

Article

## Mapping Annual Precipitation across Mainland China in the Period 2001–2010 from TRMM3B43 Product Using Spatial Downscaling Approach

Yuli Shi <sup>1,2,\*</sup>, Lei Song <sup>1</sup>, Zhen Xia <sup>1</sup>, Yurong Lin <sup>1</sup>, Ranga B. Myneni <sup>3</sup>, Sungho Choi <sup>3</sup>, Lin Wang <sup>1</sup>, Xiliang Ni <sup>2</sup>, Cailian Lao <sup>3,4</sup> and Fengkai Yang <sup>1</sup>

<sup>1</sup> School of Geography and Remote Sensing, Nanjing University of Information Science and Technology, Nanjing 210044, China; E-Mails: pinot\_nuist@163.com (L.S.); 814189209xiazhen@sina.com (Z.X.); qiufeng1012@163.com (Y.L.); 13182926782@163.com (L.W.); super8533656@sina.com (F.Y.)

<sup>2</sup> State Key Laboratory of Remote Sensing Science, Jointly Sponsored by the Institute of Remote Sensing and Digital Earth of CAS and Beijing Normal University, Beijing 100101, China; E-Mail: nixl@irsa.ac.cn

<sup>3</sup> Department of Earth and Environment, Boston University, 675 Commonwealth Avenue, Boston, MA 02215, USA; E-Mails: ranga.myneni@gmail.com (R.B.M.); schoi@bu.edu (S.C.); laowan@cau.edu.cn (C.L.)

<sup>4</sup> College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

\* Author to whom correspondence should be addressed; E-Mail: ylshi.nuist@gmail.com; Tel.: +86-025-5869-9689.

Academic Editors: Richard Gloaguen and Prasad S. Thenkabail

Received: 20 January 2015 / Accepted: 24 April 2015 / Published: 8 May 2015

---

**Abstract:** Spatially explicit precipitation data is often responsible for the prediction accuracy of hydrological and ecological models. Several statistical downscaling approaches have been developed to map precipitation at a high spatial resolution, which are mainly based on the valid conjugations between satellite-driven precipitation data and geospatial predictors. Performance of the existing approaches should be first evaluated before applying them to larger spatial extents with a complex terrain across different climate zones. In this paper, we investigate the statistical downscaling algorithms to derive the high spatial resolution maps of precipitation over continental China using satellite datasets, including the Normalized Distribution Vegetation Index (NDVI) from the Moderate Resolution Imaging

Spectroradiometer (MODIS), the Global Digital Elevation Model (GDEM) from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), and the rainfall product from the Tropical Rainfall Monitoring Mission (TRMM). We compare three statistical techniques (multiple linear regression, exponential regression, and Random Forest regression trees) for modeling precipitation to better understand how the selected model types affect the prediction accuracy. Then, those models are implemented to downscale the original TRMM product (3B43; 0.25° resolution) onto the finer grids (1 × 1 km<sup>2</sup>) of precipitation. Finally we validate the downscaled annual precipitation (a wet year 2001 and a dry year 2010) against the ground rainfall observations from 596 rain gauge stations over continental China. The result indicates that the downscaling algorithm based on the Random Forest regression outperforms, when compared to the linear regression and the exponential regression. It also shows that the addition of the residual terms does not significantly improve the accuracy of results for the RF model. The analysis of the variable importance reveals the NDVI related predictors, latitude, and longitude, elevation are key elements for statistical downscaling, and their weights vary across different climate zones. In particular, the NDVI, which is generally considered as a powerful geospatial predictor for precipitation, correlates weakly with precipitation in humid regions.

**Keywords:** TRMM 3B43; NDVI; spatial downscaling; precipitation

---

## 1. Introduction

Precipitation is a vital element of water cycle in the Earth System, which is closely related to ecological, hydrological, and meteorological processes [1,2]. Its spatial and temporal variations generally influence vegetation distribution, soil moisture and surface runoff [3,4]. Thus, high-quality of precipitation dataset is needed for the development of ecological and hydrological models at corresponding scales.

Precipitation dataset used in recent ecological and hydrological researches is mainly derived from three sources: (a) outputs from various numerical climate/weather models; (b) rain gauge observations; and (c) estimates from space-based observing systems [5–7]. For example, the Global Climate Models (GCMs) can simulate the changes of precipitation at large scales [5]; the Regional Climate Models (RCMs) can be used to predict the distribution of precipitation at a regional scale [8]. Neither the GCMs nor the RCMs, however, can provide precipitation fields with higher spatial resolutions reflecting the spatio-temporal variations at small scales [9]. Since the numerical weather models normally function at a coarse spatial resolution, scientists alternatively use *in situ* rain gauge (spatially sparse) or satellite-based observations to build the precipitation dataset. With the help of spatial analysis technique, precipitation data from rain gauge stations can be interpolated and extrapolated over the regions with no data samplings. Daly [10] has established a statistical regression model known as the Parameter elevation Regressions on Independent Slopes Model (PRISM) to simulate the distribution of precipitation over the Olympic Peninsula, in the northwest corner of Washington State, USA. The DAYMET model uses daily weather observations (1980–1997) to produce climate grids of annual total precipitation and other

climatic variables over the continental US [11]. Still, these rain gauge-based methods are accurate only within the area where rain gauge stations are spatially, densely installed [12,13]. Furthermore, weather radars offer an enormous potential to improve the quality of rainfall at high resolution. The wider spatial coverage of weather radars compared to that of a dense network of rain-gauges is an obvious advantage. However, despite the progress in technology and methodology over past decades, radar data are still not used as broadly and efficiently as they should be [14,15].

To acquire adequate and reliable spatial representations of precipitation over a broad area, scientists begin to explore novel approaches implementing the recent remote sensing precipitation products [9,14–19]. The National Aeronautics and Space Administration (NASA) and the Japanese space agency (JAXA) launched the Tropical Rainfall Measuring Mission (TRMM) in 1997, which is one of the highest resolution products among all the current satellite precipitation datasets. TRMM has been extensively used for inter-disciplined investigations and applications, such as land surface modeling [17,20], climatic prediction [21], and hydrological simulation [22]. Nevertheless, the large pixel size (*i.e.*,  $0.25^\circ$ ) of the TRMM precipitation data is too coarse for many regional and continental models, which generally need higher resolution inputs [23]. The above precipitation products, thus, need to be, first, downscaled to meet the requirement of being in a high spatial resolution.

Statistical downscaling is a recently-developed approach in obtaining high spatial resolution of variables, based on conjugations between the variable at a coarse scale and geospatial predictors at a finer resolution [24,25]. This technique has been widely used for downscaling variables, such as land surface temperature [26,27], river flow [28], soil moisture [29,30], and vegetation fraction cover [31]. Specifically, studies of downscaling TRMM precipitation data have been conducted on at a regional scale in arid and humid regions. Fang *et al.* [32] assumed that the spatial variability of precipitation could be well explained by local topography and prestorm meteorological factors, and developed a statistical spatial downscaling scheme to disaggregate the TRMM 3B42 products into a 1 km gridded rainfall field in a mountainous area. An exponential statistical regression model between the Normalized Differential Vegetation Index (NDVI) and precipitation is useful to downscale precipitation from the TRMM monthly product in the Iberian Peninsula [33]. Based on this downscaling method, Jia *et al.* [34] have developed a statistical regression model by introducing NDVI and elevation, which are seen as the main factors affecting precipitation, and have improved the resolution of the TRMM precipitation from  $0.25^\circ$  to 1 km in the Qaidam Basin. Similar to the examples above, two previous examples of downscaling research, Duan and Bastiaanssen [24] used an integrated downscaling-calibration procedure of the TRMM 3B43 product with the limited rain gauge data sets to map monthly precipitation data at a higher spatial resolution (1 km) in humid and semi-arid regions.

Those previous approaches in recent literature can effectively downscale precipitation data at a local scale, but we need to explore the applicability of the existing algorithms for a larger spatial extent with complex terrain and different climate zones. Lovejoy *et al.*'s [35,36] suggested global scaling of TRMM satellite radar data is quite remarkable, and extends from kilometers to the size of the Earth. Kang *et al.* [37] analyzed Next Generation Weather Radar rainfall data and found that rainfall fluctuations at spatial scales smaller than a reference scale exhibit self-similarity and that at scales larger than the reference scale, rainfall fluctuations are scale dependent. A coupled stochastic space-time intermittent random cascade model was then proposed to downscale summer daily rainfall for the Central United States from a scale of 256 km to a scale of 2 km. Chen *et al.* [38] assumed that the rainfall-geospatial factors

relationship varies spatially but is similar in a region and constructed geographically weighted regression model for rainfall downscaling in North China. In addition, it is worth testing the validity of the suggested geospatial predictors (e.g., NDVI and elevation) because of certain contradictions in the literature: for instance, one analysis shows there is no clear relationship between precipitation and elevation in the Pangani Basin in Tanzania [39], while another finds a clear relationship between them in the Qaidam Basin in China [34]. Moreover, a number of geospatial predictors are still not explicitly examined in current research, and such predictors include geo-locations and topography (slope and aspect) that are associated with solar radiation and moisture conditions, in turn, granting significant impacts on precipitation [40–42]. The analysis of the importance of predictors would be useful to investigate the potential of geospatial predictors with regard to downscaling precipitation over complex terrain and climate zones.

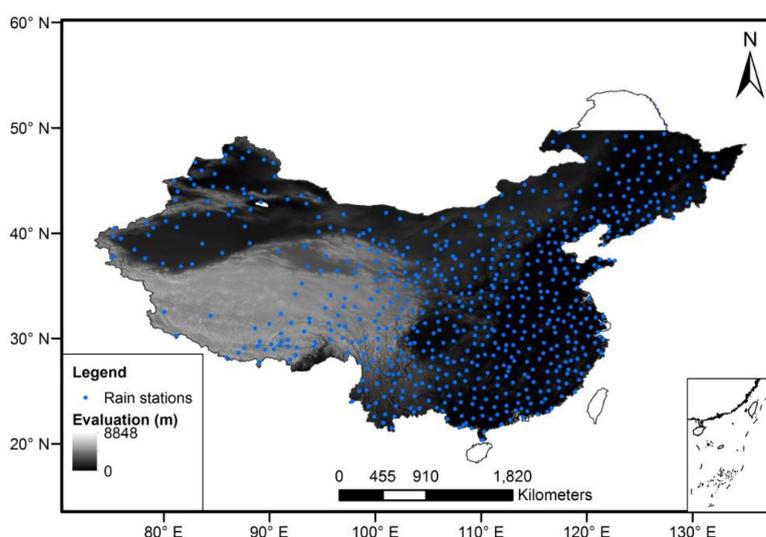
The Random Forest (RF) approach is an ensemble, purely statistically-based modeling, which constructs numerous small regression trees that vote on predictions using the random samplings of data and model layers [43]. Amongst many non-parametric regression approaches, RF is receiving a considerable attention of ecological and other applications [44–46]. This is because: (1) the robustness of RF may avoid over-fitting; (2) many different types of input variables can be implemented without variable deletion and regularization; and (3) it has tremendous analytical and operational flexibility. Thus, RF would be useful to build regression models for precipitation in relation to a number of geospatial predictors across different spatial scales, and to downscale the original TRMM 3B43 data to 1 km resolution.

In this study, our main goal is to investigate several downscaling approaches generating annual precipitation with a high spatial resolution over a variety of arid-humid regions in China Mainland. In this ultimate goal, three sub-objectives are: (1) to compare three statistical techniques (multiple linear regression, exponential regression, and RF regression trees) of modeling precipitation for better understanding how the model selection affects the performance of downscaling; (2) to explore the importance of input geospatial predictors in the regression models across different climatic zones; and (3) to generate the annual precipitation map of China mainland at a 1 km resolution, based on the best-performed downscaling approach. We fill the gap between the observation and estimates, and validate the final results with an independent precipitation dataset from 596 meteorological stations in China mainland. This research, in particular, demonstrates a unique application of machine learning technique (*i.e.*, RF) to downscale the coarse TRMM 3B43 product, and, in turn, to provide a contextual framework for interpreting the fine spatial distribution of precipitation at a sub-continental scale. This study can have practical implications, particularly for biological researchers who require annual precipitation estimates at finer resolution as an input of tree height or biomass models over complex terrain, where they are characterized by the sparseness of *in situ* networks for precipitation measurement.

## 2. Study Area

China is located in the eastern part of Asia, on the west coast of the Pacific Ocean. It extends from 73°40'E to 135°2'E and from 3°52'N to 53°33'N. Mainland China (except the island of Taiwan, Hainan Island, and other small islands) encompasses complex terrain and a distinct temperate monsoon climate: areas of mountains, hills, and plateaus take up the two thirds of the entire mainland (Figure 1);

the direction of the seasonal wind changes significantly within a year. These geographical and climatic traits lead to a strong spatial pattern of precipitation and, thus, form arid and humid conditions [47]. The arid areas (annual precipitation less than 200 mm/yr), semi-arid areas (annual precipitation between approximately 200 and 400 mm/yr), semi-humid areas (annual precipitation between approximately 400 and 800 mm/yr), and humid areas (annual precipitation between approximately 800 and 2000 mm/yr) are distributed from the northwest inner part of the land to the southeast coast. Similar to this pattern, vegetation coverage in the continental area of China reflects the distribution of annual precipitation. Combined with other evidence [48,49], it naturally indicates the strong positive correlations between precipitation and topography/vegetation cover. This assumption consists of the fundamental of the algorithm for downscaling spatial precipitation. It should be noted that the region greater than latitude 50° N is not included in our study due to the fact that TRMM does not cover this area of China.



**Figure 1.** Elevation over Mainland China and location of 596 meteorological stations.

### 3. Dataset and Methodology

#### 3.1. Dataset

##### 3.1.1. Tropical Rainfall Measuring Mission

NASA and JAXA launched a joint project called the Tropical Rainfall Measuring Mission (TRMM) on 27 November 1997. TRMM aims to measure the intensity and area coverage of rainfall around the tropical and semi-tropical area where two thirds of the world's rainfall happens [50,51]. TRMM can provide accurate precipitation data between latitude 50° N and latitude 50° S at the resolution of  $0.25^\circ \times 0.25^\circ$  (approx. 28 km  $\times$  28 km), which is high compared to other satellite-based products. There are two TRMM precipitation products commonly used, namely TRMM 3B42 and TRMM 3B43 datasets. TRMM 3B42 provides three-hour averaged precipitation values. It is converted into monthly TRMM 3B43 precipitation data, which was used in this study. We aggregated the TRMM 3B43 monthly product of Mainland China (12 months) into annual total precipitation of 2001 or 2010 at a spatial resolution of  $0.25^\circ$  (Figure A1). The equation deriving annual total precipitation is as follows (Equation (1)):

$$P_{TRMM,i} = \sum_{j=1}^m P_{TRMM\_monthly,j} \quad (1)$$

where  $P_{TRMM,i}$  is the  $i$ -th annual precipitation of Mainland China from 2001 or 2010,  $P_{TRMM\_monthly,j}$  is the monthly TRMM 3B43 precipitation product for  $j$ -th month ( $m = 12$ ). The TRMM 3B43 dataset used in this study was provided by the International Scientific and Technical Data Mirror Site, Computer Network Information Center, Chinese Academy of Sciences (<http://www.gscloud.cn>). The annual total precipitation in China from 2001 to 2010 is illustrated below (Figure A1). We compared the scatterplots between annual rain gauge precipitation from rain gauge stations and the products from original annual TRMM 3B43 precipitation for 2001 and 2010 over Mainland China. We found original TRMM precipitation is in agreement with measured rain gauge data ( $R^2 = 0.92$ , RMSE = 176 mm, MAE = 108 mm, Bias = 0.05 for the year 2001, and  $R^2 = 0.94$ , RMSE = 157 mm, MAE = 52 mm, Bias = 0.06 for the year 2010).

### 3.1.2. Normalized Difference Vegetation Index

Based on differences in pigment absorption features in the red ( $\sim 0.660 \mu\text{m}$ ) and near-infrared ( $\sim 0.860 \mu\text{m}$ ) band, Rous *et al.* (1974) proposed an vegetation index, called Normalized Distribution Vegetation Index (NDVI) [52] (Equation (2)).

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}} \quad (2)$$

NDVI can be effective in responding to changes in the amount of green biomass, chlorophyll content, and canopy water stress, and have been widely used in various applications since it was proposed. Several investigations have indicated that there exists a strong positive correlation between precipitation and NDVI [53–55]. In this study we chose the MODIS NDVI products as a geospatial predictor for the downscaling of TRMM precipitation. The characters of the MODIS NDVI products, MOD13A3, are listed in the Table 1. We aggregated the MODIS monthly products, MOD13A3 (12 months), into the annual average NDVI of Mainland China for 2001 and 2010 at a spatial resolution of 250 m. We also generated the annual maximum, minimum, and range NDVI (maximum NDVI minus minimum NDVI) of Mainland China. Then, we resampled those products at a spatial resolution of 1 km by applying the nearest neighbor resize method. The equation deriving the annual average NDVI is as follows (Equation (3)):

$$NDVI_{avg.,i} = \frac{\sum_{j=1}^n NDVI_{monthly,j}}{n} \quad (3)$$

where  $NDVI_{avg.,i}$  is the  $i$ -th annual average NDVI of Mainland China for 2001 to 2010,  $NDVI_{monthly,j}$  is the monthly maximum NDVI product for  $j$ -th month ( $n = 12$ ). We analyzed the NDVI trends during the research period and found the NDVI trends for all growing seasons increased moderately, except for NDVI values during the period of research years. This is consistent with the research conducted by Li *et al.* [56]. The MOD13A3 dataset used in this study is provided by the International Scientific and Technical Data Mirror Site, Computer Network Information Center, Chinese Academy of Sciences (<http://globalchange.nasdc.cn>).

### 3.1.3. ASTER Global Digital Elevation Model

ASTER GDEM was released by the Ministry of Economy, Trade, and Industry (METI) of Japan and the United States National Aeronautics and Space Administration (NASA). ASTER GDEM covers land surface between latitude 83° N and 83° S of the Earth, which includes the entire area of Mainland China. GDEM version 2, with a resolution of one arc-second (roughly 30 m), was used in this study to analyze the influence of topography on precipitation (Figure 1). In addition to the elevation data, two terrain attributions, slope and aspect, were derived from ASTER GDEM. The ASTER GDEM data was downloaded from the website, <http://gdem.ersdac.jspacesystems.or.jp/>.

### 3.1.4. Rain Gauge Data

Monthly total precipitation datasets for 2001 and 2010 were used in this study to independently validate the final downscaled precipitation product. Those data were downloaded from the website, <http://cdc.cma.gov.cn>. The precipitation values have been measured from 596 rain gauges of the national weather stations of Mainland China, which are distributed in a relatively sparse network. Most of the meteorological stations are mainly located in Eastern China, and few cover Western and Northwestern China (Figure 1). Those monthly data were used to calculate the annual total precipitation of China over ten years. Our analysis indicates that the annual total precipitation for the period of 2001 and 2010 varied from 0 to 2672 mm/yr. Table 1 lists datasets that were used in this study to build the statistical regression models and to downscale the precipitation.

**Table 1.** Variables used in this study to construct the statistical regression models and to downscale the TRMM 3B43 precipitation product.

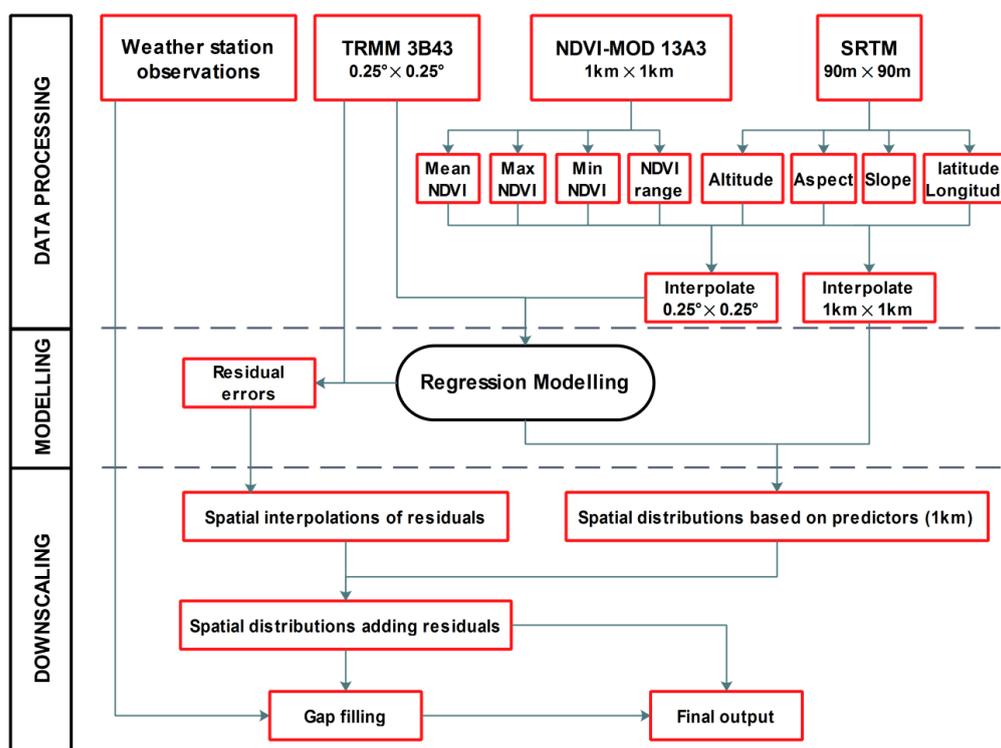
Variables	Dataset	Year	Resolution
Precipitation (mm)	TRMM3B43	2001, 2010	Monthly, 0.25°
NDVI	MOD13A3	2001, 2010	Monthly, 1000 m
Max_NDVI	MOD13A3	2001, 2010	Annual, 1000 m
Min_NDVI	MOD13A3	2001, 2010	Annual, 1000 m
Range_NDVI	MOD13A3	2001, 2010	Annual, 1000 m
Elevation (m)	GDEM	2010	-, 30 m
Slope	GDEM	2010	-, 30 m
Aspect	GDEM	2010	-, 30 m
Latitude	-	-	-, -
Longitude	-	-	-, -

## 3.2. Methodology

Statistical downscaling methods recently developed at regional scales were employed to downscale the TRMM dataset of Mainland China. The algorithms have successfully predicted variables at finer scales mainly based on the relationships between variables (such as land surface temperature and precipitation) and geospatial predictors (such as NDVI and elevation) [24,26,34]. Commonly used regression models are the multiple linear and exponential models. The two models have generally performed well, using several combinations of different predictors at a regional scale in humid and arid

areas. In addition to the above two regression models, RF may be also useful to construct the relationships between precipitation and geospatial predictors. In this study, we compared these three statistical techniques for modeling precipitation, and examined how the model selection affects the prediction accuracy (Figures A2–A4). Pre-processing has been used to model precipitation. The predictors have been centered and scaled, and transformed to a smaller sub-space with principal component analysis (PCA), where the new predictors are uncorrelated with one another. PCA transforming, centering and scaling of predictors were performed using in-house routines developed for R statistical software [57,58] and its caret package [59]. Finally, the pre-processing progress changes the column names to PC1, PC2, etc.

Similar to the processes proposed by Immerzeel [33] and by Jia *et al.* [34], we designed the aggregation procedure in which several pre-processing steps also had to be carried out, as explained in the following section (Figure 2).



**Figure 2.** Flow chart of the downscaling algorithm used in the study.

(1) Pre-processing of the DEM and NDVI related datasets

Elevation, slope, and aspect at the 30 m resolution were extracted from the DEM of Mainland China and resized into a 1 km resolution by applying the pixel averaging technique. NDVI and NDVI related variables also are resized to a 1 km resolution by applying pixel averaging. Additionally, we extracted the annual average NDVI, annual maximum NDVI, annual minimum NDVI, and annual range of NDVI, TRMM 3B43-derived annual precipitation, elevation, slope, aspect, and latitude and longitude at 0.25 ° resolution by pixel averaging, respectively.

## Regression modeling and prediction

Three statistical regression models (multiple linear, exponential, and RF) were then respectively constructed and tested at 0.25 ° scales. Here, independent predictors of the three models were annual average NDVI, annual maximum NDVI, annual minimum NDVI, annual range of NDVI, elevation, slope, aspect, latitude, and longitude. We then predicted the annual precipitation fields at a 1 km resolution by implementing the optimal coefficients of three statistical models to the geospatial predictors at a 1 km resolution ( $P_{prt.}$ ).

### (2) Generating final precipitation estimates

Part of the precipitation variability cannot be explained by the regression models. Those parts were generated by computing the difference between the predicted values at the 0.25 ° scale and the original TRMM 3B43, and then spatially interpolated into the residual (Pres.) maps at a 1 km resolution using a simple spline tense technique. Because of that, the residual data are regular-spaced and the spline technology is usually used for this type of data [33], and the initial testing analysis shown that this kind interpolator performed better than other interpolators. The final precipitation estimates (P) of Mainland China at a 1 km resolution was obtained by adding the residual terms to the predicted terms (Equation (4)).

$$P = P_{prt.} + P_{res.} \quad (4)$$

It should be noted that we excluded the precipitation values over water bodies and urban areas in the downscaling processes, because the three regression models generally produced no significant relationships between precipitation and geospatial predictors over water bodies and urban areas.

The gaps between final precipitation estimates (P) and measured precipitation at each rain gauge were filled by calculating the difference between them. The difference was summarized as follows [24,25,60,61]:

$$error = P_{station} - P \quad (5)$$

### 3.2.1. Linear Regression Model

Jia *et al.* [34] investigated valid relationships of precipitation with NDVI and elevation in a semi-arid area of Qilian Mountain, China, in the process of downscaling the TRMM 3B43 product. Similar to their work, we constructed multiple linear regression models at different scales and applied them to the downscaling process over the continental area of China. The regression linear model is written as follows (Equation (6)):

$$y = a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (6)$$

where  $a_1, a_2, \dots, a_n$  are the linear regression coefficients.

### 3.2.2. Exponential Regression Model

Immerzeel [33] shows a strong coefficient of the correlation between NDVI and TRMM precipitation in Iberian Peninsula. Based on the Immerzeel's method, we built an exponential regression model to downscale the TRMM 3B43 data of Mainland China (Equation (7)).

$$y = a \cdot e^{bx} + c \quad (7)$$

where  $a$ ,  $b$ ,  $c$  are the fitting coefficients of the regression model.

### 3.2.3. Random Forest Model

We employed RF to downscale the TRMM 3B43 precipitation at six different scales using geospatial predictors, including DEM, NDVI related variables, latitude, longitude, aspect, and slope at a 1 km resolution.

Random Forest (RF) is developed by Breiman [43], based on the classification and regression trees (CART) algorithm. For regression in CART, vector  $Y$  represents the response values for each observation in variable matrix  $X$  [19]. The matrix  $X$  and vector  $Y$  can be split into different subsets to regress a tree with a certain number of nodes. In each of the terminal nodes of the tree, a simple and accurate model is built to explain the relationship of  $X$  and  $Y$ . If the regression tree is built with sampled data, the tree then is able to predict another  $Y$ .

Splitting in regression trees is made in accordance with squared residuals minimization algorithm, which implies that expected sum variances for two resulting nodes should be minimized [19] (Equation (8)).

$$\arg \min_{x_j \leq x_j^R, j = 1, \dots, M} [P_l \text{Var}(Y_l) + P_r \text{Var}(Y_r)] \quad (8)$$

where  $P_l$ ,  $P_r$  are probabilities of left and right nodes;  $\text{Var}(Y_l)$ ,  $\text{Var}(Y_r)$  are response vectors for corresponding left and right child nodes;  $x_j \leq x_j^R, j = 1, \dots, M$  is optimal splitting satisfying the condition.

Averaging estimates is one way to reduce the variance of an estimate [62] (Equation (9)).

$$f(x) = \sum_{m=1}^M \frac{1}{M} f_m(x) \quad (9)$$

where  $f_m$  is the  $m$ th tree. This is called bagging [43], which stands for the procedure of bootstrap aggregating.

However, simply re-running the same learning algorithm on different subsets of the data can result in highly correlated predictors. This may limit the amount of variance reduction that is possible. Thus, Breiman [43] proposed a new technique called Random Forest (RF) to decorrelate the base predictors by learning trees on a randomly chosen subset of input variables.

The RF regression algorithm performs as follows [63]: firstly,  $n$  tree bootstrap samples  $X_i$  ( $i$  is bootstrap iteration) were randomly draw with replacement from the training dataset. The elements not included in  $X_i$  are referred to as out-of-bag data (OOB) for that bootstrap sample; then, for each bootstrap sample, an un-pruned regression tree was grown with the modification that at each node,  $m$  try of the predictors, were randomly sampled and the best split from among those variables chosen. Finally, new data (out-of-bag elements) were predicted by averaging predictions of the  $n$  tree trees. The out-of-bag (OOB) samples in the training data were used to estimate prediction error, in which, the OOB samples were predicted by the respective trees and by aggregating the predictions. The out-of-bag (OOB) estimate of the error rate (ERROOB), were calculated as [64]:

$$MSE_{oob} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{f}_{oob}(x_i))^2 \quad (10)$$

where  $\hat{f}_{oob}(x_i)$  is the out-of-bag prediction observation  $i$ .

The orderings of the variable importance is an important issue in problems selecting variable by interpretation issues. The RF algorithm can also provide a measurement of variable importance by looking at how much the prediction error increases when left-out OOB data for that specific predictor variable is permuted while keeping the values of other predictors unchanged. These variable importance values are then used to rank orderings of those independent variables in term of their contributions to the regression model.

Here we used the R package of RF to model precipitation with geospatial predictors [58]. Since precipitation is the continuous value, the RF regression tree approach was implemented in the study. We established “RF models” with the original TRMM 3B43 and other predictor data at six different scales in order to choose the best-performed RF model. In each RF model construction, the tree number of each “RF forest” was set as 500, and we conducted an iterative sampling process where each observation in the input dataset has an equal chance to be selected. A total of six geospatial predictors (NDVI, elevation, slope, aspect, latitude, and longitude) were used to grow “RF trees”.

In this study, the RF regression models were developed for both sub-regions and the entire study area. Annual precipitation at a 1 km resolution for the period 2001–2010, for the entire study area, was obtained by implementing the nation-wide RF model. On the other hand, we divided the whole study area into four zones: arid, semi-arid, semi-humid, and humid zones, depending on annual precipitation in the past regionalization studies of Mainland China [65,66]. The reasons for developing the sub-regional regression models were two-fold. First was to eliminate, as far as possible, the uncertainty caused by the nation-wide RF model when applied to each sub-region. The second reason was to analyze the importance of input predictor variables in the downscaling procedures across different regions.

### 3.3. Validation

Evaluations of the models and downscaling algorithms were divided into two parts: (a) the two-fold cross validation approach was implemented to investigate the performance of the three statistical models, respectively, and (b) the final downscaled precipitation products were validated by an independent precipitation dataset. The dataset provides the monthly precipitation measures from 596 meteorological stations over Mainland China in the period 2001–2010.

#### 3.3.1. Two-Fold cross Validation

In the two-fold cross validation, we randomly divided the original input data into two sample sets at different scales. The first half data was used to train the models at each scale, while exploring the stability of model performances using the independent test sets (the second half data). For the purpose of model evaluations, we used the  $R^2$  and RMSE. The  $R^2$  between the modeled and TRMM estimates and RMSE at six different scales were calculated using Equations (11) and (12).

$$R^2 = \left( \frac{\sum_{i=1}^n (O_i - \bar{O}) \sum_{i=1}^n (P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (P_i - \bar{P})^2}} \right)^2 \quad (11)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - P_i)^2}{n}} \quad (12)$$

where  $O_i$  is the TRMM precipitation and  $P_i$  is the modeled precipitation of the  $i$ -th pixel,  $n$  is the total number of the TRMM precipitation pixel at each scale, and are the value of the TRMM precipitation and the modeled precipitation, respectively.

### 3.3.2. Validation with Ground Observations

The two-fold cross validation was only used to select the best-performed model at a certain scale to downscale the TRMM 3B43 precipitation. The accuracy of the final downscaled products should be further validated against the *in situ* measurements from the rain gauge stations, based on the  $R^2$ , RMSE, mean absolute error (MAE), and bias. Here, we assume that the ground-measured precipitation could well represent the regional precipitation at a scale of 1 km. The MAE and bias were defined as follows (Equations (13) and (14)):

$$MAE = \frac{\sum_{i=1}^n |O_i - P_i|}{n} \quad (13)$$

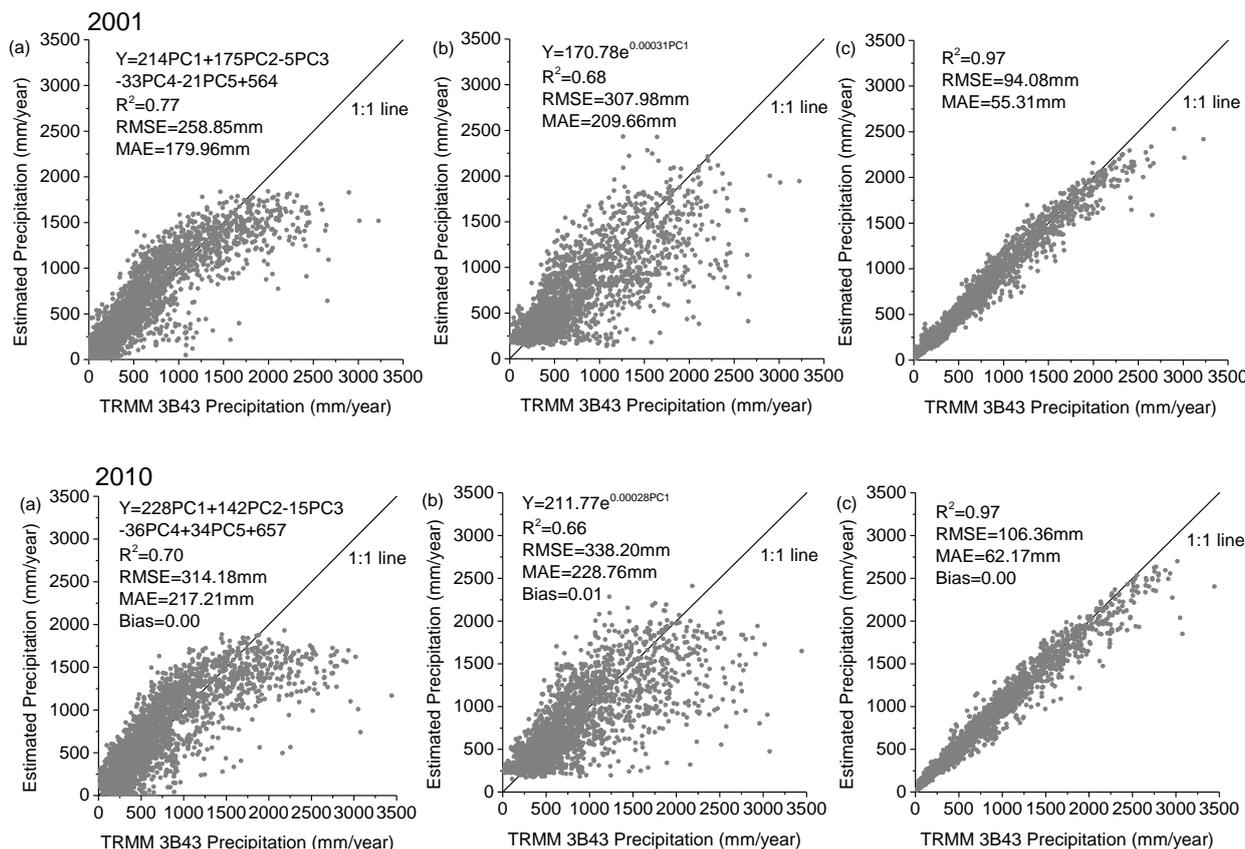
$$Bias = \frac{\sum_{i=1}^n P_i}{\sum_{i=1}^n O_i} - 1 \quad (14)$$

where  $O_i$  indicates the observed annual precipitation from the  $i$ -th rain gauge stations and  $P_i$  is the downscaled annual precipitation extracted at the location of the  $i$ th meteorological stations,  $n$  is the total number of the meteorological stations ( $n = 596$ ), and are the value of the observed and the downscaled precipitation, respectively. Several studies [2,3] demonstrated that the  $R^2$  value is not enough to evaluate the prediction accuracy. Thus, in this study, we gave more weight to the MAE and bias in the evaluations.

## 4. Results and Validation

Figure 3 demonstrates the scatter diagrams for the annual precipitation derived from TRMM 3B43 and the predicted precipitation with the multiple linear, exponential, and RF models at  $0.25^\circ$  scale for 2001 and 2010. All the models qualified the significance test (all the  $p$  values  $< 0.007$  and all the  $R^2 > 0.66$ ). However, for the linear model the underestimations were found where the TRMM precipitation was over 1200 mm/year in 2001 and 2010. For the remaining two models, the underestimations were also observed where the TRMM precipitation was over 2000 mm/yr in 2001 and 2010. All the models seem to dampen potential extremes; a possible reason is that the predictors used in the models, such as NDVI, elevation, are not so sensitive to the extreme rainfall. Extreme rainfall normally occurs in humid area; in this case, the NDVI value will not greatly change with increasing amounts of precipitation as vegetation can only absorb a certain amount of rain water. As shown in

Figure 3, the RF model performed best in terms of the  $R^2$ , RMSE, and MAE in 2001 and 2010, when compared to the two previous models.



**Figure 3.** Scatter plots of the agreements between the annual precipitations of China mainland for the year of 2001 and 2010 derived from the TRMM 3B43 and (a) the multiple linear regression model, (b) exponential regression model, and (c) random Forest model at spatial resolutions of 0.25 °, respectively.

The residual of the regression model was generated by subtracting the predictive value from the original TRMM 3B43, which represented the amount of precipitation that could not be explained by the regression model. The performance of the regress models, thus, would impact on the residual value. We calculated the relative weight of the predicted and residual rainfall fields in terms of the following equation:

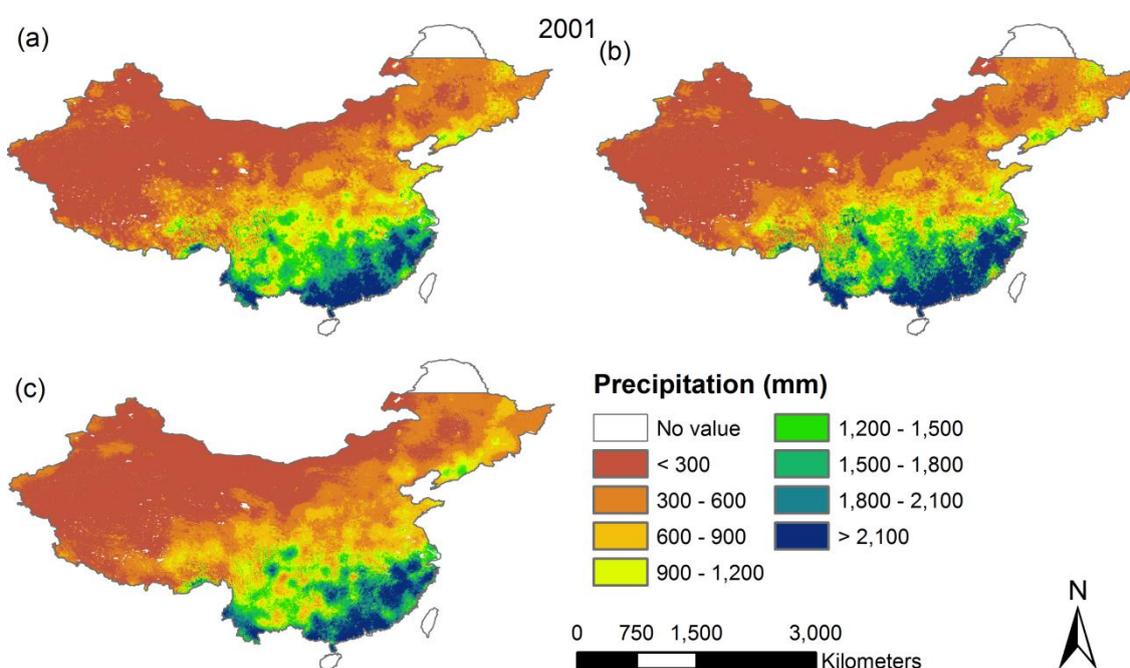
$$\begin{aligned}
 RW_{prd} &= P_{prd} / (P_{prd} + P_{res}) \\
 RW_{res} &= P_{res} / (P_{prd} + P_{res})
 \end{aligned}
 \tag{15}$$

The results for the three techniques are listed in the following table (Table 2). Table 2 implies the predicted rainfall derived with the RF method is very close to the final results. The addition of the residual term seems not to greatly improve the prediction accuracy of the results of the RF downscaling approach for the RF method. However, substantial residuals are found for the linear and exponential models, the ignorance of those residual values would finally reduce the accuracy of downscaling.

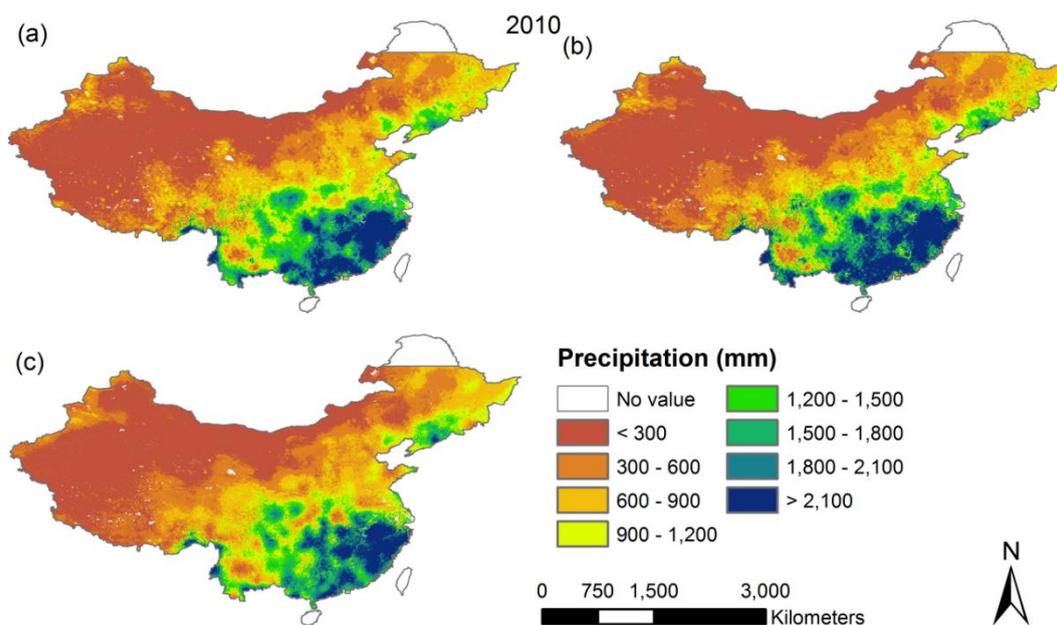
**Table 2.** The relative weight of the predicted and residual annual rainfall fields for the three techniques for the year of 2001 and 2010 over Mainland China.

		Linear Regression	Exponential Regression	RF Regression
2001	RW_prd.	2.16	3.44	1.14
	RW_res.	-1.17	-2.44	-0.14
2010	RW_prd.	2.14	2.01	1.07
	RW_res.	-1.14	-1.01	-0.07

The three models' coefficients were then employed to produce annual precipitation of Mainland China at a 1 km resolution. Here, we also produced the residual map at a 1 km resolution by interpolating residuals using a simple spline tension interpolator, which is the difference between the original TRMM 3B43 and the predicted precipitation at the resolution of 0.25°. The final downscaled product was obtained by combining the predicted values of annual precipitation and the residuals of precipitation at a 1 km resolution (Figure 4). From the final map of downscaled precipitation derived from the three models, we noticed that the distribution of annual precipitation for the period of 2001 and 2010 over Mainland China reflects a strong gradient from the northwest to southeast. Figure 4 reveals that in the wet year and dry year, the exponential method generates more extremes of high precipitation in humid areas, with a significantly larger standard deviation. The probable explanation for this is that the exponential method is markedly more sensitive to predictors of high value in humid areas. The RF method produces the least extremes of high precipitation. Furthermore, the RF downscaling product shows a smoothed pattern and did not present heterogeneous patches when compared to the final products derived from the other two models. The likely explanation is that many more decision trees built in the RF algorithm minimize changes in gradients, giving rise to smoother transitions in surface values across the research area.

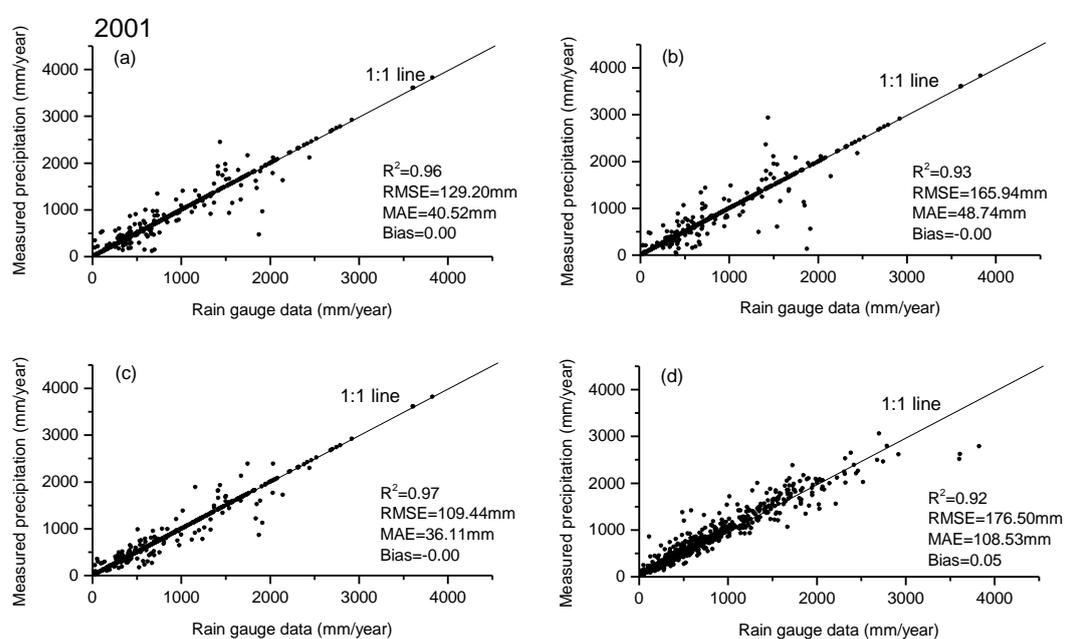


**Figure 4.** Cont.

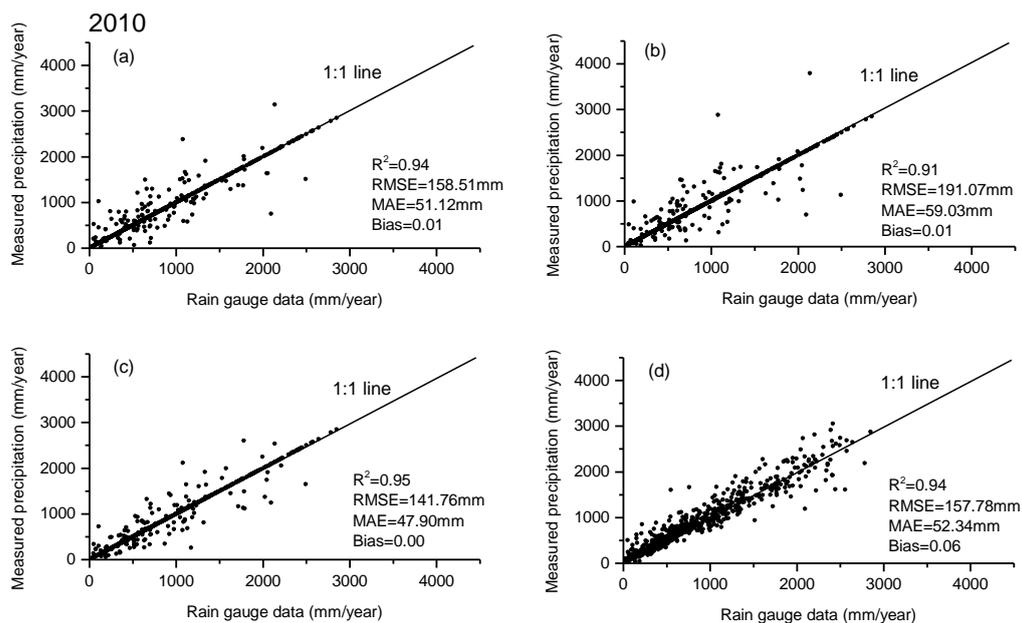


**Figure 4.** The final predicted annual precipitation of Mainland China at a 1 km resolution for the years of 2001 and 2010, using (a) the multiple linear; (b) exponential; (c) Random Forest and regression model, respectively.

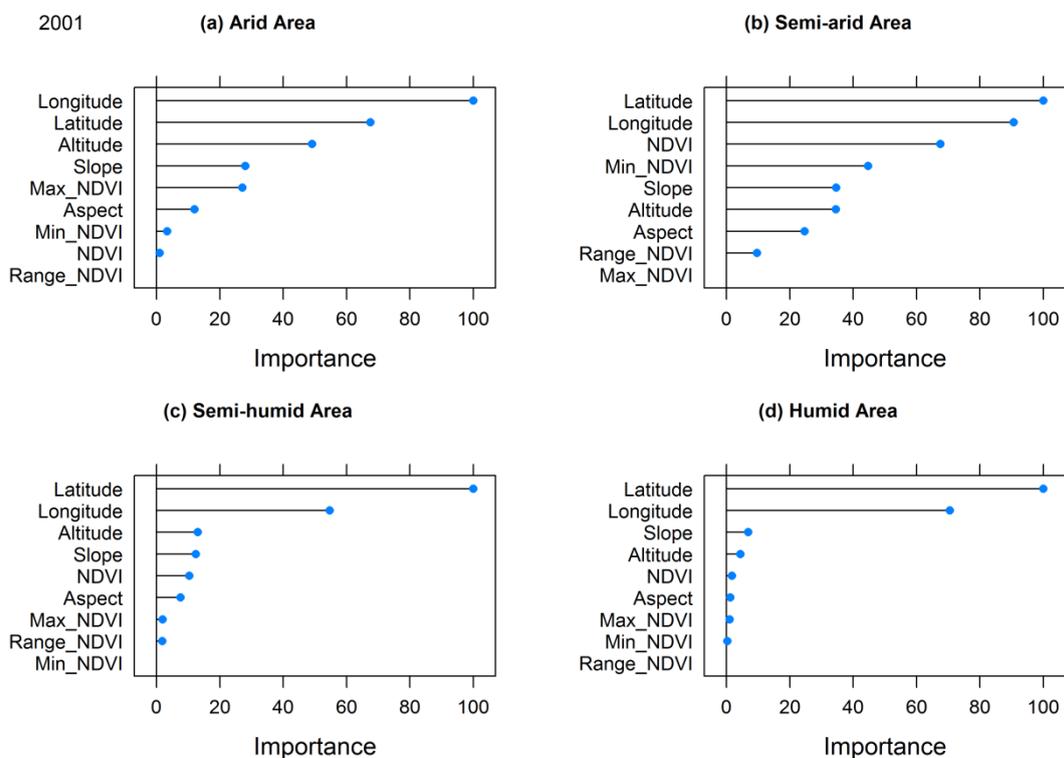
When the downscaled precipitation estimates were compared with the observed station precipitation, the  $R^2$  values are 0.74 in 2001 and 0.75 in 2010 for all the three methods. The efficacy of the three methods was finally validated using the ground observations, collected from 596 meteorological stations of Mainland China for the years of 2001 and 2010 (Figure 5). As seen from Figure 5, the RF downscaling approach slightly outperformed among the three regression models in this study in terms of the statistics of  $R^2$ , RMSE, MAE, and bias. The results of all years from 2001 to 2010 are shown in Figure A5.



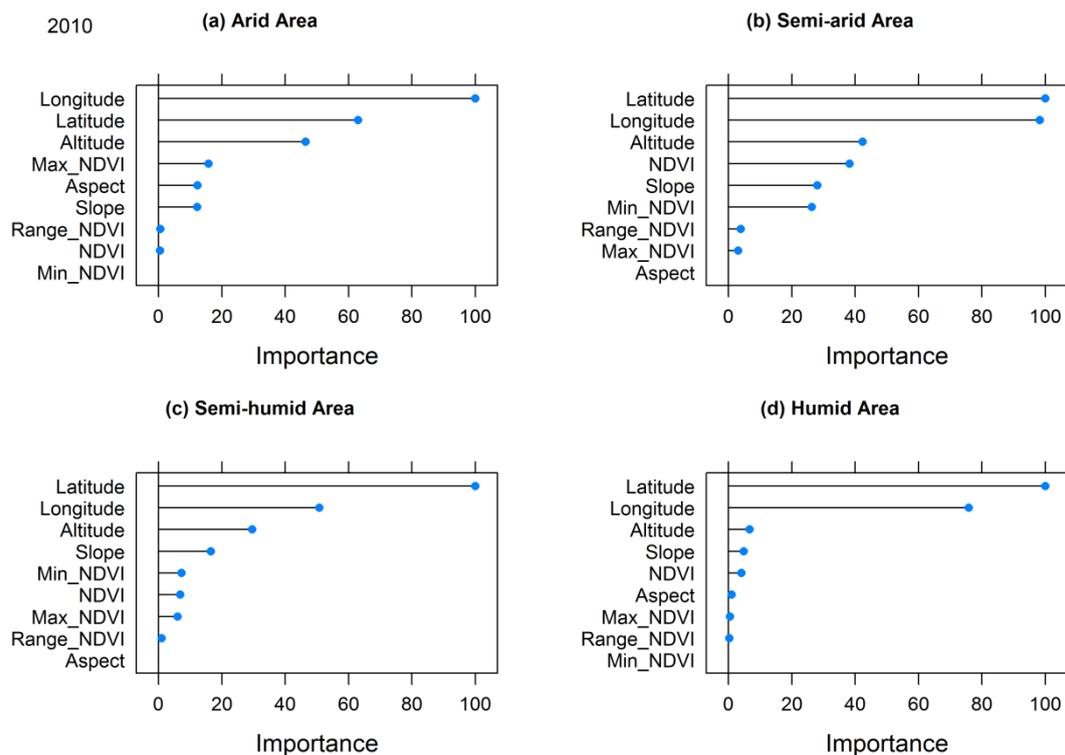
**Figure 5. Cont.**



**Figure 5.** Scatter plot of the measured annual precipitation from 596 meteorology stations versus the predicted precipitation derived from (a) the multiple linear; (b) exponential; (c) Random Forest regression model, and (d) original TRMM 3B43 for the years of 2001 and 2010 over Mainland China, respectively.



**Figure 6.** Cont.



**Figure 6.** Importance of geospatial predictors for precipitation downscaling in (a) arid; (b) semi-arid; (c) semi-humid; and (d) humid regions for the years of 2001 and 2010, illustrated by the MDG of attributes as assigned by RF.

The importance of input predictors for the RF model is illustrated in Figure 6. It shows that all nine geospatial predictors play corresponding roles across four climate zones over Mainland China. NDVI-related predictors, such as NDVI, Max\_NDVI, and Min\_NDVI, represents more importance in arid, semi-arid, and semi-humid areas than in humid areas. Two input predictors, latitude and longitude, make steady contributions to the model simulations unlike the other factors, including altitude, aspect, and slope in four climate zones. The topography-related predictors (altitude, aspect, and slope) were also significant factors for the precipitation downscaling in some areas.

## 5. Discussion

In this study, three statistical regression approaches were investigated to downscale the TRMM 3B43 precipitation data, and to derive annual precipitation at a 1 km resolution over continental China. We compared the multiple linear, exponential, and RF regression models to better understand how the choice of model type affected the prediction accuracy. We found that (a) the RF regression outperformed, with better  $R^2$ , RMSE, MAE, and bias values, when compared to the other two statistical regressions; (b) the RF downscaling can well predict annual precipitation of continental China at a 1 km resolution; and (c) NDVI-related variables, latitude and longitude, and elevation are key elements of downscaling method on estimation of annual precipitation from the analyses of variable importance.

We compared the performance of three statistic models in the process of downscaling TRMM 3B43 precipitation and found that RF regression achieved the best results in modeling annual total precipitation in a broad area. It was supported by cross validation of the three models. The RF regression model

consistently performed best (Figure 3) both for the wet year and dry year in term of  $R^2$  (0.97 for two years) and RMSE (94 mm for 2001, and 106 mm for 2010), when compared to the multiple linear regression model ( $R^2$  is 0.77 for 2001 and 0.70 for 2010, RMSE is 258 mm, 314 mm), and exponential regression model ( $R^2$  is 0.68 and 0.66, RMSE is 307 mm and 338 mm). Unlike previous studies [24,34,39], which only used one or two independent variables (NDVI and elevation), the three methods used nine different predictors in this study to establish the regression models. Those newly introduced predictors seem to be necessary for well-fitting regression models in a broad spatial extent since there are relatively weak empirical relationships between precipitation and the saturated NDVI in humid regions. Despite being the most accurate among three statistical models for predicting precipitation and it exhibited considerable improvement from fitting, there are still certain limitations in the RF algorithm. For example, other uncontrolled variables, such as hydrological conditions and human activities, might influence the distribution of NDVI and, hence, cause some errors in the final results; the processing of variables aggregation (e.g., NDVI, elevation, slope, aspect of the terrain) had not taken into account the uncertainties derived from resampling and re-projection of maps and data. In addition, other unknown predictor variables may also contribute to the efficacy of the RF model, thus further investigation would be needed to confirm which and how many factors are related to the total annual precipitation in Mainland China.

From Figure 3, we can see that all the models seemingly perform better for the areas characterized by low amounts of rainfall. A possible reason may be that NDVI-related predictors are better indicators of precipitation in arid and semi-arid areas. The NDVI values would not increase with the increased rainfall amount in humid areas, which makes a relatively weak empirical relationship between precipitation and saturated NDVI. This could also be seen from the analysis of variable importance (Figure 6). However, substantial difference of the precipitation maps is found in some regions of Mainland China. The area with extreme precipitation predicted by the RF model is smaller than that predicted by multiple linear, and much smaller than that by the exponential model. The differences among these maps are possibly caused by the added residual terms, or by the performance of the models. Normally, the model's performance in the validation period is the best indicator as to its ability to predict the future. If a regression model, such as the RF model, can make accurate predictions of precipitation ( $R^2 > 0.97$ ), the predicted values will be very close to the original TRMM precipitation, and the residual values will be small. The relative weights of the residuals will also become small (Table 2). Especially in South Mainland China, where a large amount of precipitation (annual precipitation greater than 1200 mm/yr) is received, the relative weight of the residual calculated from the RF model will be significantly smaller than that from the other two models.

The downscaling algorithm with three regression models can well predict annual precipitation of continental China at a 1 km resolution. In terms of RMSE, MAE, and bias, the downscaling approach with the RF model produced the lowest amount of errors compared to the other two statistic regression models when validated using precipitation dataset collected from the meteorological stations for both the wet year and dry year (Figure 5). The general precipitation patterns are well captured in this estimate by all three regression models. The northwestern part of the Chinese mainland is drier, and the southeastern part and the southern coast are clearly wetter. Those patterns satisfactorily resemble the general spatial distribution patterns of annual precipitation of Mainland China [67,68].

Through importance analysis, we found that latitude and longitude are very important factors in simulating the annual precipitation over Mainland China. Rainfall in the study area mainly results from the moisture-laden air masses from the Western Pacific and the Indian Ocean. Latitude and longitude may represent the relative distance from the Pacific and Indian Oceans, and, thus, significantly affect precipitation and its spatial distribution. NDVI is generally considered as a powerful predictor for precipitation [24,34,39], but our study resulted in correlations only in arid, semi-arid, and semi-humid zones (annual precipitation less than 800 mm/yr). The higher NDVI does not always represent more precipitation in humid zones because of the saturated NDVI. This leads to a declined importance of NDVI-related predictors in humid areas.

The elevation, another geospatial predictor in downscaling approaches [24,34,39], was used as an indispensable predictor in four climate zones over continental China. Interestingly, the elevation was even more important than the NDVI in humid, semi-arid, and arid areas due to the uplift precipitation effects of mountains and/or the saturated NDVI. Moreover, the importance of the aspect and slope steadily ranked elevation, and these two predictors certainly contributed to the prediction accuracy of precipitation. This is because aspect is linked to the prevalent wind orientation, and, thus, determines the potential relative water excess or deficit; a gradient in the speed of vertical air movements may control the intensity and area of precipitation [69,70].

Since the RF technique only pursues the valid regressions between precipitation and geospatial predictors, the geo-physical mechanism is actually not well represented in this approach. Thus, we are still not able to give the explicit mechanistic understanding of precipitation and related predictors. More investigations about the geo-physical mechanism in relation to precipitation would be needed to understand the feedbacks between precipitation and those predictors. This would be also helpful to explore additional geospatial predictors that may improve the RF algorithm. Furthermore, our downscaling framework is currently not applicable to shorter temporal periods (e.g., monthly or seasonal). In the future, monthly or seasonal precipitation can be estimated using, for instance, the fraction method proposed by Duan and Bastiaanssen [24]. It should be also noted that there is still a scale gap between rain-gauge data and final rainfall product in this study because a rain gauge typically collects rainfall at ground level with a sample area of roughly 50 cm<sup>2</sup> [71]. It is, therefore, suggested that gap be taken into account in future studies, especially at finer time scales.

## 6. Conclusions

This study explored three spatial downscaling methods implementing satellite-based TRMM 3B43 precipitation and a number of geospatial predictors over a complex territory. The ultimate goal was to generate the annual precipitation over continental China at a 1 km resolution, from 2001 to 2010. The results of this investigation show that, (1) the regression model based on Random Forest performs best among all three models investigated when modeling TRMM precipitation at coarse resolution; (2) all the downscaled results seemingly captured the spatial patterns of annual precipitation of Mainland China, and the validations against the *in situ* measurements from rain gauge stations showed that the RF approach is better than the other two approaches in terms of RMSE, MAE, and bias; (3) NDVI (and NDVI-related variables) is not an important predictor in the humid areas. We believe that our framework may be useful to produce fields of annual precipitation at fine spatial resolutions, which is one of the

key input parameters of some biological models [51–53]. Our precipitation production can make contributions to the estimation of tree height and biomass in a study area. However, the mechanistic understanding and temporal resolution is still limited in our framework. Future studies, thus, should account for the geo-physical mechanisms of precipitation and its monthly/seasonal variations for a better understanding of the process of water and energy balances.

### Acknowledgments

This study was partially funded by the National Natural Science Foundation of China (Grants No. 41471312 and 41175077), the Foundation of National Key Laboratory of Remote Sensing of China (OFSLRSS201309). Thanks are given to the Tropical Rainfall Measuring Mission (TRMM), the Moderate Resolution Imaging Spectroradiometer (MODIS) mission, and the Shuttle Radar Topography Mission (SRTM) for their data support. We also thank the anonymous reviewers for their valuable comments and suggestions.

### Author Contributions

Yuli Shi conceived and designed the experiments; Yuli Shi, Lei Song, Zhen Xia and Yurong Lin performed the experiments; Yuli Shi, Lei Song, Zhen Xia and Yurong Lin analyzed the data; Ranga B. Myneni, Sungho Choi, Lin Wang, Xiliang Ni, Cailian Lao, Fengkai Yang contributed materials and analysis tools; Yuli Shi and Sungho Choi wrote the paper.

### Appendix

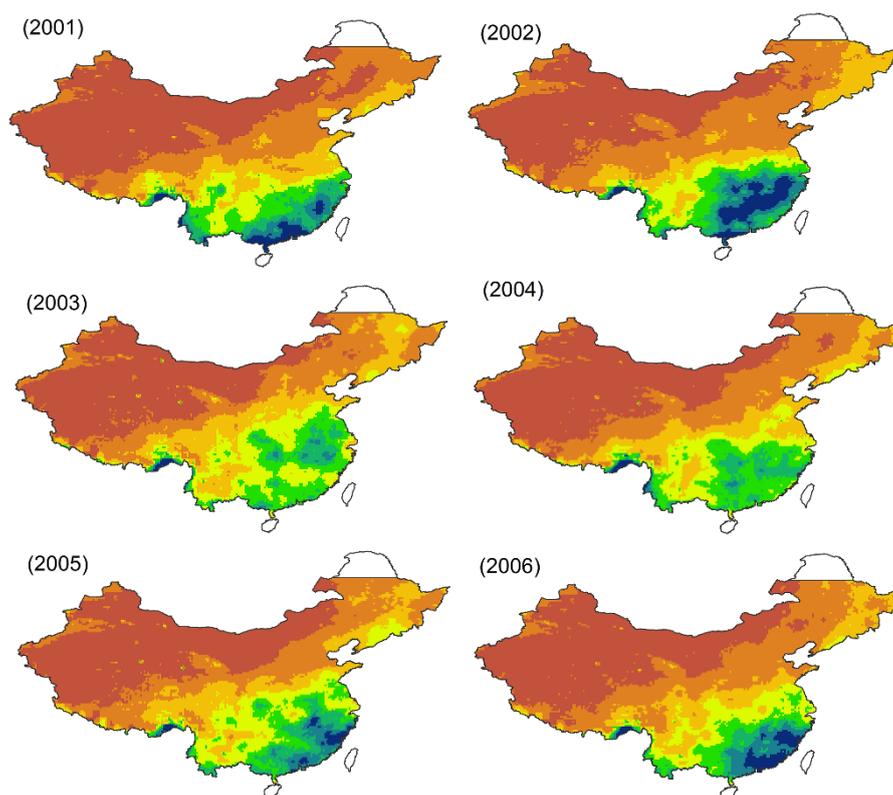
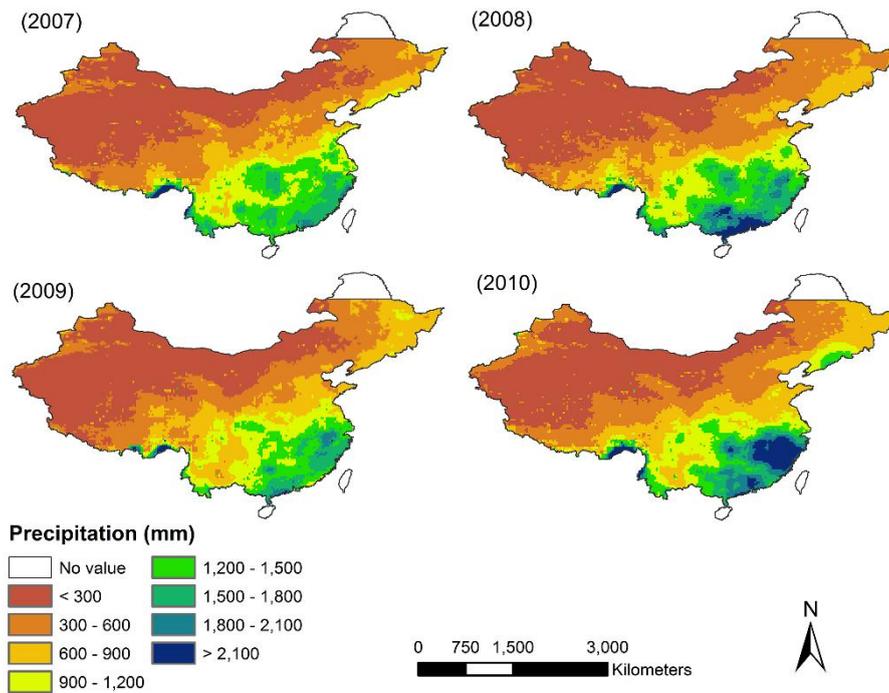
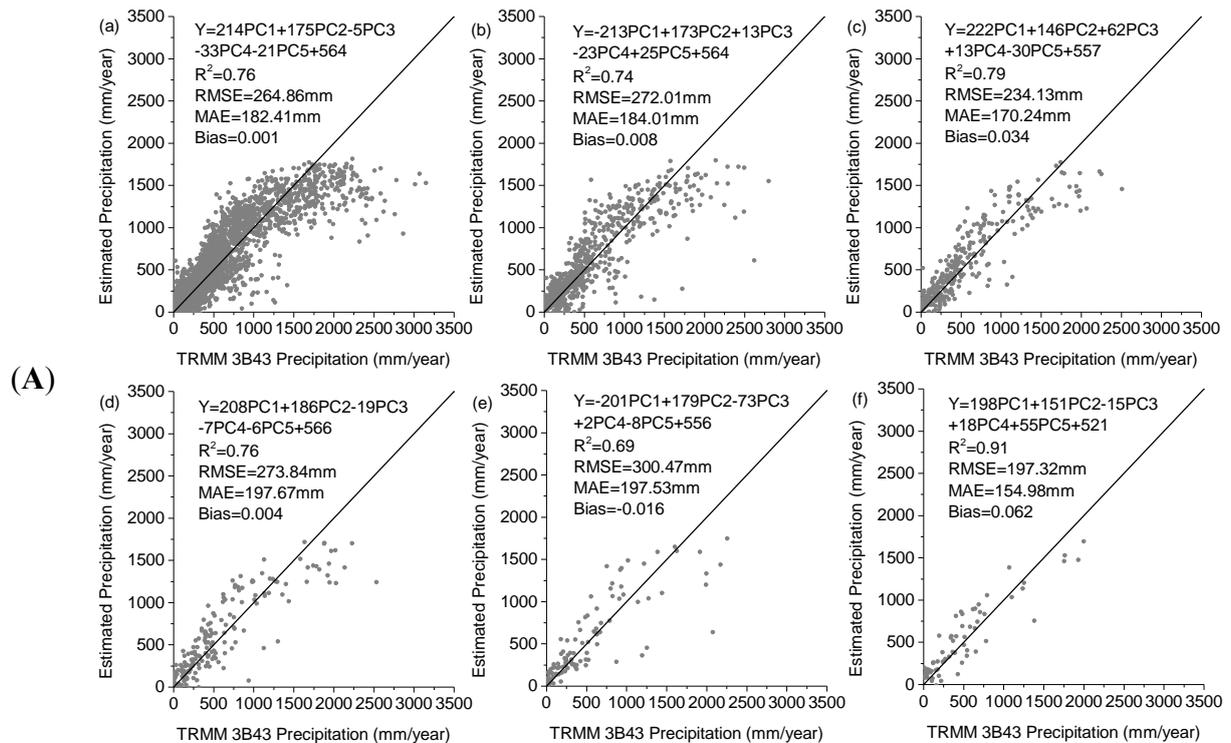


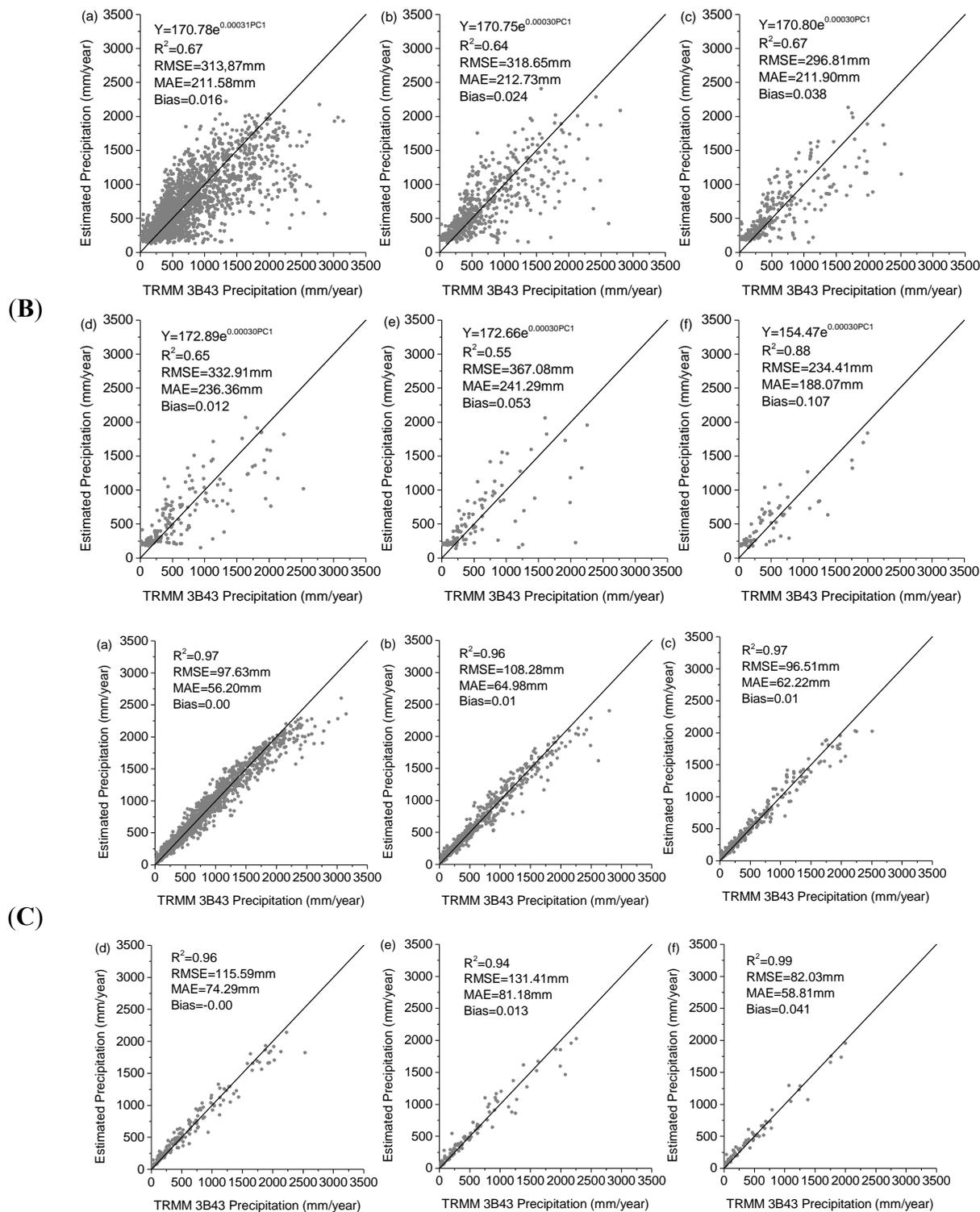
Figure A1. Cont.



**Figure A1.** Annual precipitation of China mainland at spatial resolution  $0.25^\circ \times 0.25^\circ$  for the period years from 2001 to 2010, aggregated from the original monthly TRMM 3B43 V7 precipitation product.



**Figure A2.** Cont.



**Figure A2.** Scatter plots of the agreements between the annual precipitations of China mainland in 2001 derived from the TRMM 3B43 and (A) the multiple linear regression models, (B) exponential models, and (C) random forest models at spatial resolutions of (a) 0.25 °, (b) 0.50 °, (c) 0.75 °, (d) 1.0 °, (e) 1.25 °, and (f) 1.50 °, respectively. Principal component analysis has been used in the multiple linear and exponential regression modeling.

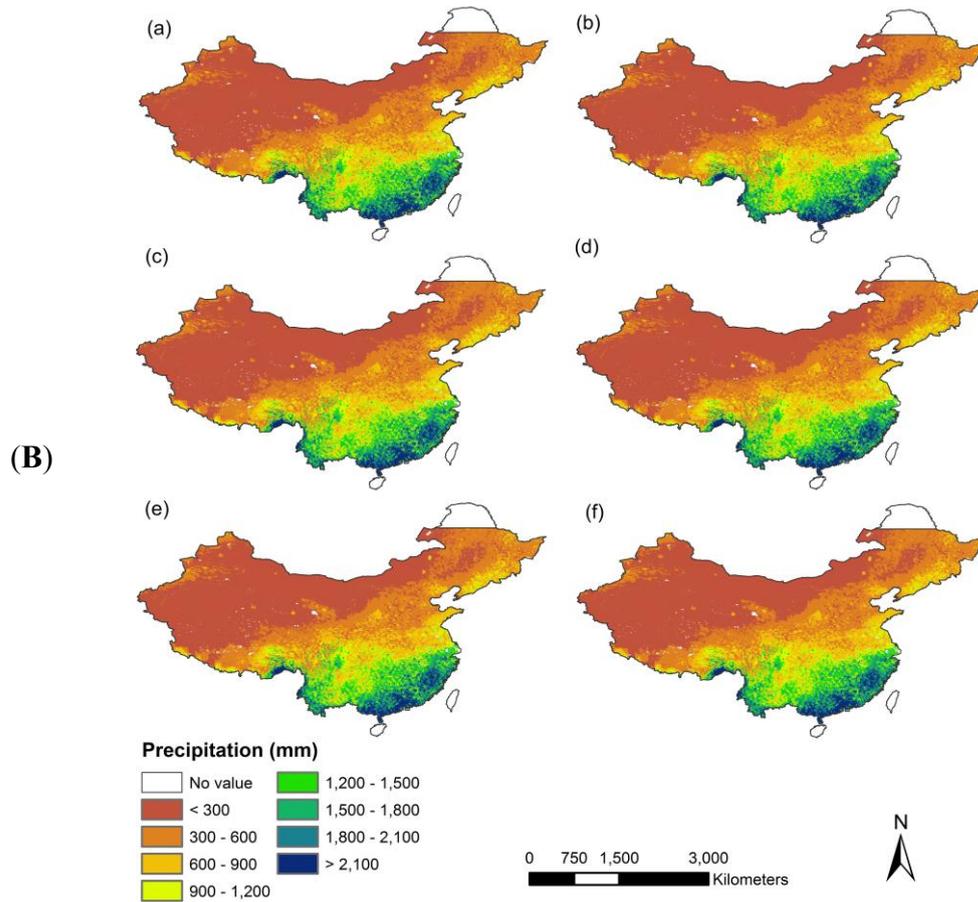
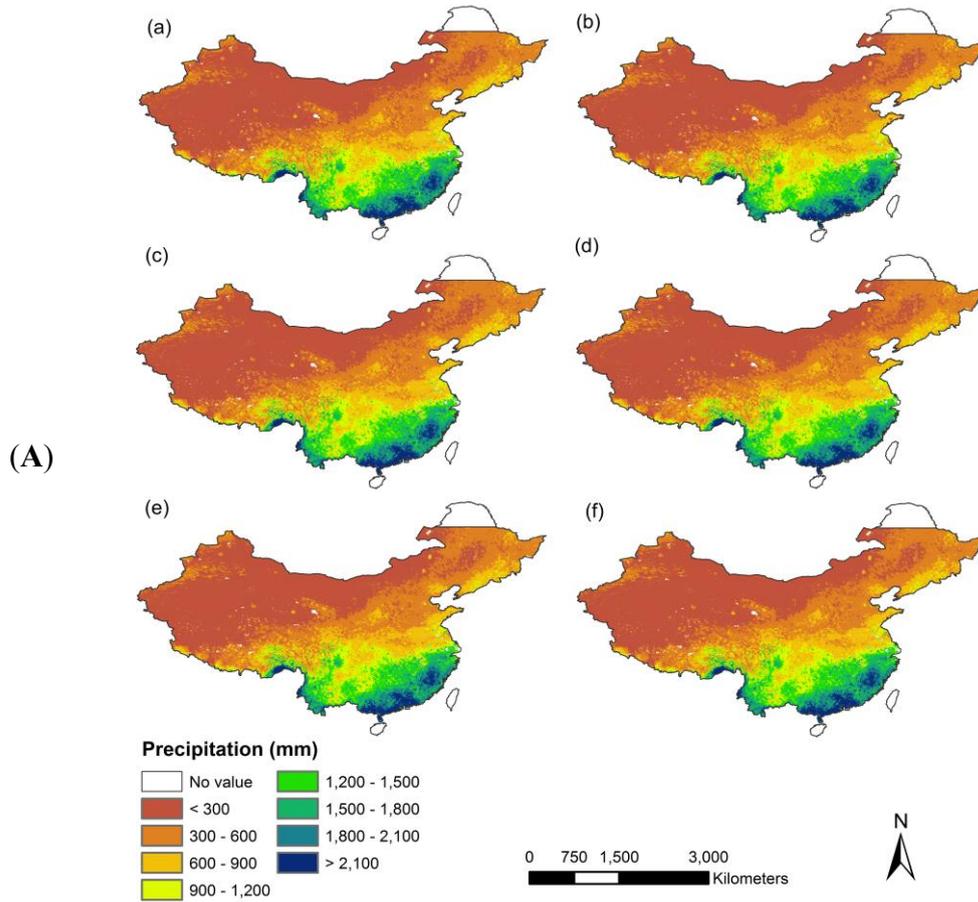
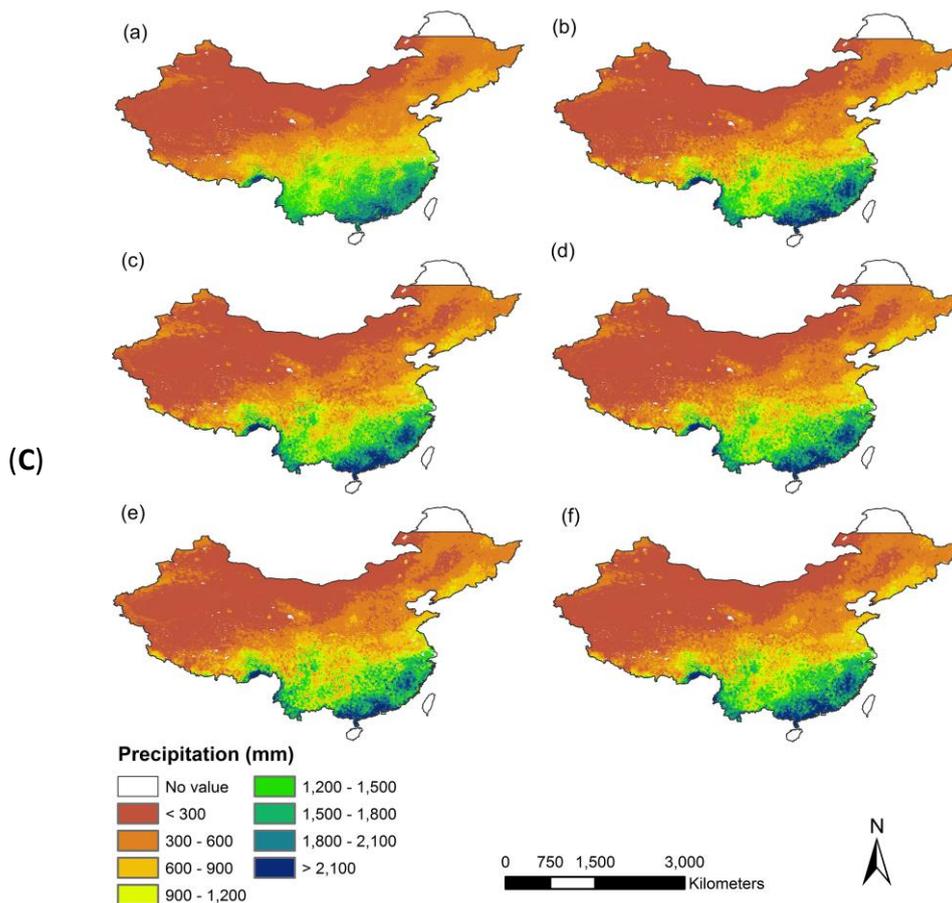
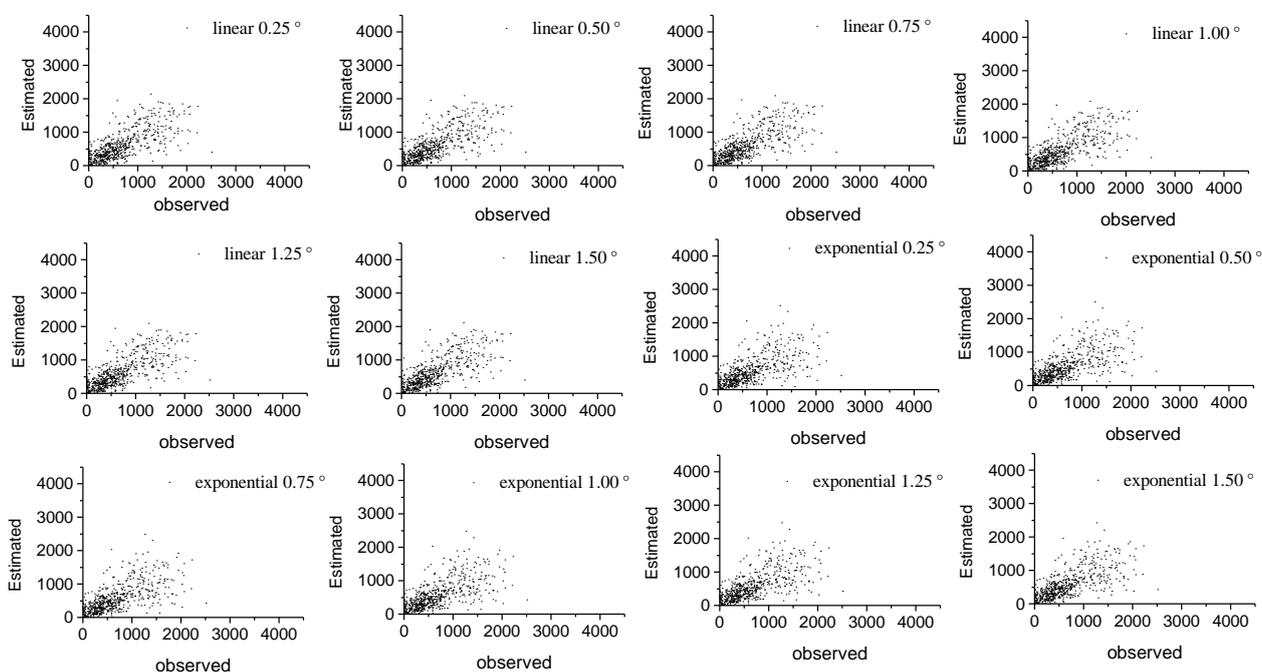


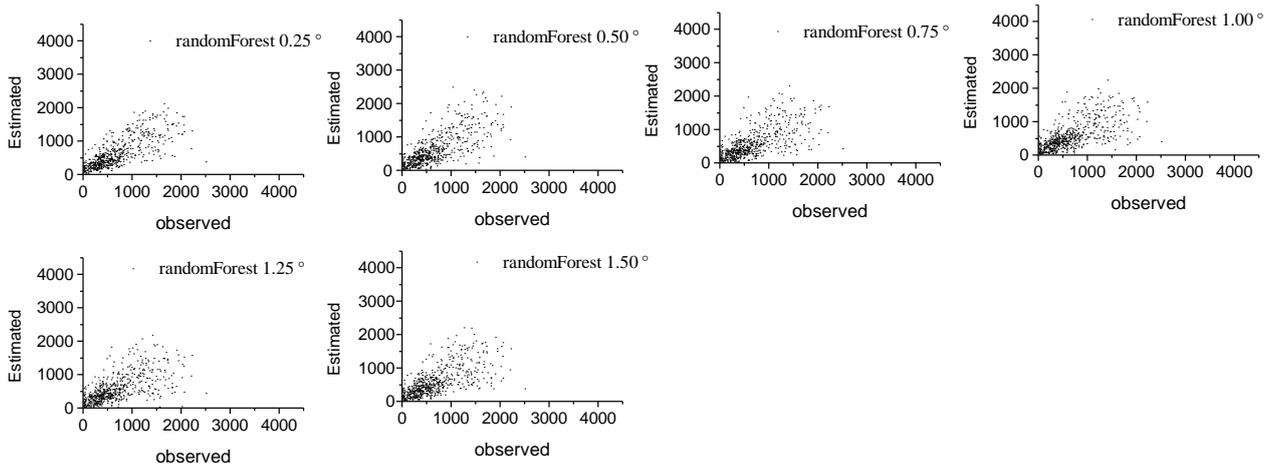
Figure A3. Cont.



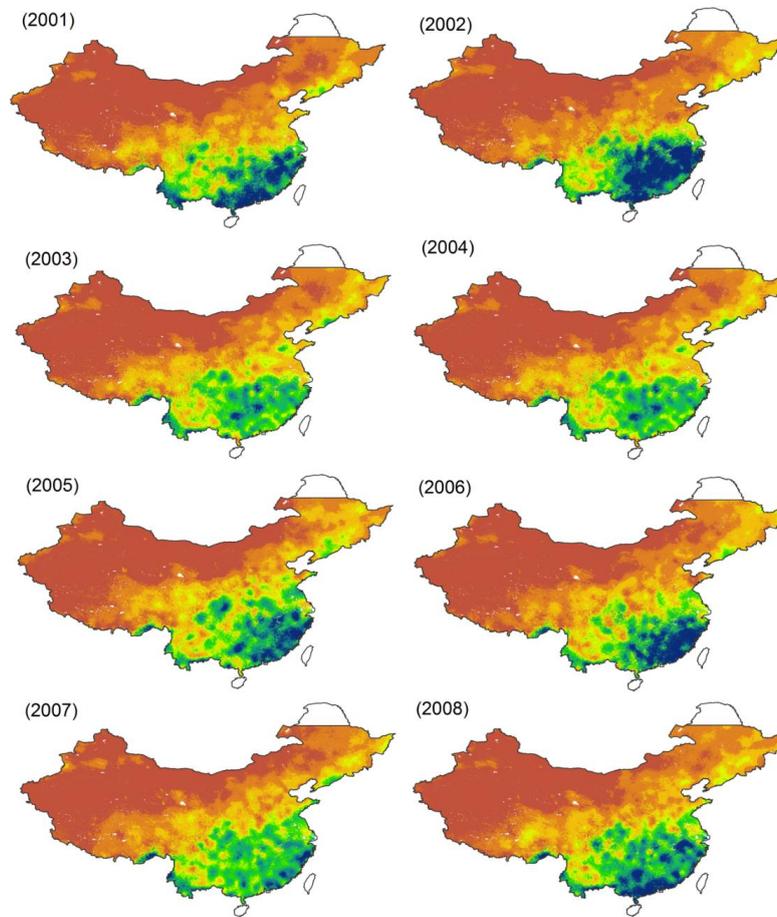
**Figure A3.** The final predicted annual precipitation of China mainland at the 1 km resolution for the year 2001 using the multiple linear regression models (A); exponential models (B); and random forest models (C) built at spatial resolutions of (a) 0.25 °, (b) 0.50 °, (c) 0.75 °, (d) 1.0 °, (e) 1.25 °, and (f) 1.50 °, respectively.



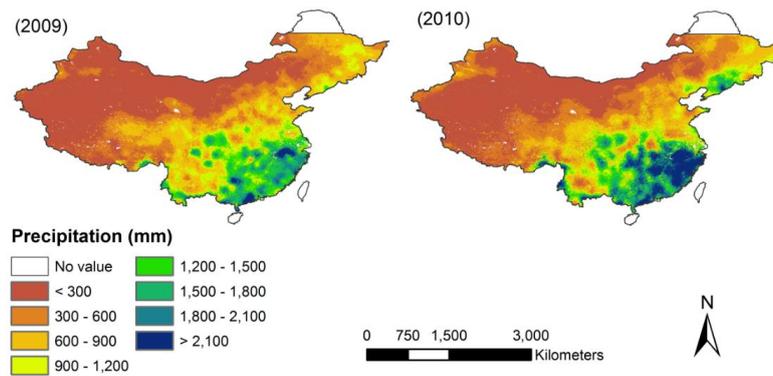
**Figure A4.** Cont.



**Figure A4.** Scatter plot of the measured annual precipitation from 596 meteorology stations versus the predicted precipitation extracted from the final downscaled using the multiple linear regression models, exponential models, and random forest models built at spatial resolutions of 0.25 °, 0.50 °, 0.75 °, 1.0 °, 1.25 °, and 1.50 °, respectively, for the year 2001 over China mainland without gap-filling processing.



**Figure A5.** Cont.



**Figure A5.** The final predicted annual precipitation of China mainland at the 1 km resolution from 2001 to 2010 using random Forest regression model.

### Conflicts of Interest

The authors declare no conflict of interest.

### References

1. Goovaerts, P. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.* **2000**, *228*, 113–129.
2. Langella, G.; Basile, A.; Bonfante, A.; Terribile, F. High-resolution space-time rainfall analysis using integrated ANN inference systems. *J. Hydrol.* **2010**, *387*, 328–342.
3. Li, M.; Shao, Q. An improved statistical approach to merge satellite rainfall estimates and raingauge data. *J. Hydrol.* **2010**, *385*, 51–64.
4. Goodrich, D.; Faurès, J.; Woolhiser, D.; Lane, L.; Sorooshian, S. Measurement and analysis of small-scale convective storm rainfall variability. *J. Hydrol.* **1995**, *173*, 283–308.
5. McGuffie, K.; Henderson-Sellers, A.; Holbrook, N.; Kothavala, Z.; Balachova, O.; Hoekstra, J. Assessing simulations of daily temperature and precipitation variability with global climate models for present and enhanced greenhouse climates. *Int. J. Climatol.* **1999**, *19*, 1–26.
6. Xie, P.; Arkin, P.A. Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Am. Meteorol. Soc.* **1997**, *78*, 2539–2558.
7. Michaelides, S.; Levizzani, V.; Anagnostou, E.; Bauer, P.; Kasparis, T.; Lane, J.E. Precipitation: Measurement, remote sensing, climatology and modeling. *Atmos. Res.* **2009**, *94*, 512–533.
8. Giorgi, F.; Marinucci, M.R.; Bates, G.T. Development of a second-generation regional climate model (RegCM2). Part I: Boundary-layer and radiative transfer processes. *Mon. Weather Rev.* **1993**, *121*, 2794–2813.
9. Huffman, G.J.; Adler, R.F.; Arkin, P.; Chang, A.; Ferraro, R.; Gruber, A.; Janowiak, J.; McNab, A.; Rudolf, B.; Schneider, U. The Global Precipitation Climatology Project (GPCP) combined precipitation dataset. *Bull. Am. Meteorol. Soc.* **1997**, *78*, 5–20.
10. Res, C.; Daly, C.; Gibson, W.P.; Taylor, G.H.; Johnson, G.L.; Pasteris, P. A knowledge-based approach to the statistical mapping of climate. *Climate Res.* **2002**, *22*, 99–113.
11. Thornton, P.E.; Running, S.W.; White, M.A. Generating surfaces of daily meteorological variables over large regions of complex terrain. *J. Hydrol.* **1997**, *190*, 214–251.

12. Steiner, M.; Smith, J.A.; Burges, S.J.; Alonso, C.V.; Darden, R.W. Effect of bias adjustment and rain gauge data quality control on radar rainfall estimation. *Water Resour. Res.* **1999**, *35*, 2487–2503.
13. Xie, P.; Chen, M.; Yang, S.; Yatagai, A.; Hayasaka, T.; Fukushima, Y.; Liu, C. A gauge-based analysis of daily precipitation over east Asia. *J. Hydrometeorol.* **2007**, *8*, 607–626.
14. Huffman, G.J.; Adler, R.F.; Bolvin, D.T.; Gu, G. Improving the global precipitation record: GPCP version 2.1. *Geophys. Res. Lett.* **2009**, *36*, 1–5.
15. Huffman, G.J.; Adler, R.F.; Morrissey, M.M.; Bolvin, D.T.; Curtis, S.; Joyce, R.; McGavock, B.; Susskind, J. Global precipitation at one-degree daily resolution from multisatellite observations. *J. Hydrometeorol.* **2001**, *2*, 36–50.
16. Huffman, G.J.; Bolvin, D.T.; Nelkin, E.J.; Wolff, D.B.; Adler, R.F.; Gu, G.; Hong, Y.; Bowman, K.P.; Stocker, E.F. The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.* **2007**, *8*, 38–55.
17. Sheffield, J.; Goteti, G.; Wood, E.F. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Clim.* **2006**, *19*, 3088–3111.
18. Huffman, G.J.; Bolvin, D.T.; Braithwaite, D.; Hsu, K.; Joyce, R.; Xie, P.; Yoo, S.-H. NASA Global Precipitation Measurement (GPM) Integrated Multi-Satellite Retrievals for GPM (IMERG). Available online: [http://pmm.nasa.gov/sites/default/files/document\\_files/IMERG\\_ATBD\\_V4.4.pdf](http://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V4.4.pdf) (accessed on 15 September 2014).
19. Timofeev, R. Classification and Regression Trees (CART) Theory and Applications. Available online: <http://edoc.hu-berlin.de/master/timofeev-roman-2004-12-20/PDF/timofeev.pdf> (accessed on 20 December 2004).
20. Tian, Y.; Peters-Lidard, C.D.; Choudhury, B.J.; Garcia, M. Multitemporal analysis of TRMM-based satellite precipitation products for land data assimilation applications. *J. Hydrometeorol.* **2007**, *8*, 1165–1183.
21. Kubota, T.; Shige, S.; Hashizume, H.; Aonashi, K.; Takahashi, N.; Seto, S.; Hirose, M.; Takayabu, Y.N.; Ushio, T.; Nakagawa, K.; *et al.* Global precipitation map using satellite-borne microwave radiometers by the GSMAP project: Production and validation. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 2259–2275.
22. Su, F.; Hong, Y.; Lettenmaier, D.P. Evaluation of TRMM multisatellite precipitation analysis (TMPA) and its utility in hydrologic prediction in the La Plata basin. *J. Hydrometeorol.* **2008**, *9*, 622–640.
23. Shi, Y.; Choi, S.; Ni, X.; Ganguly, S.; Zhang, G.; Duong, H.; Lefsky, M.; Simard, M.; Saatchi, S.; Lee, S.; *et al.* Allometric scaling and resource limitations model of tree heights: Part 1. Model optimization and testing over continental USA. *Remote Sens.* **2013**, *5*, 284–306.
24. Duan, Z.; Bastiaanssen, W.G.M. First results from version 7 TRMM 3B43 precipitation product in combination with a new downscaling-calibration procedure. *Remote Sens. Environ.* **2013**, *131*, 1–13.
25. 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.
26. Agam, N.; Kustas, W.P.; Anderson, M.C.; Li, F.; Neale, C.M.U. A vegetation index based technique for spatial sharpening of thermal imagery. *Remote Sens. Environ.* **2007**, *107*, 545–558.

27. Merlin, O.; Duchemin, B.; Hagolle, O.; Jacob, F.; Coudert, B.; Chehbouni, G.; Dedieu, G.; Garatuza, J.; Kerr, Y. Disaggregation of MODIS surface temperature over an agricultural area using a time series of Formosat-2 images. *Remote Sens. Environ.* **2010**, *114*, 2500–2512.
28. Grimes, D.I.F.; Diop, M. Satellite-based rainfall estimation for river flow forecasting in Africa. I: Rainfall estimates and hydrological forecasts. *Hydrol. Sci. J.* **2003**, *48*, 567–584.
29. Merlin, O.; Al Bitar, A.; Walker, J.P.; Kerr, Y. A sequential model for disaggregating near-surface soil moisture observations using multi-resolution thermal sensors. *Remote Sens. Environ.* **2009**, *113*, 2275–2284.
30. Merlin, O.; Al Bitar, A.; Walker, J.P.; Kerr, Y. An improved algorithm for disaggregating microwave-derived soil moisture based on red, near-infrared and thermal-infrared data. *Remote Sens. Environ.* **2010**, *114*, 2305–2316.
31. Zurita-Milla, R.; Kaiser, G.; Clevers, J.G.P.W.; Schneider, W.; Schaepman, M.E. Downscaling time series of MERIS full resolution data to monitor vegetation seasonal dynamics. *Remote Sens. Environ.* **2009**, *113*, 1874–1885.
32. 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.
33. 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.
34. Jia, S.; Zhu, W.; Lü, A.; Yan, T. A statistical spatial downscaling algorithm of TRMM precipitation based on NDVI and DEM in the Qaidam basin of China. *Remote Sens. Environ.* **2011**, *115*, 3069–3079.
35. Lovejoy, S.; Pinel, J.; Schertzer, D. The global space-time cascade structure of precipitation: Satellites, gridded gauges and reanalyses. *Adv. Water Resour.* **2012**, *45*, 37–50.
36. Lovejoy, S.; Schertzer, D.; Allaire, V. The remarkable wide range spatial scaling of TRMM precipitation. *Atmos. Res.* **2008**, *90*, 10–32.
37. Kang, B.; Ramírez, J.A. A coupled stochastic space-time intermittent random cascade model for rainfall downscaling. *Water Resour. Res.* **2010**, doi:10.1029/2008WR007692.
38. Chen, F.; Liu, Y.; Liu, Q.; Li, X. Spatial downscaling of TRMM 3B43 precipitation considering spatial heterogeneity. *Int. J. Remote Sens.* **2014**, *35*, 3074–3093.
39. Verlinde, J. Trmm Rainfall Data Downscaling in the Pangani Basin in Tanzania. Master's Thesis, Delft University of Technology, Delft, The Netherlands, 2011.
40. Marquínez, J.; Lastra, J.; García, P. Estimation models for precipitation in mountainous regions: The use of GIS and multivariate analysis. *J. Hydrol.* **2003**, *270*, 1–11.
41. Ranhao, S.; Baiping, Z.; Jing, T. A multivariate regression model for predicting precipitation in the Daqing mountains. *Mt. Res. Dev.* **2008**, *28*, 318–325.
42. Sanchez-Moreno, J.F.; Mannaerts, C.M.; Jetten, V. Influence of topography on rainfall variability in Santiago island, Cape Verde. *Int. J. Climatol.* **2014**, *34*, 1081–1097.
43. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
44. Falkowski, M.J.; Evans, J.S.; Martinuzzi, S.; Gessler, P.E.; Hudak, A.T. Characterizing forest succession with lidar data: An evaluation for the inland northwest, USA. *Remote Sens. Environ.* **2009**, *113*, 946–956.

45. Simard, M.; Pinto, N.; Fisher, J.B.; Baccini, A. Mapping forest canopy height globally with spaceborne lidar. *J. Geophys. Res. Biogeosci.* **2011**, *116*, 1–12.
46. Thompson, J.R.; Spies, T.A. Vegetation and weather explain variation in crown damage within a large mixed-severity wildfire. *For. Ecol. Manag.* **2009**, *258*, 1684–1694.
47. Qian, W.; Lin, X. Regional trends in recent precipitation indices in China. *Meteorol. Atmos. Phys.* **2005**, *90*, 193–207.
48. Basist, A.; Bell, G.D.; Meentemeyer, V. Statistical relationships between topography and precipitation patterns. *J. Climate* **1994**, *7*, 1305–1315.
49. Omuto, C.T.; Vargas, R.R.; Alim, M.S.; Paron, P. Mixed-effects modelling of time series NDVI-rainfall relationship for detecting human-induced loss of vegetation cover in drylands. *J. Arid Environ.* **2010**, *74*, 1552–1563.
50. Kummerow, C.; Simpson, J.; Thiele, O.; Barnes, W.; Chang, A.T.C.; Stocker, E.; Adler, R.F.; Hou, A.; Kakar, R.; Wentz, F.; *et al.* The status of the tropical rainfall measuring mission (TRMM) after two years in Orbit. *J. Appl. Meteorol.* **2000**, *39*, 1965–1982.
51. Kummerow, C.; Barnes, W.; Kozu, T.; Shiue, J.; Simpson, J. The tropical rainfall measuring mission (TRMM) sensor package. *J. Atmos. Ocean. Technol.* **1998**, *15*, 809–817.
52. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150.
53. Grist, J.; Nicholson, S.E.; Mpolokang, A. On the use of NDVI for estimating rainfall fields in the Kalahari of Botswana. *J. Arid Environ.* **1997**, *35*, 195–214.
54. Iwasaki, H. NDVI prediction over mongolian grassland using GSMAP precipitation data and JRA-25/JCDAS temperature data. *J. Arid Environ.* **2009**, *73*, 557–562.
55. Onema, K.J.-M.; Akpofure, T. NDVI-rainfall relationship in the semliki watershed of the equatorial Nile. *Phys. Chem. Earth Parts A/B/C* **2009**, *34*, 711–721.
56. Li, F.; Zeng, Y.; Li, X.; Zhao, Q.; Wu, B. Remote sensing based monitoring of interannual variations in vegetation activity in China from 1982 to 2009. *Sci. China Earth Sci.* **2014**, *57*, 1800–1806.
57. Team, R.C. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria, 2012. Available online: <http://cran.r-project.org> (accessed on 20 January 2015).
58. Freeman, E.; Frescino, T.; Moisen, G. Modelmap: An R Package for Modeling and Map Production Using Random Forest and Stochastic Gradient Boosting. Available online: <http://cran.r-project.org/web/packages/ModelMap/vignettes/VModelMap.pdf> (accessed on 15 April 2014).
59. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **2008**, *28*, 1–26.
60. Almazroui, M. Calibration of trmm rainfall climatology over Saudi Arabia during 1998–2009. *Atmos. Res.* **2011**, *99*, 400–414.
61. Cheema, M.J.M.; Bastiaanssen, W.G. Local calibration of remotely sensed rainfall from the TRMM satellite for different periods and spatial scales in the Indus basin. *Int. J. Remote Sens.* **2012**, *33*, 2603–2627.
62. Murphy, K.P. Adaptive basis function models. In *Machine Learning: A Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2012; pp. 544–551.
63. Liaw, A.; Wiener, M. Classification and regression by randomforest. *R News* **2002**, *2*, 18–22.

64. Cutler, A.; Cutler, D.R.; Stevens, J.R. Random forests. In *Ensemble Machine Learning*; Springer: Berlin, Germany, 2012; pp. 157–175.
65. Liu, J.Y.; Zhuang, D.F.; Luo, D.; Xiao, X. Land-cover classification of China: Integrated analysis of AVHRR imagery and geophysical data. *Int. J. Remote Sens.* **2003**, *24*, 2485–2500.
66. Zheng, J.; Yin, Y.; Li, B. A new scheme for climate regionalization in China. *Acta Geogr. Sin.* **2010**, *65*, 3–12.
67. Gao, X.; Shi, Y.; Song, R.; Giorgi, F.; Wang, Y.; Zhang, D. Reduction of future monsoon precipitation over China: Comparison between a high resolution RCM simulation and the driving GCM. *Meteorol. Atmos. Phys.* **2008**, *100*, 73–86.
68. Shen, Y.; Xiong, A.; Wang, Y.; Xie, P. Performance of high-resolution satellite precipitation products over China. *J. Geophys. Res.* **2010**, *115*, 1–17.
69. Badas, M.G.; Deidda, R.; Piga, E. Orographic influences in rainfall downscaling. *Adv. Geosci.* **2005**, *2*, 285–292.
70. Reid, I. The influence of slope aspect on precipitation receipt. *Weather* **1973**, *28*, 490–494.
71. Gires, A.; Tchiguirinskaia, I.; Schertzer, D.; Schellart, A.; Berne, A.; Lovejoy, S. Influence of small scale rainfall variability on standard comparison tools between radar and rain gauge data. *Atmos. Res.* **2014**, *138*, 125–138.

© 2015 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 license (<http://creativecommons.org/licenses/by/4.0/>).