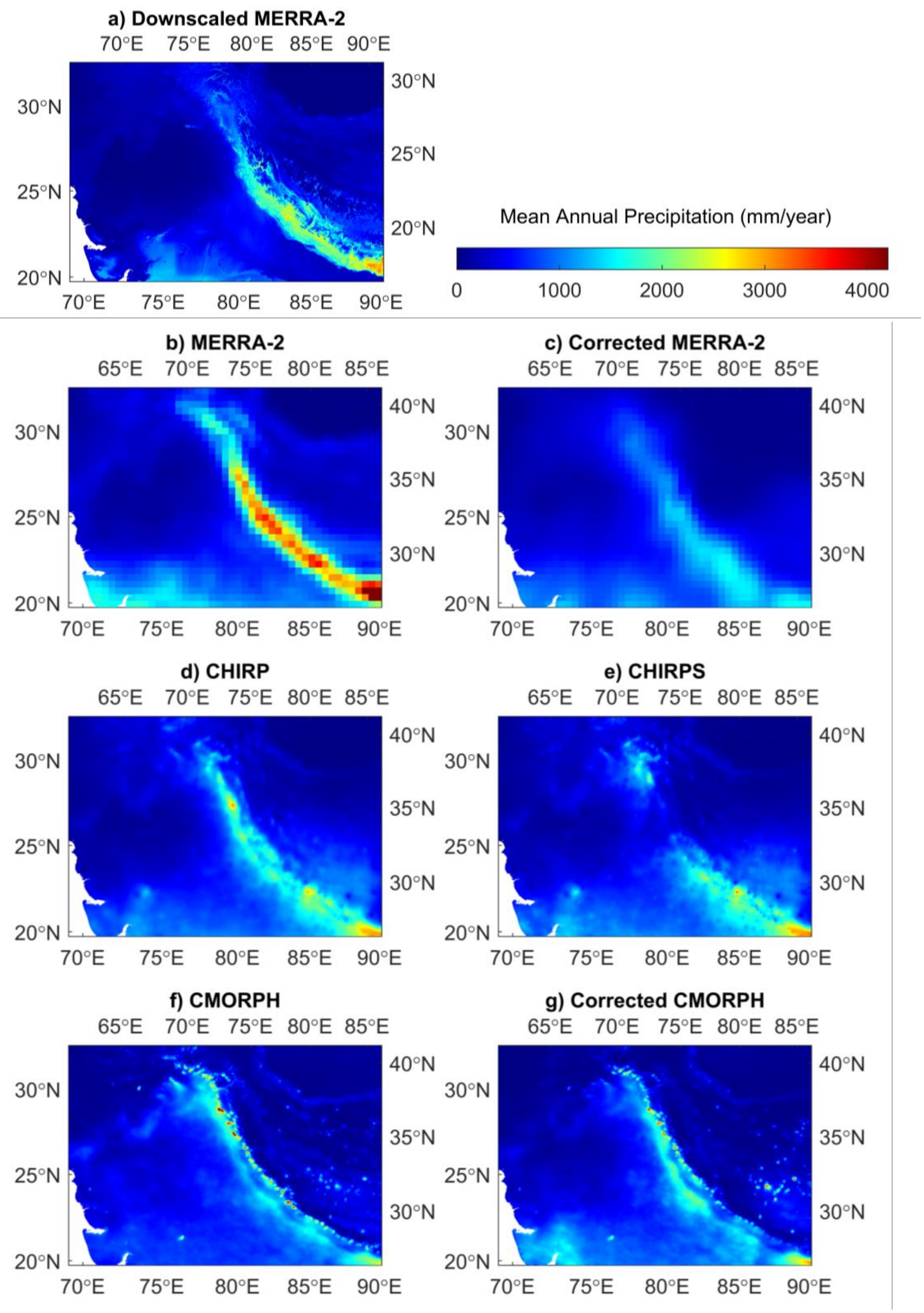


Figure 5. Mean annual precipitation during 2006 to 2008 for different precipitation datasets.

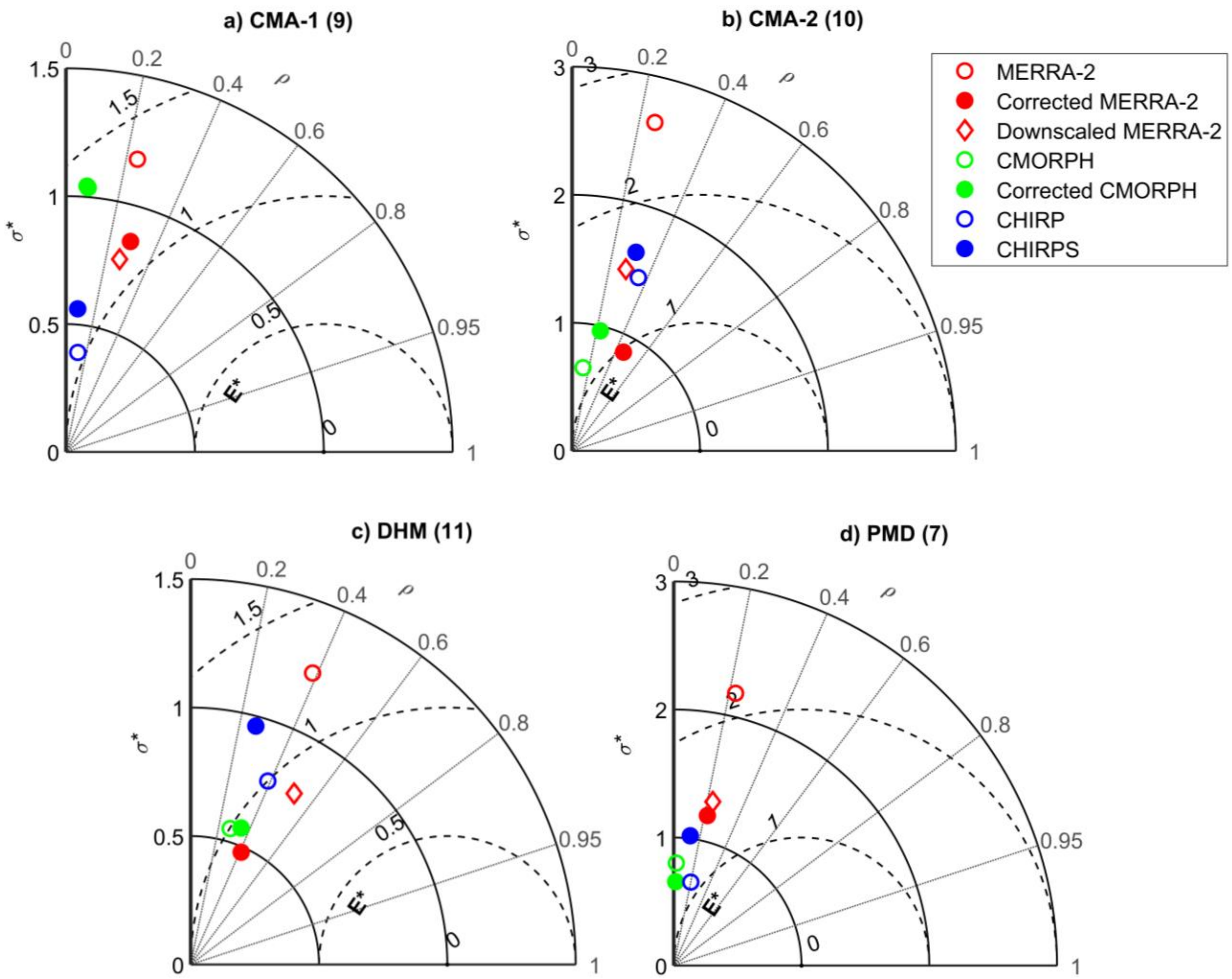


Figure 6. Taylor's diagrams for different precipitation products vs. a) the CMA-1, b) the CMA-2, c) the DHM, and d) the PMD gauge networks. The number of gauges used to plot the diagrams are shown in parenthesis in the panel titles.

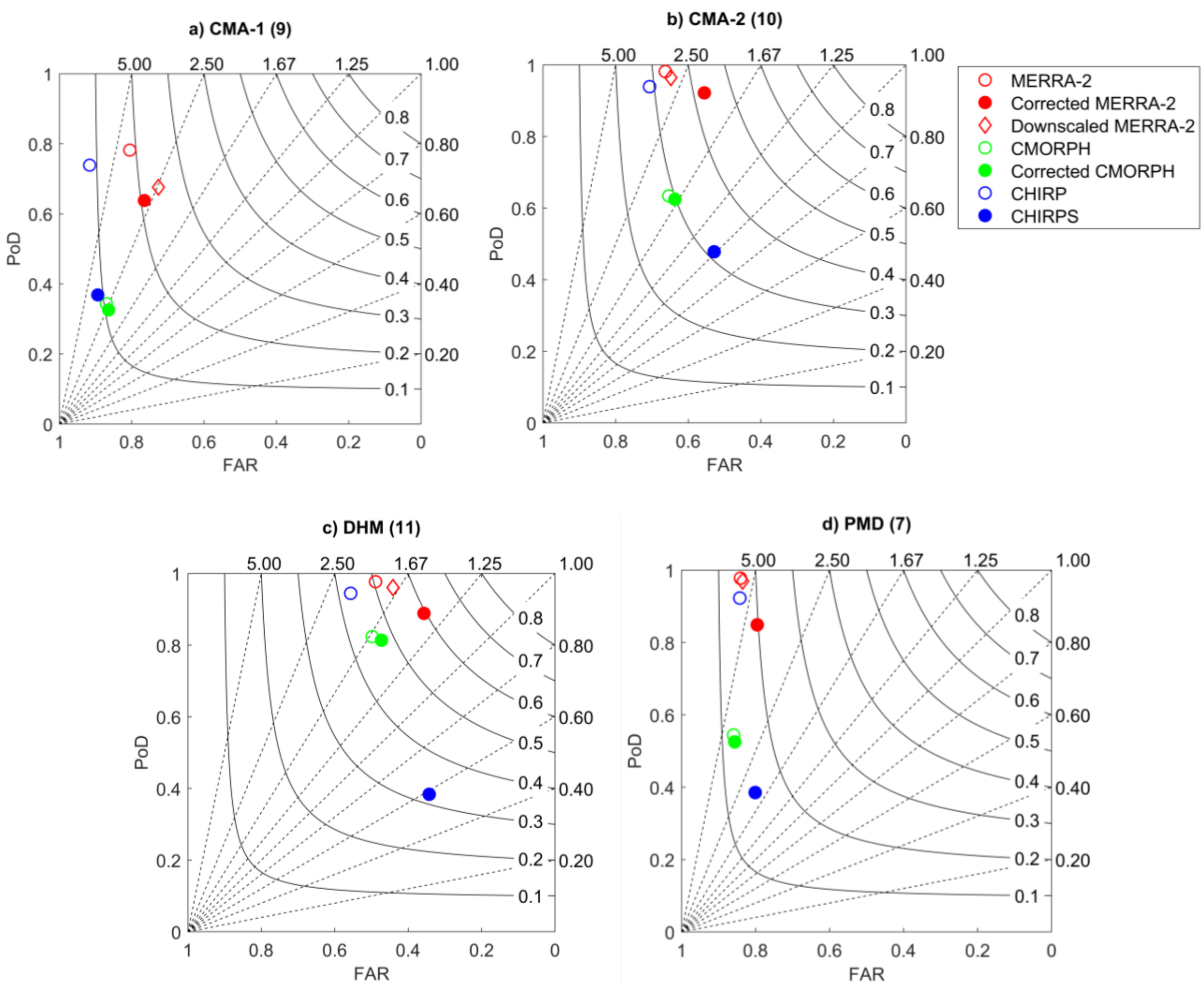


Figure 7. Same as Figure 6, but for performance diagrams. Solid curves represent *CSI* while the dash lines represent *BR*. A precipitation rate of 0.01mm/d is used as the rain/no-rain threshold.

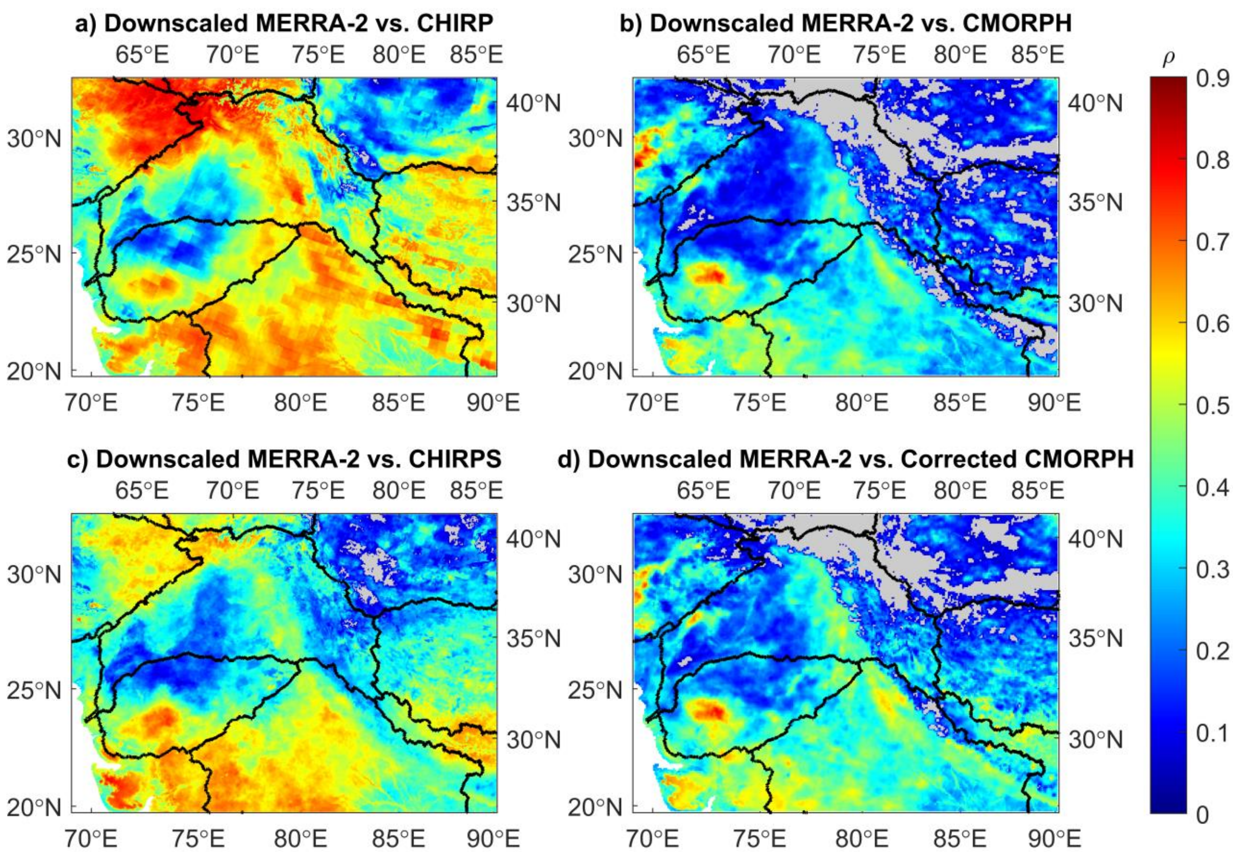


Figure 8. Spatial distributions of ρ derived for the downscaled MERRA-2 with each of the satellite precipitation product for 2006 to 2008 (locations with value that is not significantly larger than 0 are in gray). The major river basins boundaries are superimposed.

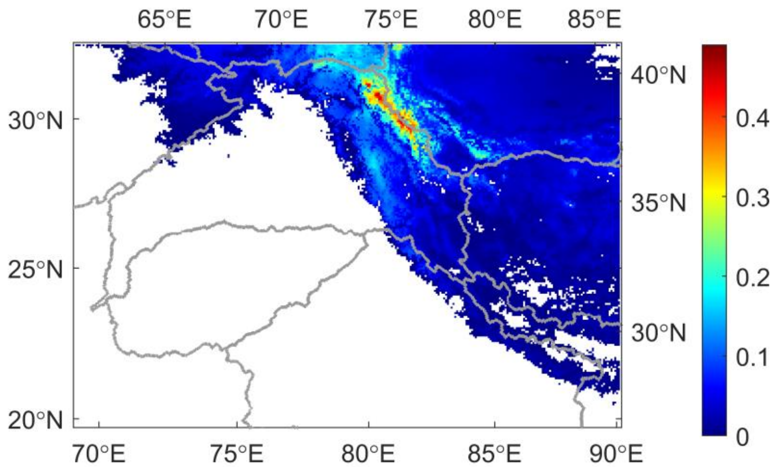


Figure 9. Fraction of missing values in CMORPH products during the December to February period. Area in white indicates no missing values. The major river basins boundaries are superimposed.

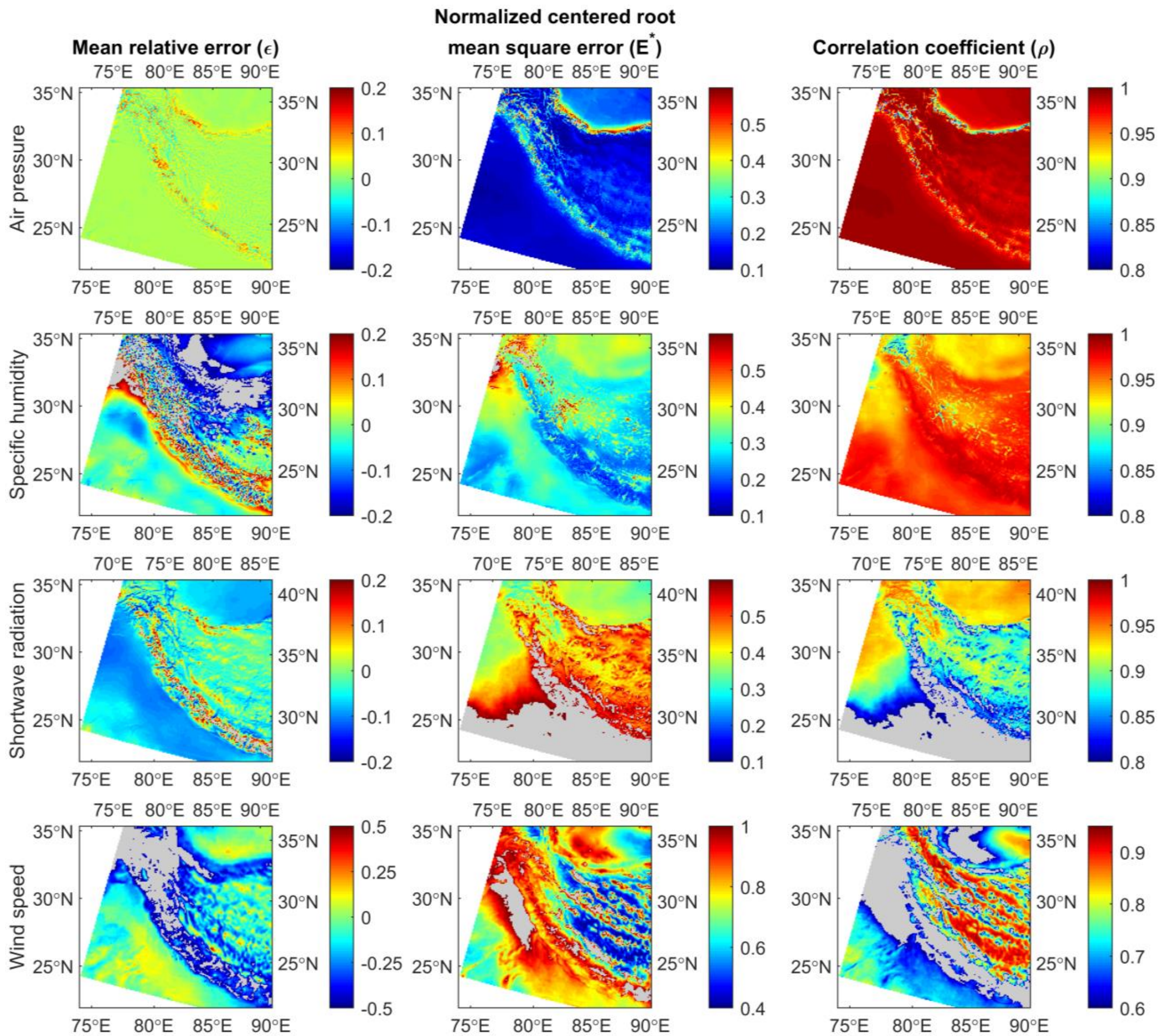
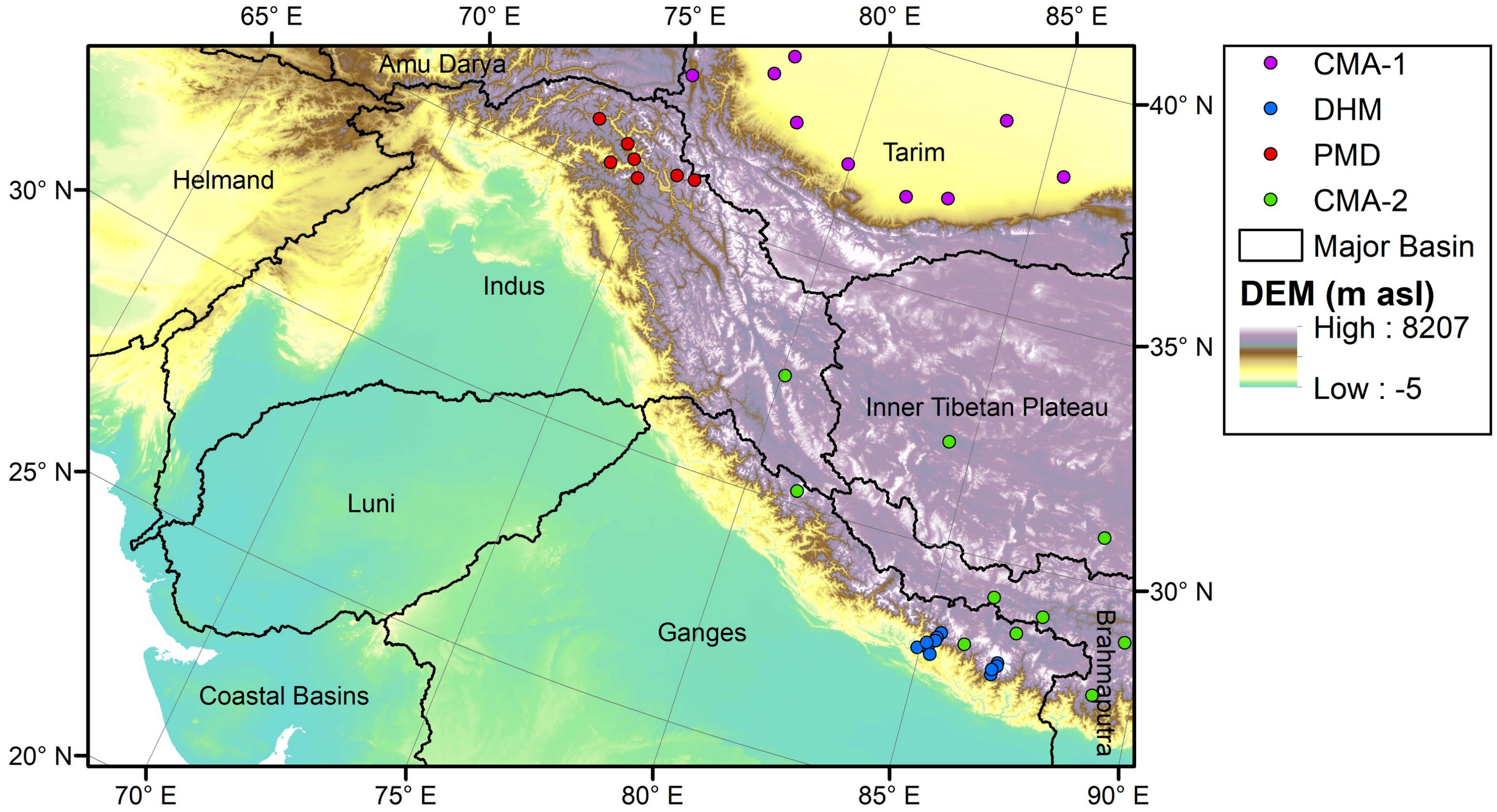


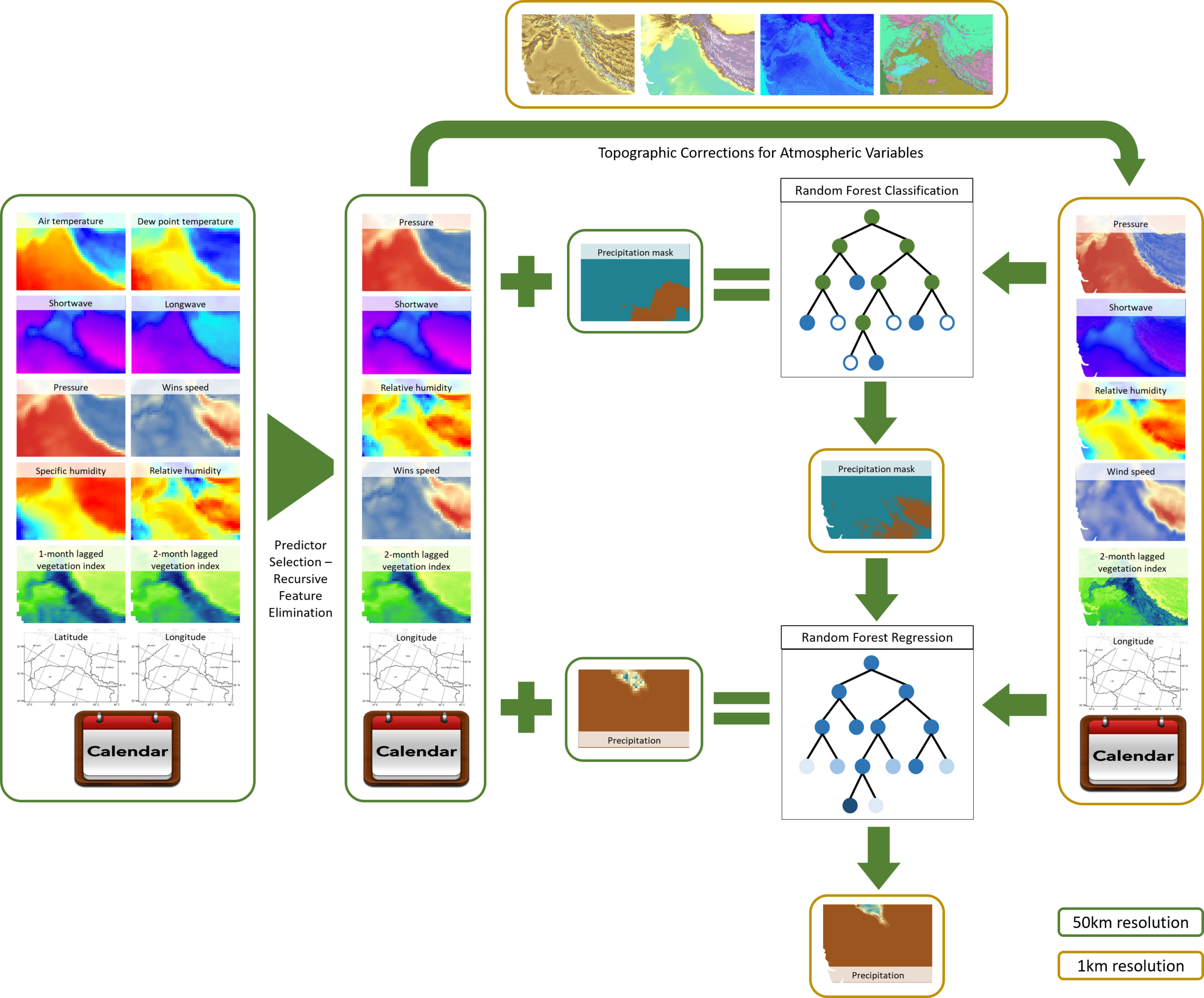
Figure A 1. Spatial distributions of error metrics derived for the topographically-corrected atmospheric variables with the corresponding HAR ones. Areas in gray indicate that the ε, E^{*}, or ρ are not significantly within ±0.2, lower than 0.6, or larger than 0.8 (those thresholds change to ±0.5, 1, or 0.6 for the case of wind speed). Note that the comparison domain is smaller than the precipitation downscaling domain given the smaller coverage of HAR.

Figure 1.



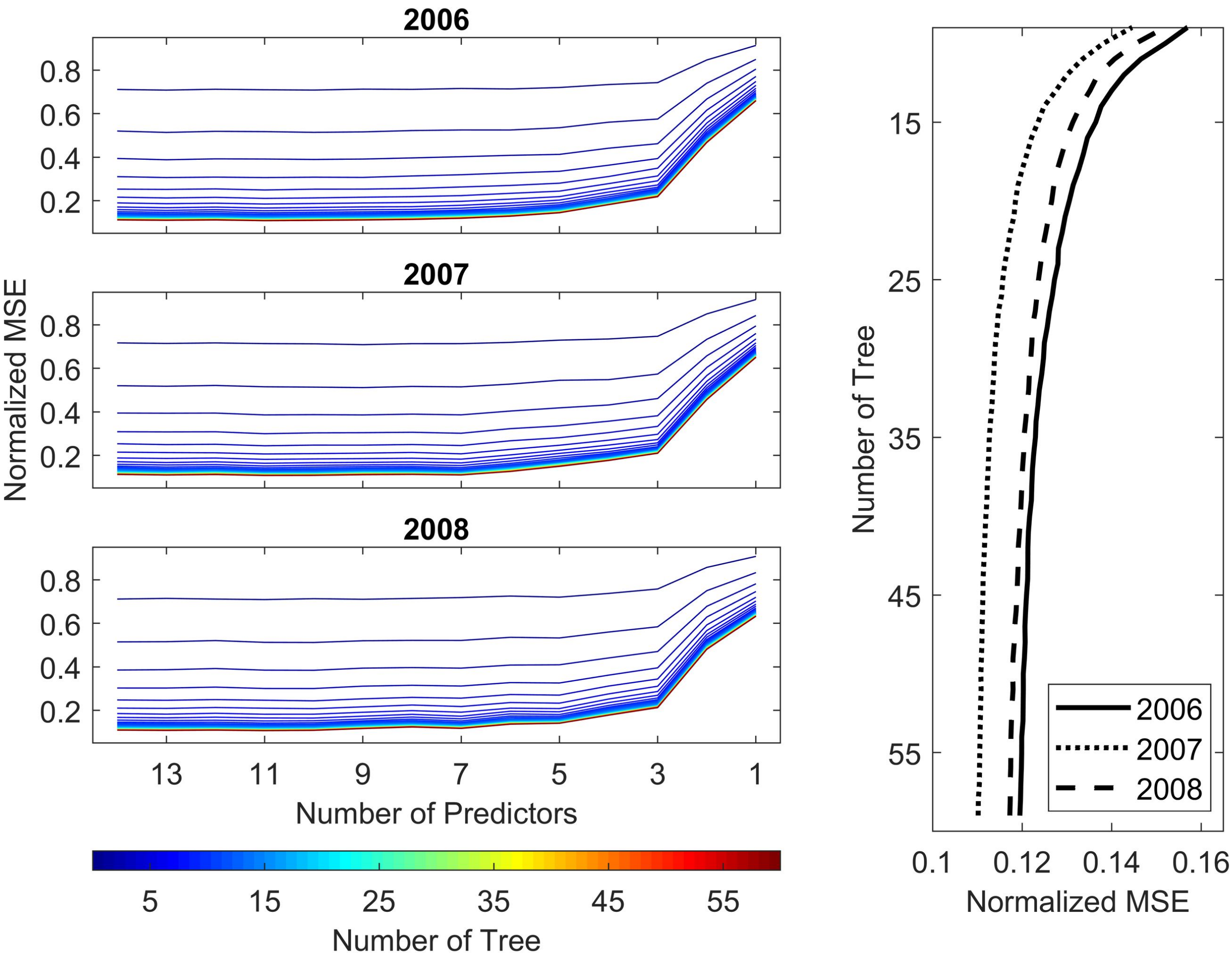
65° E
70° E
75° E
80° E
85° E
Amu Darya
Tarim
Helmand
Indus
Inner Tibetan Plateau
Luni
Ganges
Brahmaputra
Coastal Basins
40° N
35° N
30° N
30° N
25° N
20° N
70° E
75° E
80° E
85° E
90° E
CMA-1
DHM
PMD
CMA-2
Major Basin
DEM (m asl)
High : 8207
Low : -5

Figure 2.



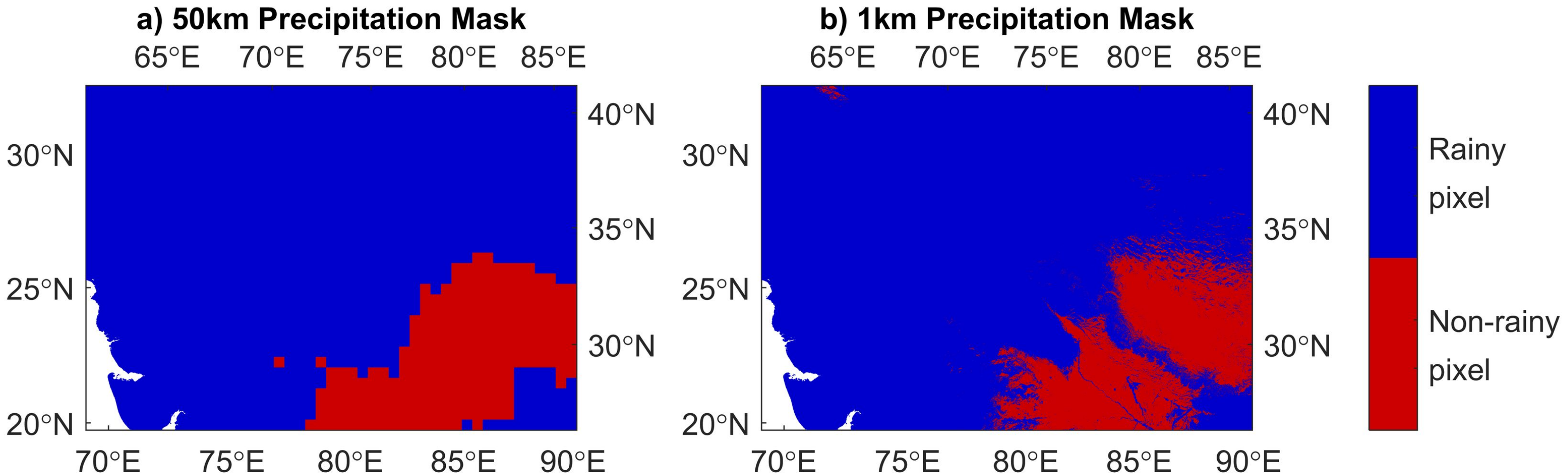
Topographic Corrections for Atmospheric Variables
Air temperature
Dew point temperature
Shortwave
Longwave
Pressure
Wins speed
Specific humidity
Relative humidity
1-month lagged vegetation index
2-month lagged vegetation index
Latitude
Longitude
Calendar
Predictor Selection – Recursive Feature Elimination
Pressure
Shortwave
Relative humidity
Wins speed
2-month lagged vegetation index
Longitude
Calendar
Precipitation mask
Random Forest Classification
Precipitation mask
Random Forest Regression
Precipitation
Precipitation
Pressure
Shortwave
Relative humidity
Wind speed
2-month lagged vegetation index
Longitude
Calendar
50km resolution
1km resolution

Figure 3.



2006
2007
2008
Normalized MSE
0.2
0.4
0.6
0.8
13
11
9
7
5
3
1
Number of Predictors
5
15
25
35
45
55
Number of Tree
15
25
35
45
55
0.1
0.12
0.14
0.16
Normalized MSE
2006
2007
2008

Figure 4.



a) 50km Precipitation Mask
b) 1km Precipitation Mask
65°E
70°E
75°E
80°E
85°E
90°E
20°N
25°N
30°N
35°N
40°N
Rainy pixel
Non-rainy pixel

Figure 5.

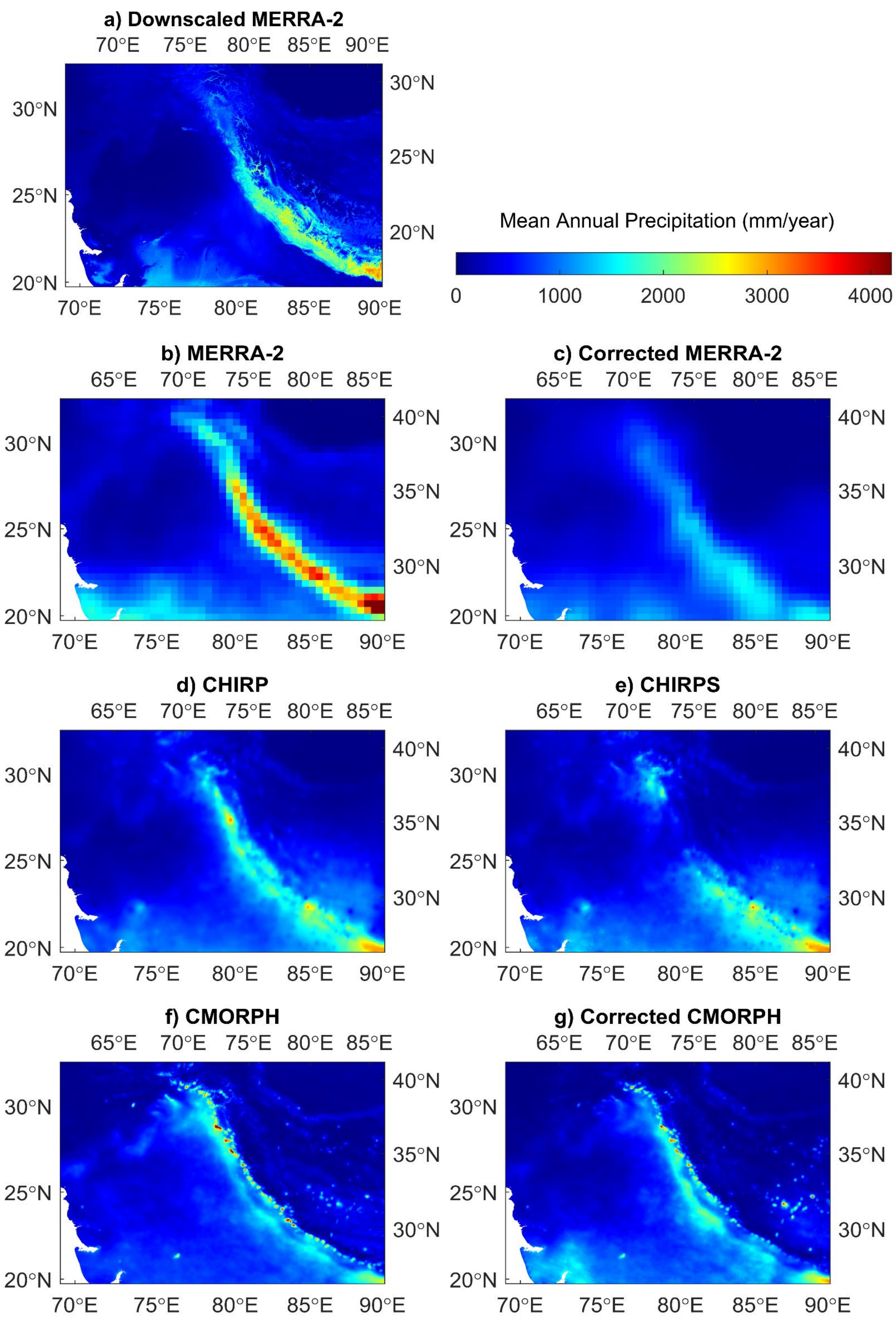
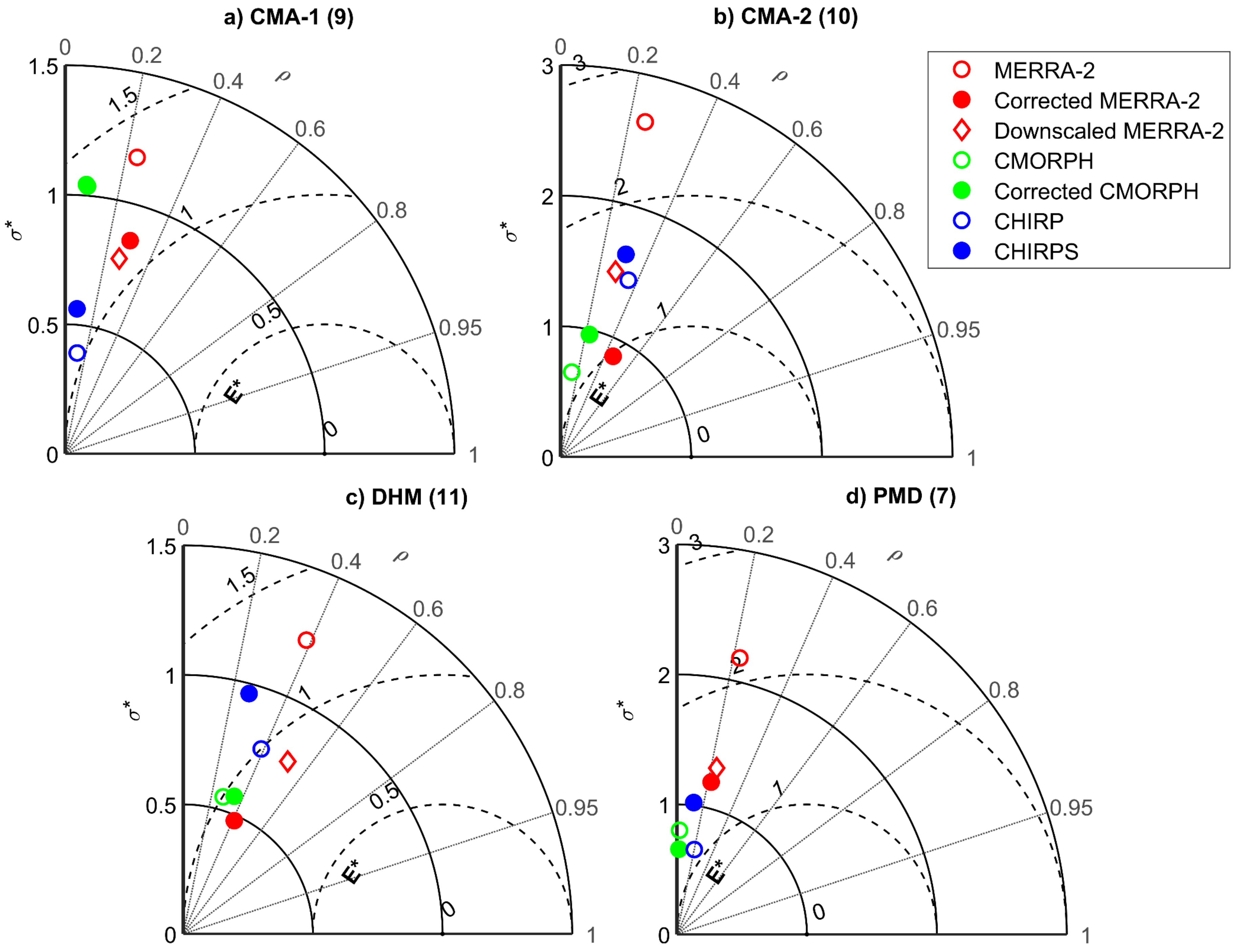
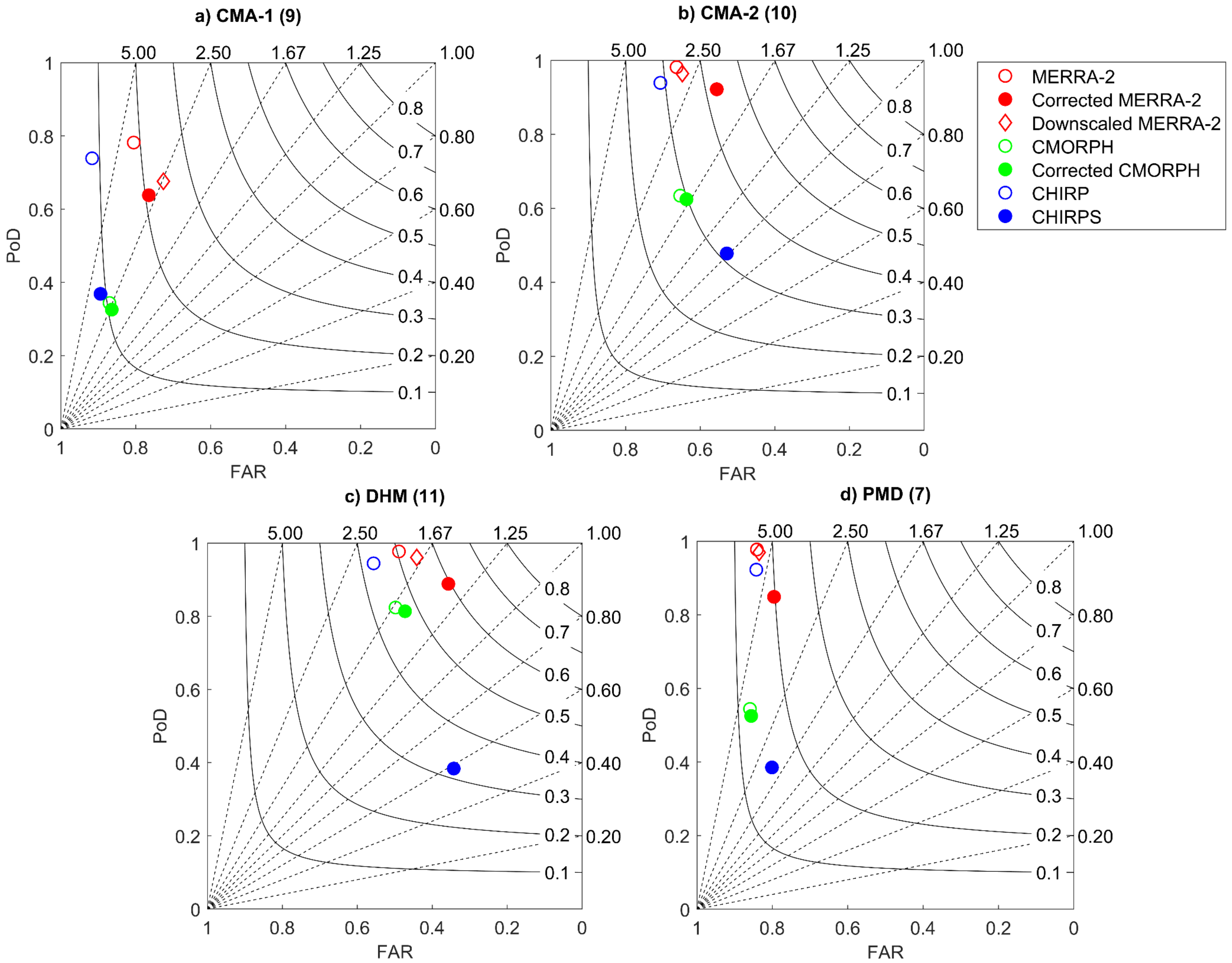

a) Downscaled MERRA-2
Mean Annual Precipitation (mm/year)
0
1000
2000
3000
4000
b) MERRA-2
c) Corrected MERRA-2
d) CHIRP
e) CHIRPS
f) CMORPH
g) Corrected CMORPH

Figure 6.



a) CMA-1 (9)
b) CMA-2 (10)
c) DHM (11)
d) PMD (7)
MERRA-2
Corrected MERRA-2
Downscaled MERRA-2
CMORPH
Corrected CMORPH
CHIRP
CHIRPS
σ^*
ρ
E*

Figure 7.



a) CMA-1 (9)
b) CMA-2 (10)
c) DHM (11)
d) PMD (7)
PoD
FAR
5.00
2.50
1.67
1.25
1.00
MERRA-2
Corrected MERRA-2
Downscaled MERRA-2
CMORPH
Corrected CMORPH
CHIRP
CHIRPS

Figure 8.

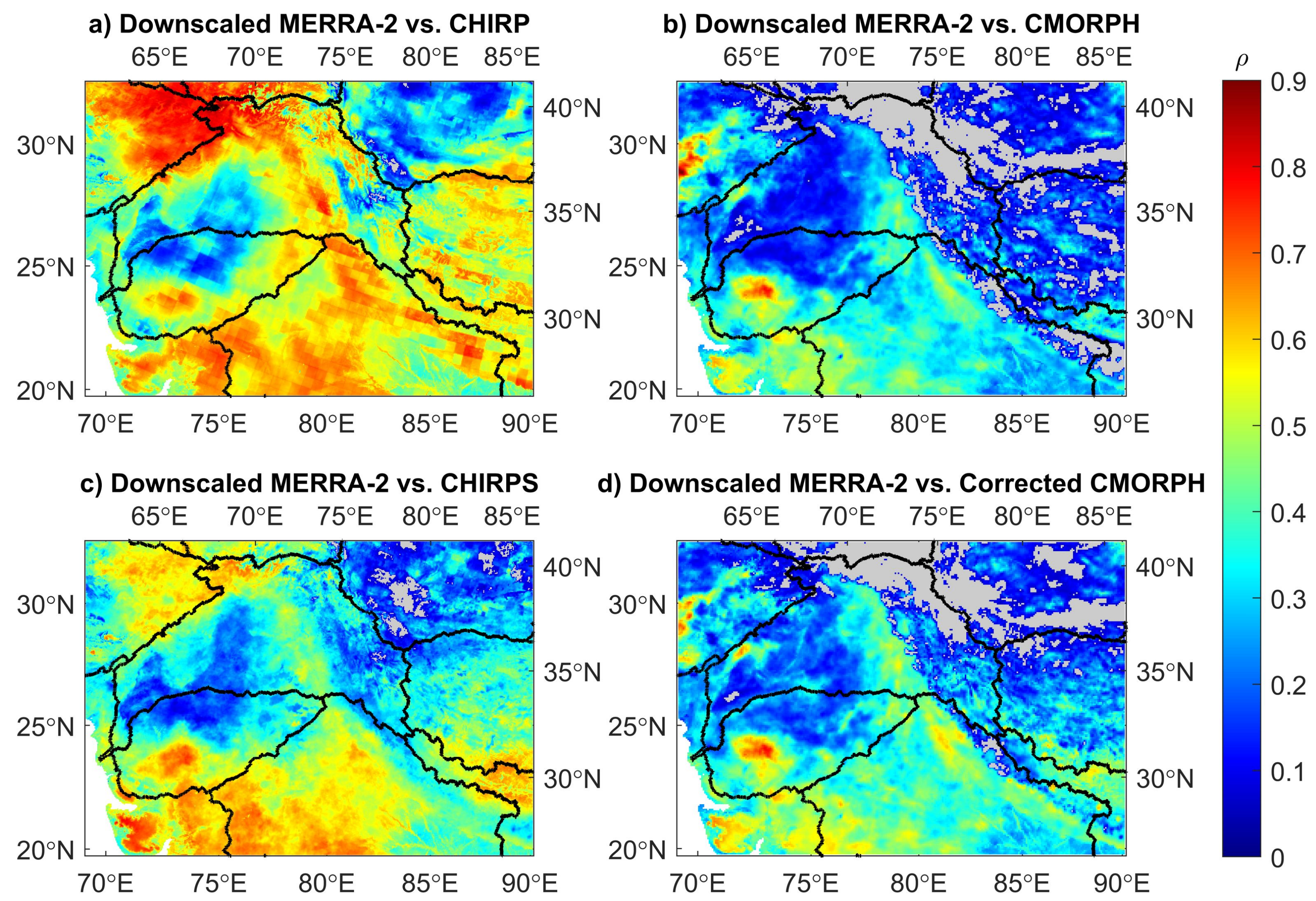

a) Downscaled MERRA-2 vs. CHIRP
b) Downscaled MERRA-2 vs. CMORPH
c) Downscaled MERRA-2 vs. CHIRPS
d) Downscaled MERRA-2 vs. Corrected CMORPH
65°E
70°E
75°E
80°E
85°E
90°E
20°N
25°N
30°N
35°N
40°N
ρ
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

Figure 9.

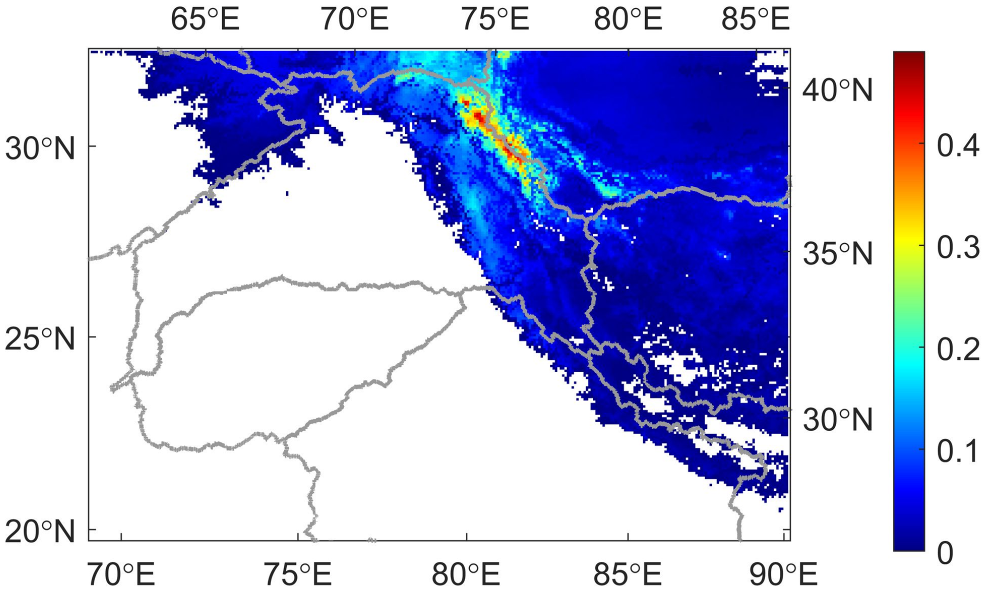
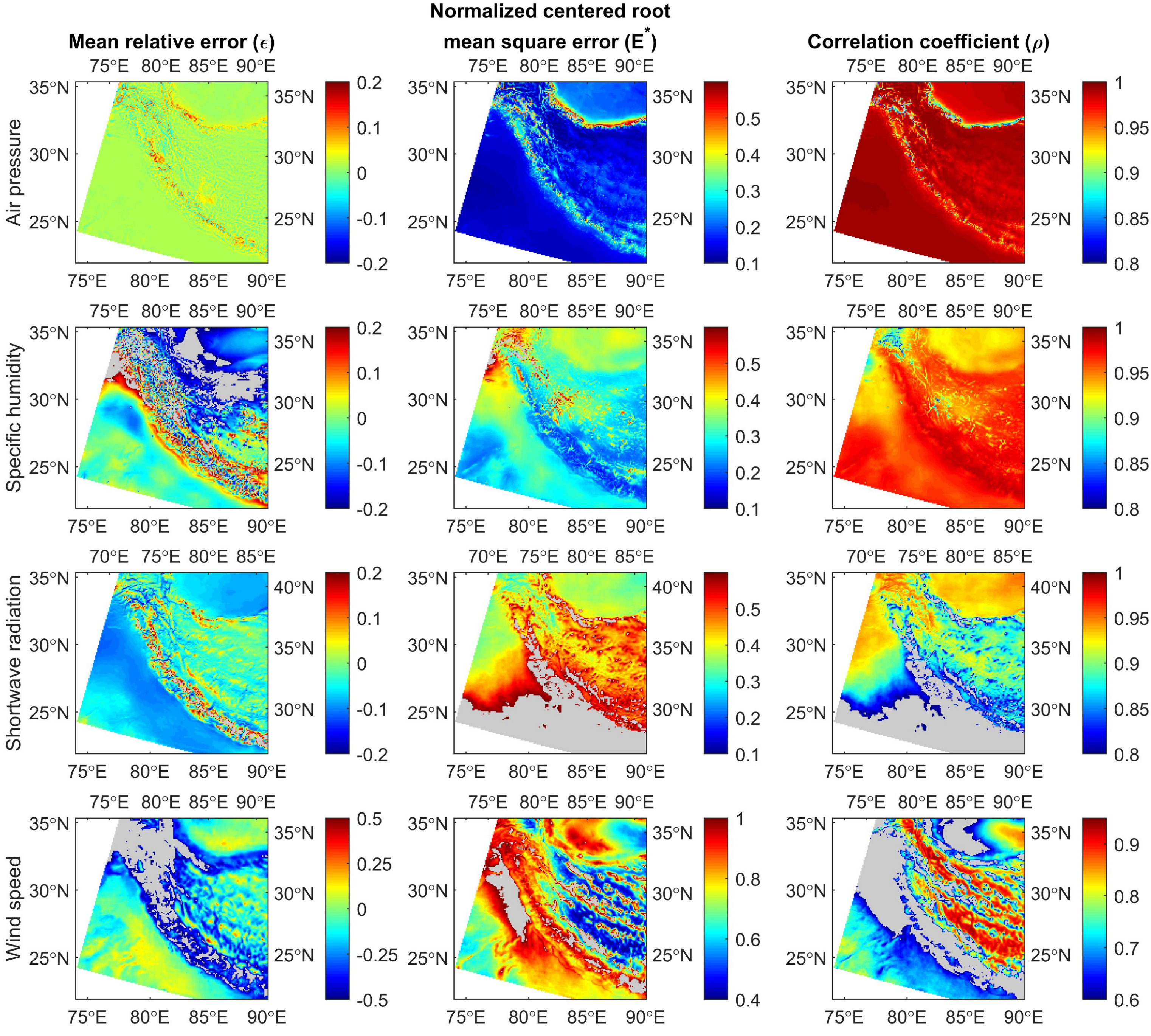

65°E
70°E
75°E
80°E
85°E
40°N
35°N
30°N
30°N
25°N
20°N
70°E
75°E
80°E
85°E
90°E
0.4
0.3
0.2
0.1
0

Figure A1.


Mean relative error (ϵ)
Normalized centered root mean square error (E*)
Correlation coefficient (ρ)
Air pressure
Specific humidity
Shortwave radiation
Wind speed
75°E 80°E 85°E 90°E
70°E 75°E 80°E 85°E
35°N 30°N 25°N
40°N 35°N 30°N
0.2 0.1 0 -0.1 -0.2
0.5 0.4 0.3 0.2 0.1
1 0.95 0.9 0.85 0.8
0.5 0.25 0 -0.25 -0.5
1 0.8 0.6 0.4
0.9 0.8 0.7 0.6