




Article

Assimilation of Satellite-Based Data for Hydrological Mapping of Precipitation and Direct Runoff Coefficient for the Lake Urmia Basin in Iran

Mahdi Akbari ^{1,*}, Ali Torabi Haghighi ¹, Mohammad Mahdi Aghayi ²,
Mostafa Javadian ³, Masoud Tajrishy ² and Bjørn Kløve ¹

¹ Water, Energy and Environmental Engineering Research Unit, University of Oulu, PO Box 4300, FIN-90014 Oulu, Finland

² Department of Civil Engineering, Sharif University of Technology, Tehran 11155-9313, Iran

³ Department of Hydrology and Atmospheric Sciences, University of Arizona, PO Box 210011, Tucson, AZ 85721, USA

* Correspondence: mahdi.akbari@oulu.fi

Received: 25 June 2019; Accepted: 2 August 2019; Published: 6 August 2019



Abstract: Water management in arid basins often lacks sufficient hydro-climatological data because, e.g., rain gauges are typically absent at high elevations and inflow to ungauged areas around large closed lakes is difficult to estimate. We sought to improve precipitation and runoff estimation in an arid basin (Lake Urmia, Iran) using methods involving assimilation of satellite-based data. We estimated precipitation using interpolation of rain gauge data by kriging, downscaling the Tropical Rainfall Measuring Mission (TRMM), and cokriging interpolation of in-situ records with Remote Sensing (RS)-based data. Using RS-based data application in estimations gave more precise results, by compensating for lack of data at high elevations. Cokriging interpolation of rain gauges by TRMM and Digitized Elevation Model (DEM) gave 4–9 mm lower Root Mean Square Error (RMSE) in different years compared with kriging. Downscaling TRMM improved its accuracy by 14 mm. Using the most accurate precipitation result, we modeled annual direct runoff with Kennessey and Soil Conservation Service Curve Number (SCS-CN) models. These models use land use, permeability, and slope data. In runoff modeling, Kennessey gave higher accuracy. Calibrating Kennessey reduced the Normalized RMSE (NRMSE) from 1 in the standard model to 0.44. Direct runoff coefficient map by 1 km spatial resolution was generated by calibrated Kennessey. Validation by the closest gauges to the lake gave a NRMSE of 0.41 which approved the accuracy of modeling.

Keywords: hydrological modeling; downscaling TRMM; direct runoff coefficient; water scarcity; lakes; Urmia; desiccation; land use; ungauged basin; SCS-CN; Kennessey

1. Introduction

Mapping hydrological parameters such as runoff and precipitation requires a variety of approaches and the most widely used tools are data-driven [1]. However, in most catchments around the world runoff is not gauged [1] and, especially in developing countries, rain gauge networks have poor spatial and temporal resolution [2]. A good example is Lake Urmia in north-west Iran, where water overuse has led to lake desiccation [3] and where a marked lack of in-situ data for the basin is affecting lake restoration policy.

In Lake Urmia basin, there are two main limitations in runoff estimation: (i) the lake is surrounded by an ungauged buffer zone with a high concentration of agricultural activities, which has an important effect on water reaching to the lake body [4]; and (ii) there is a lack of sufficient data on water withdrawal by intensive irrigation in central sub-basins, as an overall survey is only performed every 10 years by

the Iran Water Resource Management Company (WRM) to determine the amount of water withdrawal. There are also problems in estimating precipitation, another important element of water resource management in the basin, owing to lack of inadequate rain gauges, especially at high elevations where a large proportion of precipitation in the basin occurs. A possible solution is use of satellite-based data in estimating precipitation [2,5,6].

Global precipitation data sets, including gauge-based, satellite-based, and reanalysis data sets, have proved useful across a wide range of fields of research [6]. The accuracy of these data sets has been investigated in different case studies. [2,7–11]. Moazami et al. [11] studied the performance of the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN), and the Tropical Rainfall Measuring Mission (TRMM) in Iran. They found that TRMM gives better performance based on statistical measures. More specifically, Ghajania et al. [2] compared six high resolution precipitation data sets (PERSIANN, CMORPH-RAW, CMORPH-CRT, TRMM 3B42, TRMM 3B43 V7, and APHRODITE) in Lake Urmia Basin. They found that APHRODITE and TRMM 3B43 V7 present better estimations in this basin.

Due to a scarcity of in situ observations in Lake Victoria, the largest lake in Africa, Inne Vanderkelen et al. [12] presented a water balance model for Lake Victoria, using state-of-the-art remote sensing observations, high-resolution reanalysis downscaling, and outflow values to estimate water balance of the lake. The computation of the individual water balance terms yields lake level fluctuations that closely match the levels retrieved from satellite altimetry. Georgy Ayzel and Alexander Izhitskiy [13], addressing discontinuous nature of hydrological regime of Aral Sea basin, have developed a continuous prediction system for assessing freshwater inflow into the Small Aral Sea based on coupling stack of hydrological and data-driven models. Their results show a good prediction of water assessment tool which utilizes classical physically based and modern machine learning models. Tilahun and Merkel [14] simulated groundwater recharge, surface runoff, and evapotranspiration in Dire Dawa (Ethiopia) by WetSpss. WetSpss as a physically-based methodology is a distributed water balance model.

To overcome the lack of runoff data in the ungauged buffer zone around Lake Urmia and precipitation records in high elevations, we present a novel data assimilation and modeling approach. First, precipitation is estimated by coupling TRMM, NDVI, and Digitized Elevation Model (DEM), as satellite-based data sets, with station data. Next, direct runoff is estimated using the Kennessey [15] and Soil Conservation Service Curve Number (SCS-CN) [16] models. The spatial variation in models' parameters is estimated from GIS layers covering slope, land use, and soil permeability. The approach maps precipitation and direct runoff coefficient (RC) as gridded outputs by 1 km spatial resolution.

2. Datasets and Study Area

2.1. Study Area

The study was carried out using the case of Lake Urmia basin (35°40'–38°30' N, 44°07'–47°53' E), one of six main basins in Iran. Mean annual precipitation in the basin is 340 mm and mean annual temperature is 12 °C. The area of the basin is about 52,000 km², consisting of around 34,000 km² of mountainous terrain and 13,000 km² of flat land supporting around 6000 km² of farms or orchards. The remaining 5000 km² of the basin are occupied by Lake Urmia (Figure 1a). This lake is mainly fed by the Zarinehroud, Siminehroud, Ajichay, Gadarchay, Brandooz, Shahrchay, Nazluchay, Mahabad, Rozechay, Ghalechay, and Zulachay rivers, but mean annual inflow has declined from 12 to 2.4 billion cubic meters (BCM) over the past five decades [17]. Due to unrecorded water withdrawal by farmers in the basin, for runoff modeling we used sub-areas of the basin that are less affected by human activities (hereafter called 'marginal sub-basins'), which were well distributed over the basin (Figure 1b).

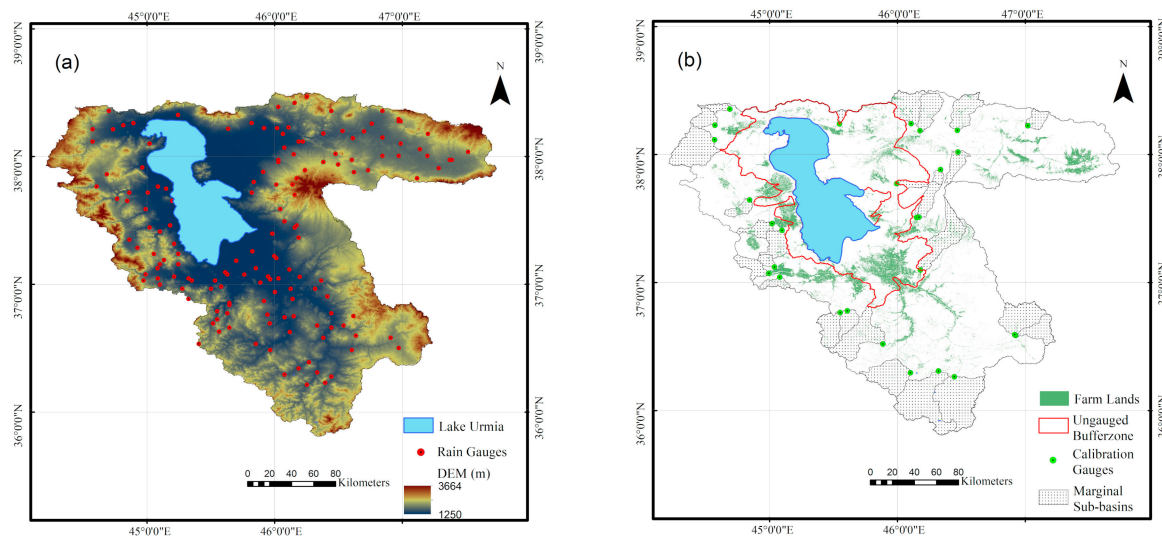


Figure 1. Map of the Lake Urmia basin showing the location of (a) rain gauges and (b) the ungauged buffer zone and marginal sub-basins designated in this study.

Lake Urmia is the second most hypersaline (about 300 P.S.U) lake in the world and the largest lake in Iran. However, the lake has lost more than 90% of its water volume in the past 25 years [18]. This is because farmers in the basin use as much water as possible during the irrigation season [19] which means consumption of 70% of the basin's renewable water resources [18,20]. This lake has been designated a protected area by UNESCO and is recognized as a Biosphere Reserve and Ramsar site in the Ramsar Convention [18]. One of the most important concerns raised by the desiccation of Lake Urmia is its transformation into an active center of salty dust and there have been warnings that Lake Urmia could become similar to the Aral Sea over time [21,22]. This is problematic since the population around Lake Urmia is much denser than that around the Aral Sea [23].

2.2. Data Sources

Monthly records of temperature were taken from 18 synoptic stations of Meteorological Organization (IMO) in Lake Urmia basin. Additionally, 157 rain gauges (Figure 1a) monthly values were obtained from WRM and IMO. Daily runoff values, measured in 27 gauges of marginal sub-basins, were obtained from WRM and used for calibration (Figure 1b). In addition, 10 gauges on main rivers of the basin (Figure 2d) were used for validation of the calibrated direct runoff model. Additionally, the most recent data on water withdrawal in Lake Urmia basin is available for 2008 which is published by WRM and utilized in this study. This data is updated each 10 years for all major basins in Iran.

TRMM Multi Satellite Precipitation Analysis [5], product 3B43 V7 which had monthly temporal resolution and 0.25° (about 25 km) of spatial resolution was used for estimation of precipitation. In TRMM 3B43 V7, Global Precipitation Climatology Project (GPCP) monthly rain gauge data are used for bias reduction [9]. In this study, monthly TRMM data from 2006 to 2011 were utilized. In addition, satellite-based Digitized Elevation Model (DEM) from Shuttle Radar Topography Mission (SRTM) initiated by NGA and NASA, which provide high quality DEM at global scale with 30 m spatial resolution [24], was used in precipitation estimation and generating the slope map of the basin (Figure 2c).

MOD13-Gridded Vegetation Indices by 1 km spatial resolution and 16-day temporal resolution [25] was used as required Normalized Difference Vegetation Index (NDVI) in the TRMM downscaling process. Additionally, the utilized land use map, produced in supervised classification with field survey data by Iran Water Research Institute (WRI) [26], has 30 m spatial resolution (Figure 2a). Soil texture data was taken from the Harmonized World Soil Database [27]. Additionally, a permeability map (Figure 2b) generated by the Remote Sensing Research Center of Sharif University of Technology [28]

by 1 km spatial resolution was used as required data for the Kennessey model. In this research, all gridded data were upscaled to 1 km to homogenize spatial resolution for all inputs of the models.

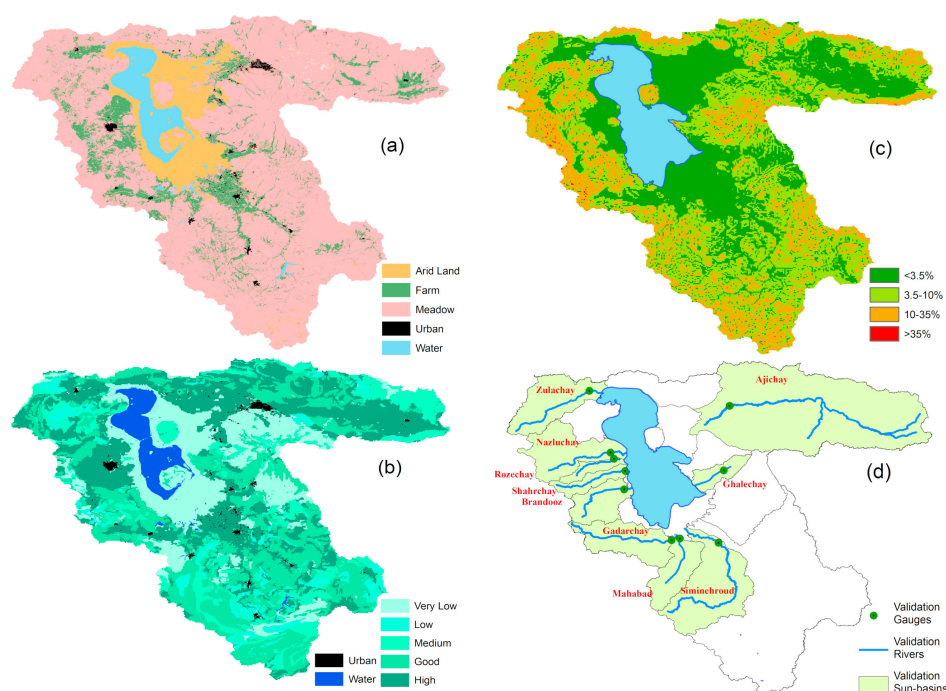


Figure 2. Lake Urmia basin data: (a) land use, (b) permeability, (c) slope, and (d) validation gauges.

3. Methodology

The work comprised two main steps (Figure 3):

1. Estimation of precipitation: observed precipitation data statistically examined in terms of data quality by Ghajarnia et al. [29] were utilized. These data were randomly selected and split, with 70% (110 rain gauges) used for precipitation estimation and 30% (47 rain gauges) for validation of the applied methods (using 50% for validation would have left too few rain gauges in the 52,000 km² basin for estimation). We did not involve 30% of rain gauges as validation dataset in the estimation process to prevent affectability of estimated precipitation from the validation dataset. Using 70% of rain gauge data and satellite images of TRMM, NDVI, and DEM for whole of the basin, annual precipitation was estimated by 1 km spatial resolution. The best method was then defined by Root Mean Square Error (RMSE), based on other pristine 30% of rain gauge data. The methods applied for estimating precipitation were geostatistical interpolation and downscaling. In a first estimation of the accuracy of each method, the year 2007 was selected. Then, to consider the effect of wet or dry years on the results of precipitation estimation, the most accurate method determined in 2007 was applied for the three water years (September–August) 2006–2007, 2007–2008, and 2010–2011, which were wet, dry, and normal years, respectively.

2. Direct runoff modeling: first, direct RC modeling was performed using the SCS-CN and Kennessey models in 2006. The results of these models were compared to marginal sub-basins gauge data (Figure 1b) because these sub-basins had a suitable situation in terms of data continuity and low impact of human activities on their observed gauges' data. Having specified the most accurate runoff model among Kennessey and SCS-CN in their standard format, we tried to calibrate it. Only marginal sub-basins were chosen for calibration because in these areas water withdrawal from surface resources, based on WRM data, is less than 5% of observed runoff in gauges. Furthermore, less than 5% of these sub-basins are covered by farm lands. Finally, the annual direct RC map was generated for the whole basin as gridded data by 1 km spatial resolution. We chose the closest flow gauges to the lake, which had continuous data during 2006–2011 and cover more than 45% of the basin. As shown

in Figure 2a, in validation sub-basins water withdrawal is high because of the wide area covered by farm lands. Since water withdrawal data is only available for 2008, the validation was done by using five-year means of annual estimated direct RC and observed values in the validation gauges (Figure 2d). Upstream withdrawal was deducted from the mean five-year hydrograph base flow of validation stations. Logically, if withdrawal is higher than base flow, the extra amount should be provided by direct runoff. Thus, the withdrawal value remaining after deduction from the base flow was added to the observed direct runoff, which resulted in naturalized direct runoff. The validation process was then completed by comparing naturalized observed RC and modeled values.

3.1. Estimation of Precipitation

3.1.1. Geostatistical Interpolation

Ordinary kriging [30] was used for interpolation. Kriging weights the surrounding measured values to derive an estimation for an unmeasured location. The general formula for interpolators is formed as a weighted sum of the data:

$$Z(x_0) = \sum_{i=1}^N \lambda_i Z(x_i), \quad (1)$$

where $Z(x_i)$ is the measured value at the i th location; λ_i is an unknown weight for the measured value at the i th location; x_0 is the estimation location, and N is the number of measured values. In the kriging method, λ_i depends on the distance between the measured points and the estimation location, and also on the spatial autocorrelation of the measured points.

Additionally, cokriging interpolation was utilized in this study to estimate the precipitation because this method can explore more than two variables and data from auxiliary variables to make estimations. Cokriging is commonly applied in environmental studies [31]:

$$S(d) = \mu(d) + \delta(d), \quad (2)$$

where $\mu(d)$ is an average vector with fixed effects and $\delta(d)$ is a random vector with a zero mean. $S(d)$ is set as data $\{S(d_1), \dots, S(d_n)\}$ detected at locations $\{d_1, \dots, d_n\}$. For m different variables:

$$\begin{pmatrix} S_1 \\ \dots \\ S_m \end{pmatrix} = \begin{pmatrix} (EV)_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & (EV)_m \end{pmatrix} \begin{pmatrix} \theta_1 \\ \dots \\ \theta_m \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \dots \\ \delta_m \end{pmatrix}, \quad (3)$$

where $(EV)_i$ is a matrix in which k th row is $(ev)_i(d)_k$; $k = 1, \dots, n$; $i = 1, \dots, m$. $[(ev)(d)]$ represents explanatory variables for the i th variable to be predicted and θ_i is a vector of unknown regression parameters [31].

3.1.2. Downscaling

To enhance the spatial resolution of the satellite precipitation product, we downscaled TRMM 3B43 V7 using an approach developed by Agam [32] and a method proposed by Immerzeel [33] that involves finding an exponential relationship between TRMM as a low resolution (LR) data (0.25° which is about 25 km) and NDVI as a high resolution (HR) data (1 km). We also tested use of DEM for downscaling TRMM instead of NDVI.

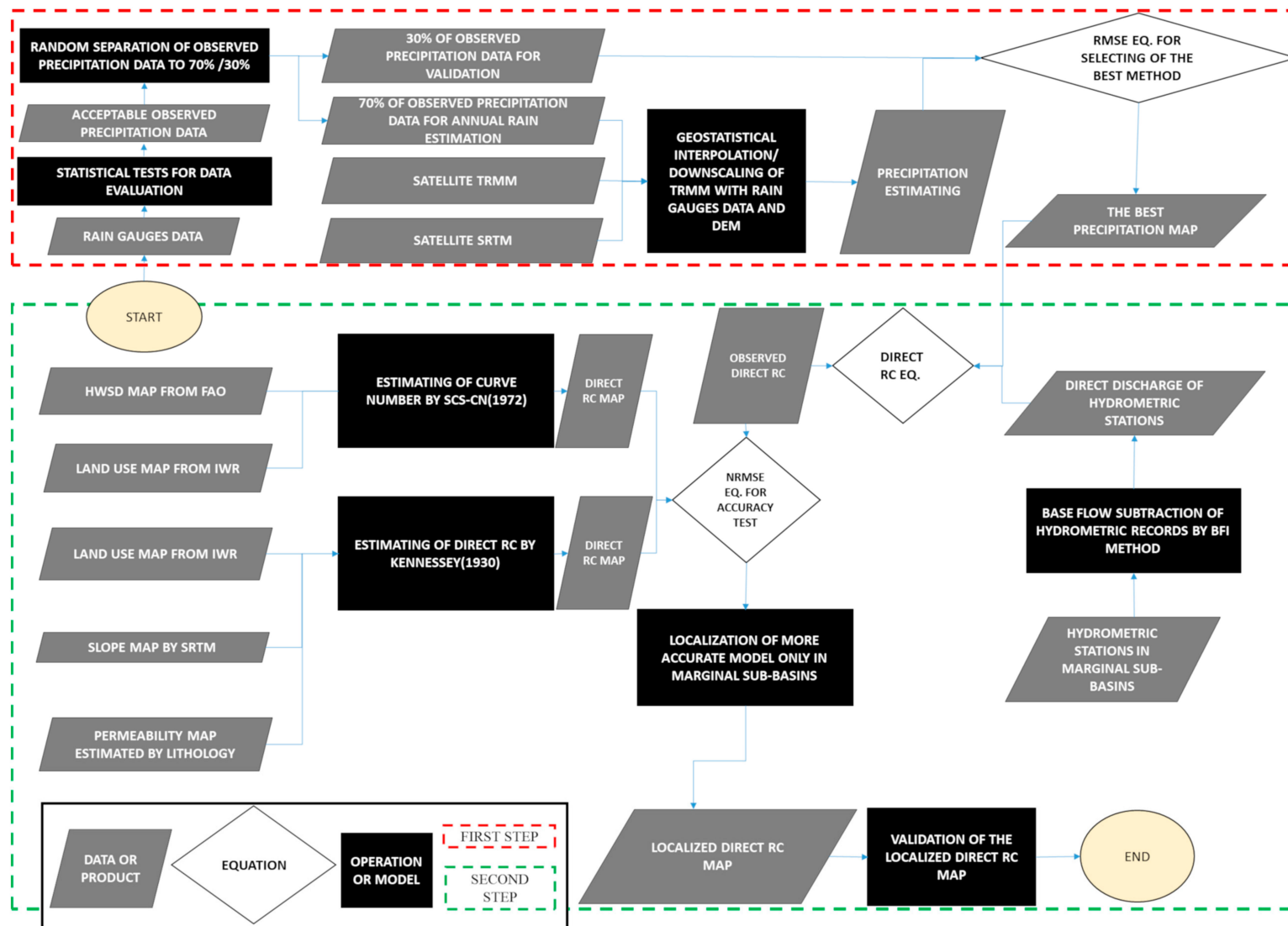


Figure 3. Flowchart showing steps followed in estimation and mapping of annual precipitation and direct runoff coefficient for Lake Urmia basin.

For downscaling: (1) the spatial resolution of DEM and NDVI were decreased to 0.25° to produce DEM_{LR} and $NDVI_{LR}$; (2) two exponential relationships achieved between DEM_{LR} -TRMM and $NDVI_{LR}$ -TRMM which had 0.25° spatial resolution; (3) we applied obtained exponential relationships on DEM_{LR} and $NDVI_{LR}$; (4) the product of step 3 was deduced from TRMM. The obtained value was called Res_{LR} ; (5) through spline interpolation, spatial resolution of Res_{LR} was enhanced from 0.25° to 1 km to produce Res_{HR} ; (6) eventually, the exponential relationships obtained from step 2 was applied on original high resolution DEM and NDVI; (7) by summing up Res_{HR} and the product of step 6, downscaled TRMM was finally produced.

3.2. Direct Runoff Modeling

3.2.1. Baseflow Subtraction

To model direct runoff, first the base flow should be subtracted from observed runoff. In this study, baseflow was subtracted by Base Flow Index (BFI) method with the USGS-GW-Toolbox [34]. The BFI program [35] is based on a set of procedures developed by the Institute of Hydrology in which the streamflow record is partitioned into intervals of length N-days. The minimum streamflow during each N-day interval then is identified and compared to adjacent minimums to determine “turning points”. If 90% of a given minimum (the “turning point test factor”) is less than both adjacent minimums, then that minimum is a turning point. The base-flow hydrograph is completed by connecting the turning points [34].

3.2.2. SCS-CN

For estimation of direct runoff, we used the SCS-CN method, which has been applied and modified by several authors [36–41]. The SCS-CN consists of the water balance equation and two fundamental hypotheses expressed, respectively, as:

$$P = Ia + F + Q, \quad (4)$$

$$\frac{Q}{P - Ia} = \frac{F}{S'} \quad (5)$$

$$Ia = \lambda S, \quad (6)$$

where P is total precipitation (mm); Ia is initial abstraction (mm); F is cumulative infiltration (mm); Q is direct runoff (mm); S is potential maximum retention (mm); and λ is initial abstraction coefficient. Potential maximum retention can be transformed to CN scale using the following empirical relationship:

$$CN = 25400 / (S + 254), \quad (7)$$

where S is in mm and CN is a non-dimensional parameter. Furthermore, CN of a region can be determined by Curve Number tables [16]. A sample of these tables is presented in Table 1. To model SCS-CN in the present study, we derived soil texture from the Harmonized World Soil Database to determine soil group map. Soil group can be A, B, C, and D which represent the runoff potential of the soil according to its soil texture (Table 1). By CN , direct runoff can be calculated using above equations.

Table 1. Curve number for different land uses in the Soil Conservation Service Curve Number (SCS-CN) model.

Land Use	Curve Number			
	A (Low Runoff Potential)	B (Moderate Infiltration)	C (Slow Infiltration)	D (High Runoff Potential)
Arid land	77	86	91	94
Meadow	30	58	71	78
Forest	35	56	70	77
Crop	74	83	88	90

3.2.3. Kennessey

We used the Kennessey model in our study, which has also been applied in other studies [42–45]. According to this model, three major factors influence runoff: slope, land use, and soil permeability. Later, Tardi and Vittorini [46] added a climatic parameter, Ia , which represents the extent of wetness of the water year in this model:

$$Ia = \frac{\frac{P}{(T+10)} + \frac{12p}{(t+10)}}{2}, \quad (8)$$

where P is mean annual precipitation (mm); T is mean annual temperature ($^{\circ}\text{C}$); p is precipitation in the driest month of the year (mm); and t is temperature of the driest month of the year ($^{\circ}\text{C}$). The partial coefficients of the model are categorized in three different classes according to Ia .

Direct RC is quantified by partial coefficients of the Kennessey (Table 2). By this model and gridded inputs (shown in Figure 2a–c), each grid, according to Ia of the year, gets three specific partial coefficients representing its slope, permeability, and land use. Finally, annual direct RC of each grid is calculated by summing up these three partial coefficients.

Table 2. Partial runoff coefficient values for the four different parameter classes used in the Kennessey.

Classes	Aridity Index (Ia)		
	$Ia < 25$	$25 < Ia < 40$	$Ia > 40$
Slope			
>35%	0.22	0.26	0.30
10%–35%	0.12	0.16	0.20
3.5%–10%	0.01	0.03	0.05
<3.5%	0.00	0.01	0.03
Land cover			
Arid land	0.26	0.28	0.30
Meadow	0.17	0.21	0.25
Farm	0.07	0.11	0.15
Forests	0.03	0.04	0.05
Permeability			
Very low	0.21	0.26	0.30
Low	0.16	0.21	0.25
Medium	0.12	0.16	0.20
Good	0.06	0.08	0.10
High	0.03	0.04	0.05

3.2.4. Calibration of Kennessey

Having specified the most accurate runoff model by RMSE in 2006, we calibrated it only in marginal sub-basins for the studied time span from 2006–2007 to 2010–2011. To assess the accuracy of calibrated model, the weighted average (Equation (9)) of modeled direct RC in each marginal sub-basin were compared against the observed values in gauges shown in Figure 1b, from which base flow was first subtracted by the BFI method. To calibrate the Kennessey model, we estimated direct RC in each marginal sub-basin:

$$\widehat{RC}_{ik} = \sum_{j=1}^n \alpha_{ik,j} C_j, \quad (9)$$

where \widehat{RC}_{ik} is modeled direct RC of the i th marginal sub-basins in year k ; C_j is partial coefficients of the model (Table 3); $\alpha_{ik,j}$ is the percent of marginal sub-basin i covered by C_j in year k . $\alpha_{ik,j}$ is calculated according to gridded inputs of the Kennessey model (Figure 2a–c) classified in Table 3 by Ia ; n is equal to 39, the number of partial coefficients in Kennessey. For example, if Ia of the i th sub-basin in year k is

in class I (Table 3), $\alpha_{ik,j}$ s for C_j s when $j \geq 14$ will get zero values. Furthermore, if C_j does not exist in the i th sub-basin, corresponding $\alpha_{ik,j}$ will be zero as well.

Table 3. Parameters of optimization for calibration of the Kennessey model.

Classes	Aridity Index (<i>Ia</i>)		
	Class I of <i>Ia</i>	Class II of <i>Ia</i>	Class III of <i>Ia</i>
Slope			
>35%	C_1	C_{14}	C_{27}
10%–35%	C_2	C_{15}	C_{28}
3.5%–10%	C_3	C_{16}	C_{29}
<3.5%	C_4	C_{17}	C_{30}
Land cover			
Arid land	C_5	C_{18}	C_{31}
Meadow	C_6	C_{19}	C_{32}
Farm	C_7	C_{20}	C_{33}
Forests	C_8	C_{21}	C_{34}
Permeability			
Very low	C_9	C_{22}	C_{35}
Low	C_{10}	C_{23}	C_{36}
Medium	C_{11}	C_{24}	C_{37}
Good	C_{12}	C_{25}	C_{38}
High	C_{13}	C_{26}	C_{39}

The objective and constraints of calibration are:

$$\begin{aligned} \text{Objective : } \min f(C_j) &= \sum_{k=1}^M \sum_{i=1}^N (\widehat{RC}_{ik} - RC_{ik})^2, \\ \text{s.t. } C_j &\leq C_{j+13}; 1 \leq j \leq 26, \end{aligned} \quad (10)$$

where RC_{ik} is the observed direct RC of the i th marginal sub-basins in year k ; N is equal to 27, the number of marginal sub-basins. M is equal to 5 as the number of investigated water years from 2006–2007 to 2010–2011. The parameters of optimization (Table 3) are 39 partial coefficients of the Kennessey model. The constraint of the optimization is based on *Ia* role in the Kennessey model. This is because when *Ia* increases, it means that wetness of the year has increased. Each class of *Ia* can be separated by $j + 13$ while j is less than 26.

For calibration, we tried to minimize the square difference between estimated and observed values of direct RC using the *fmincon* function of MATLAB optimization toolbox [47]. *fmincon* uses a sequential quadratic programming (SQP) method. In this method, the function solves a quadratic programming (QP) subproblem at each iteration. *fmincon* updates an estimate of the Hessian of the Lagrangian at each iteration using the BFGS formula [48].

3.3. Statistical Metrics

The error of the different methods used for estimation of precipitation was determined as RMSE. Additionally, the error of direct runoff modeling is reported by Normalized RMSE (NRMSE). Lower values of these statistical metrics indicated higher accuracy of the method investigated.

$$\text{RMSE} = \sqrt{\sum_{i=1}^N \frac{(Z_i^{\text{obs}} - Z_i^{\text{mod}})^2}{N}}, \quad (11)$$

$$\text{NRMSE} = \frac{\text{RMSE}}{(\sum_{i=1}^N Z_i^{\text{obs}} / N)}, \quad (12)$$

where Z_i^{obs} is the value of observations in the i th record and Z_i^{mod} is the value of model; N is the number of observations. Actually, NRMSE is normalized by dividing RMSE by the mean of observed values. If NRMSE is less than 1, the error is less than 100%. NRMSE can be reported in percentage.

In this study, the precipitation and direct RC were estimated using the procedure within GIS (ESRI ArcGIS 10.3(Environmental Systems Research Institute, Inc, Redlands, CA, USA)), where each information layer was converted to raster format and then processed by the methodology described above.

4. Results and Discussion

4.1. Spatiotemporal Precipitation Maps

Cokriging interpolation of rain gauge values with DEM and TRMM 3B43 V7 gave the best method accuracy, with RMSE of 81 mm for 2007 (Table 4). This was 4 mm less than the RMSE of kriging interpolation of station data. The best method was also confirmed for 2010–2011, a normal water year in terms of rainfall, for which RMSE was reduced by 9 mm (Table 5). The benefit of using satellite data was that RMSE was reduced by 4–9 mm across different years compared with kriging interpolation. Further, downscaled TRMM 3B43 V7 by finding an exponential relationship with DEM reduced its RMSE by 14 mm. Additionally, there was no significant relationship between NDVI and TRMM in the studied area, which can attribute to the high concentration of agricultural activities as an anthropogenic effect on NDVI.

Table 4. Root Mean Square Error (RMSE) of different methods for precipitation estimation in Lake Urmia basin in 2007.

Method	Mean (mm)	RMSE (mm)
TRMM raw images	361	100
Kriging interpolation of 70% of station records	297	85
Downscaling TRMM with DEM	372	86
Cokriging interpolation of station records with DEM and TRMM	290	81
Cokriging interpolation of station data with Aspect and TRMM	290	86

Table 5. RMSE of the best method for estimation of precipitation in dry, wet, and normal years.

Normal year 2010	Estimation	Method		Wet year 2006	Estimation	Method		Dry year 2007	Estimation	Method	
		Cokrig	Krig			Cokrig	Krig.			Cokrig	Krig.
	Mean	315	337		Mean	390	392		Mean	170	173
	RMSE	97	106		RMSE	147	152		RMSE	81	85

The precipitation layers for 2006–2011 obtained by cokriging interpolation of rain gauge values with DEM and TRMM 3B43 V7, as the best method, revealed high values of precipitation in marginal areas of the basin (Figure 4). As can be seen from Figure 1a, these marginal areas have high elevation and use of DEM in the estimation process resulted in higher values of precipitation for these mountainous regions.

4.2. Annual Direct Runoff Coefficient Map

The NRMSEs obtained for the SCS-CN and Kennessey models in 2006 were 3 and 1, respectively. Therefore, Kennessey was chosen as the more accurate model for calibration. In this model, Ia is an important parameter because, as described in Section 3.2.3, this index as a climatic parameter is defined in Kennessey to reflect the dryness or wetness of each water year. However, applying Equation (8) from 2006 to 2011 in Lake Urmia basin gave a maximum Ia value equal to 18. If Ia is lower than 25, the

partial coefficients lie in the same category (see Table 2). Therefore, in the studied basin, all years from 2006 to 2011 were categorized in the same class in terms of runoff production by the default intervals of Ia regardless of its precipitation or temperature variability. In reality, wet, dry and normal years have occurred (see Figure 4) from 2006–2007 to 2010–2011. As novelty of this study, attempts were made to identify new intervals for Ia to make Kennessey flexible to wetness or dryness of investigated years in Lake Urmia basin as a semi-arid region.

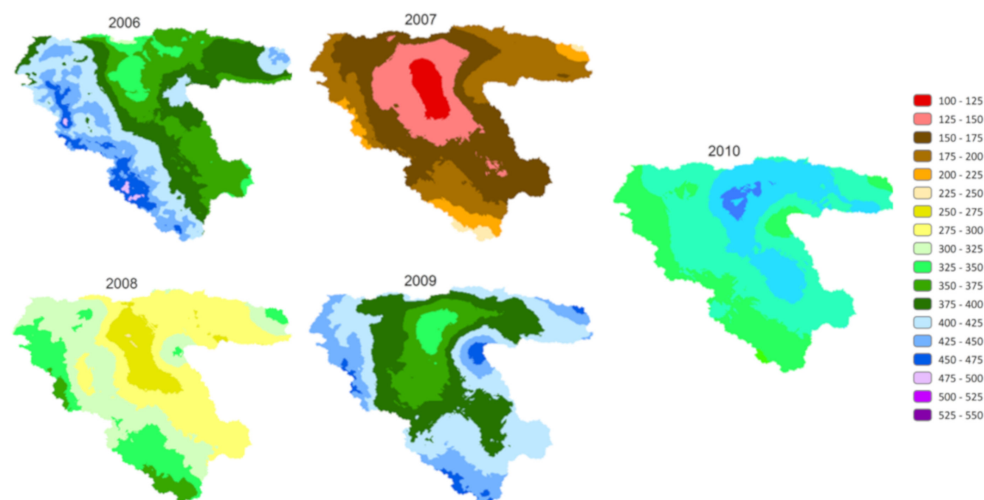


Figure 4. Precipitation (mm) in Lake Urmia basin in five investigated water years (September–August) from 2006–2007 to 2010–2011, based on cokriging interpolation of in-situ records, the Tropical Rainfall Measuring Mission (TRMM), and Digitized Elevation Model (DEM).

The values obtained indicated that in the dry year 2007–2008, Ia ranged between 3 and 6, and thus $Ia < 6$ can represent dryness. In normal years (2008–2009 and 2010–2011), Ia was on average 7–9 so $6 < Ia < 9$ was presumed to represent the normal year. In the wet years (2006–2007 and 2009–2010), Ia was more than 9, so $Ia > 9$ suggested a wet year in the studied basin.

The partial coefficients of the calibrated Kennessey model are shown in Table 6. Using the results of the calibration, a scatter plot of modeled coefficients and observed RC was created after convergence of estimated and observed values (Figure 5). There was observable agreement between modeled and observed values, such that the correlation coefficient obtained 80%. The calibrated Kennessey reduced the NRMSE from 1 to 0.44. The map of the five year mean of annual direct RC for Lake Urmia basin from 2006–2007 to 2010–2011 produced by calibrated Kennessey by 1 km spatial resolution is shown in Figure 6.

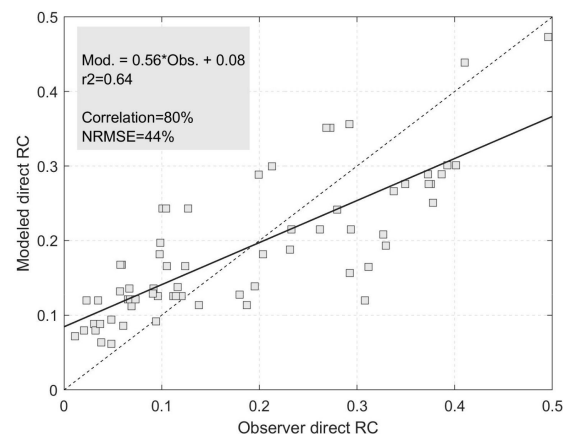
As can be seen from Figure 5, when the observed direct RC was lower than 0.2, most of the values less than 0.2 were above the 1:1 line in the first quadrant, i.e., the model overestimated direct RC in this range. At values greater than 0.2, the model underestimated direct RC.

The direct RC map generated for Lake Urmia basin (Figure 6) was validated using observed values in gauges on the main rivers of the basin. Since the water withdrawal is high in validation sub-basins by intensive farm lands (Figure 2d), we must consider it in observation of validation gauges. The water withdrawal of the basin, published by WRM, is only available for year 2008 so we validated direct RC map of calibrated Kennessey by averaging estimated values from 2006–2007 to 2010–2011 and then compared them to the five-year mean annual observed direct runoff in validation gauges which were naturalized by considering water withdrawal in observed value of gauges (see step 2 described in Section 3 ‘Methodology’).

Table 6. Calibrated partial runoff coefficient of the Kennessey model.

Classes	Aridity Index (<i>Ia</i>)		
	<i>Ia</i> < 6	6 < <i>Ia</i> < 9	<i>Ia</i> > 9
Slope			
>35%	$C_1 = 0.221$	$C_{14} = 0.249$	$C_{27} = 0.282$
10%–35%	$C_2 = 0.000$	$C_{15} = 0.001$	$C_{28} = 0.001$
3.5%–10%	$C_3 = 0.000$	$C_{16} = 0.001$	$C_{29} = 0.001$
<3.5%	$C_4 = 0.000$	$C_{17} = 0.001$	$C_{30} = 0.001$
Land cover			
Arid land	$C_5 = 0.000$	$C_{18} = 0.001$	$C_{31} = 0.015$
Meadow	$C_6 = 0.000$	$C_{19} = 0.000$	$C_{32} = 0.000$
Farm	$C_7 = 0.002$	$C_{20} = 0.205$	$C_{33} = 0.206$
Forests	$C_8 = \text{NA}^*$	$C_{21} = \text{NA}^*$	$C_{34} = \text{NA}^*$
Permeability			
Very low	$C_9 = 0.379$	$C_{22} = 0.475$	$C_{35} = 0.476$
low	$C_{10} = 0.030$	$C_{23} = 0.030$	$C_{36} = 0.060$
Medium	$C_{11} = 0.008$	$C_{24} = 0.023$	$C_{37} = 0.060$
Good	$C_{12} = 0.008$	$C_{25} = 0.021$	$C_{38} = 0.053$
High	$C_{13} = 0.000$	$C_{26} = 0.000$	$C_{39} = 0.050$

* NA means that the class does not exist in Lake Urmia basin.

**Figure 5.** Scatter plot of modeled and observed direct runoff coefficient (RC) values in marginal sub-basins.

By pristine validation gauges data which were not involved in the calibration process, the NRMSE of validation obtained 0.41, which is close to the calibration error, indicating success of the calibration process. In Table 7, the main rivers in Lake Urmia basin are listed in descending order of observed runoff. The last column on the right shows the difference between the observed and modeled values divided by the observed value. As can be seen, for most of the rivers providing higher discharge to the lake, the error was less than 50%. For the major rivers, Gadarchay, Mahabad, Rozechay, and Simineroud, the accuracy is higher. The highest error was found for the Zulachay river basin, which was due to lower value of observed direct runoff (smaller fraction denominator). The mean of validation gauges for modeled values obtained 0.19 and that of the observation was 0.16 and correlation between observed and modeled values obtained 79%.

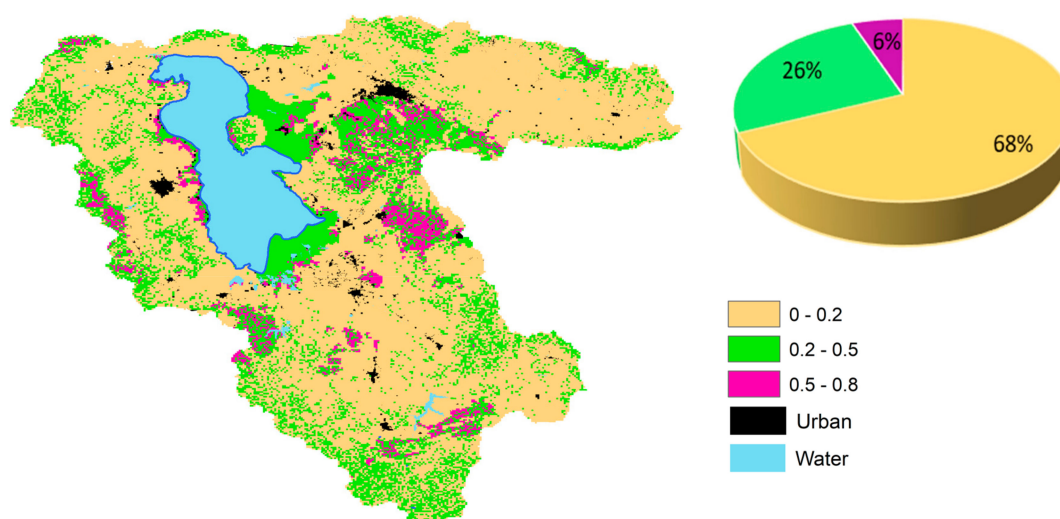


Figure 6. Mean of annual direct RC map in Lake Urmia basin from 2006–2007 to 2010–2011 produced by calibrated Kennessey.

Table 7. Gauge river discharge values used in validation of the calibrated Kennessey model.

River	Basin Area (km ²)	Withdrawal (mm)	R_{obs} (mm)	P (mm)	RC_{obs}	RC_{mod}	Diff. (%)
Gadarchay	2074	152	115	331	0.35	0.34	3%
Mahabad	1370	87	73	328	0.22	0.22	1%
Rozechay	313	82	68	338	0.2	0.17	13%
Siminehroud	3186	40	54	335	0.16	0.15	9%
Brandooz	1143	94	43	334	0.13	0.16	−24%
Ghalechay	481	35	43	305	0.14	0.2	−45%
Shahrchay	579	27	42	339	0.12	0.25	−108%
Ajichay	10,052	37	35	310	0.11	0.16	−43%
Nazluchay	1966	40	28	340	0.08	0.15	−91%
Zulachay	1892	25	15	330	0.05	0.15	−196%

R_{obs} : naturalized observed direct runoff, P : precipitation, RC_{obs} : naturalized observed direct runoff coefficient, RC_{mod} : modeled direct runoff coefficient, $Diff. = 100(RC_{obs} - RC_{mod}) / RC_{obs}$.

4.3. Limitations in Data, Modeling, and Assessment

The present study was based on data from 2006 to 2011. These years capture well the interannual variation in weather, including wet (e.g., 2006–2007, 2009–2010), dry (e.g., 2007–2008), and normal (e.g., 2008–2009, 2010–2011) water years. Additionally, the most recent data on water withdrawal and the land use map of the basin are for year 2008 which is in the middle of the studied time span. As described above, water withdrawal data were essential for validation of the calibrated direct runoff model. Under strategies approved by central government, the hydrological status of Lake Urmia basin has stabilized in recent years. These strategies prohibit any kind of additional withdrawal from the basin water resources or any new development, especially in the agricultural sector, prevent unpermitted withdrawal from surface waters, and stop all dam construction projects under study and under operation [18]. Due to these measures, it can be concluded that 2006–2011 was representative of the hydrological status in Lake Urmia basin. By calibrating the model for this period, it was possible to apply the results to the whole basin in recent years.

Expanding the time span of the assessment after solving the problem of data availability would provide more observations for calibration of the direct runoff model, increasing the accuracy of the results. Considering new intervals for Ia or generating another index to replace Ia could help the Kennessey model distinguish between wet, dry, and normal years in semi-arid basins like Lake Urmia. Another limitation in the present study was that the Zarinehroud, one of the most important rivers

in Lake Urmia basin, was not covered in the validation process because of its gauge data problem. Additionally, enhancing the quality of withdrawal data in terms of spatial and temporal resolution can help measure the validity of the calibrated model. Finally, it should be noted that water withdrawal data, available only for 2008, did not permit to validate the direct RC map annually and we had to compare the model and observed direct RC values by averaging them for five years from 2006–2007 to 2010–2011.

5. Conclusions

Assimilation of satellite-based data in in-situ observations successfully improved the accuracy of hydrological modeling for Lake Urmia basin by covering data gaps. The most important difference between the kriging and cokriging methods was use of RS-based datasets, including TRMM 3B43 V7 and DEM, in cokriging. By covering data gaps at high altitudes, cokriging allowed the high amount of precipitation observed in rain gauges located in mountainous areas to be modeled. Downscaling TRMM 3B43 V7 with the DEM of Lake Urmia basin proved as accurate as kriging interpolation of rain gauge observations and provided a real-time estimate of precipitation in the basin because rain gauge data for the basin are available after a lag of some years.

For annual direct RC modeling, Kennessey performed better than SCS-CN only in the case of the run with default parameters. Since the error was high (more than 100%), the Kennessey model required calibration. As a novel approach in modeling by Kennessey in a semi-arid basin, changing *Ia* intervals for classification of partial coefficients of the model resulted in a more accurately calibrated Kennessey. Calibration reduced the NRMSE of the model from 1 in standard format to 0.44 (correlation 80%). The results of validation (correlation 79% and NRMSE 0.41) approved the model accuracy. The calibration sub-basins represented all default classes of the Kennessey model, so the calibrated form was generalizable to the entire basin to produce a gridded direct RC map by 1 km spatial resolution and 1-year temporal resolution. Direct runoff in the ungauged buffer zone was thereby modeled without data on water use or gauge runoff data. The outputs and novel approach of this study can help water balance or aquifer recharge mapping at a grid scale.

Author Contributions: Conceptualization, M.A., A.T.H., M.M.A., M.J., M.T. and B.K.; Methodology, M.A., A.T.H., M.M.A., B.K., M.T. and M.J.; Software, M.A., A.T.H. and M.M.A.; Validation, M.A., A.T.H., M.M.A. and M.T.; Formal Analysis, M.A., A.T.H., M.M.A., M.J. and M.T.; Investigation, M.A., A.T.H., M.M.A. and M.T.; Resources, M.A., A.T.H., M.M.A. and M.T.; Data Curation, M.A., A.T.H., M.M.A., M.J. and M.T.; Writing-Original Draft Preparation, M.J. and M.A.; Writing-Review & Editing, M.A., A.T.H., M.T. and B.K.; Visualization, M.A. and M.J.; Supervision, M.T., A.T.H., M.M.A. and B.K.

Funding: This research was funded by Urmia Lake Restoration Committee and University of Oulu.

Acknowledgments: The authors are thankful to A. Daneshkar and N. Ghajarnia, who provided insights and expertise that greatly assisted the research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Blöschl, G.; Sivapalan, M.; Wagener, T.; Viglione, A.; Savenije, H. *Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales*, 2nd ed.; Cambridge University Press: New York, NY, USA, 2011; pp. 5–6.
2. Ghajarnia, N.; Liaghat, A.; Daneshkar, A.P. Comparison and evaluation of high resolution precipitation estimation products in Urmia Basin-Iran. *Atmos. Res.* **2015**, *158*, 50–65. [[CrossRef](#)]
3. Fazel, N.; Haghighi, A.T.; Klöve, B. Analysis of Land Use and Climate Change Impacts by Comparing River Flow Records for Headwaters and Lowland Reaches. *Glob. Plant. Chang.* **2017**, *4*, 47–56. [[CrossRef](#)]
4. Youneszadeh, S.; Sima, S.; Javadian, M.; Tajrishy, M. Determining Daily Variations of River Flood Plains in the Southern Buffer Zone of Urmia Lake in 2010. *Iran Water Resour. Res.* **2018**, *14*, 25–31.

5. Huffman, G.J.; Adler, R.; Bolvin, D.; Gu, G.; Nelkin, E.; Bowman, K.; Stocker, E.; Wolff, D. The TRMM multi-satellite precipitation analysis: Quasiglobal, multi-year, combined-sensor precipitation estimates at fine scale. *J. Hydrometeorol.* **2007**, *8*, 38–55. [\[CrossRef\]](#)
6. Sun, Q.; Miao, C.; Duan, Q.; Ashouri, H.; Sorooshian, S.; Hsu, K.L. A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. *Rev. Geophys.* **2018**, *56*, 79–107. [\[CrossRef\]](#)
7. Ghajarnia, N.; Daneshkar, A.P.; Liaghat, M.; Araghinejad, S. Error Analysis on PERSIANN Precipitation Estimations: Case Study of Urmia Lake Basin, Iran. *J. Hydrol. Eng.* **2018**, *23*, 05018006. [\[CrossRef\]](#)
8. Golian, S.; Javadian, M.; Behrangi, A. On the use of satellite, gauge, and reanalysis precipitation products for drought studies. *Environ. Res. Lett.* **2019**, *14*, 075005. [\[CrossRef\]](#)
9. Habib, E.; Henschke, A.; Adler, R. Evaluation of TMPA satellite-based research and real-time precipitation estimates during six tropical related heavy precipitation events over Louisiana, USA. *Atmos. Res.* **2009**, *94*, 373–388. [\[CrossRef\]](#)
10. Jamandre, C.A.; Narisma, G.T. Spatio-temporal validation of satellite-based rainfall estimates in the Philippines. *Atmos. Res.* **2013**, *122*, 599–608. [\[CrossRef\]](#)
11. Moazami, S.; Golian, S.; Kavianpour, M.R.; Yang, H. Comparison of PERSIANN and V7 TRMM multi-satellite precipitation analysis (TMPA) products with rain gauge data over Iran. *Int. J. Remote Sens.* **2013**, *34*, 8156–8171. [\[CrossRef\]](#)
12. Vanderkelen, I.; van Lipzig, N.P.M.; Thiery, W. Modelling the water balance of Lake Victoria (East Africa)—Part 1: Observational analysis. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 5509–5525. [\[CrossRef\]](#)
13. Ayzel, G.; Izhitskiy, A. Coupling physically based and data-driven models for assessing freshwater inflow into the Small Aral Sea. *Proc. IAHS* **2018**, *379*, 151–158. [\[CrossRef\]](#)
14. Tilahun, K.; Merkel, B.J. Estimation of groundwater recharge using a GISbased distributed water balance model in Dire Dawa, Ethiopia. *Hydrogeol. J.* **2009**, *17*, 1443–1457. [\[CrossRef\]](#)
15. Kennessey, B. *Lefoljasy Tènyezok ès Renenciok*; Vizugy: Koziemènyek, Hungary, 1930.
16. Soil Conservation Service. *National Engineering Handbook*; Soil Conservation Service: Washington, DC, USA, 1972.
17. Delju, A.; Ceylan, A.; Piguët, E.; Rebetez, M. Observed climate variability and change in Lake Urmia Basin, Iran. *Theor. Appl. Climatol.* **2013**, *111*, 285–296. [\[CrossRef\]](#)
18. Urmia Lake Restoration Program (ULRP). Projects Outline 2015. Available online: <http://ulrp.sharif.ir/en> (accessed on 15 May 2015).
19. Haghighi, A.T.; Fazel, N.; Hekmatzadeh, A.A.; Klöve, B. Analysis of Effective Environmental Flow Release Strategies for Lake Urmia Restoration. *Water Resour. Manag.* **2018**, *32*, 3595–3609. [\[CrossRef\]](#)
20. Javadian, M.; Kordi, F.; Tajrishy, M. Evaluation and Comparison of Estimation Methods for Actual Evapotranspiration in the Urmia Lake Basin. *Iran. J. Ecohydrol.* **2019**, *6*, 125–136. [\[CrossRef\]](#)
21. AghaKouchak, A.; Norouzi, H.; Madani, K.; Mirchi, A.; Azarderakhsh, M.; Nazemi, A.; Hasanzadeh, E. Aral Sea syndrome desiccates Lake Urmia: Call for action. *J. Great Lakes Res.* **2015**, *41*, 307–311. [\[CrossRef\]](#)
22. Torabian, E.; Gemenne, F.; Brücker, P.; Ionesco, D. *Exploring Social Vulnerability and Environmental Migration in Lake Urmia in Iran: Comparative Insights from the Aral Sea. The State of Environmental Migration*; International Organization for Migration (IOM) and SciencePo: Grand Sakonay, Switzerland, 2015; pp. 65–87.
23. United Nations Environment Programme (UNEP). *The Drying of Iran's Lake Urmia and Its Environmental Consequences*; UNEP: Nairobi, Republic of Kenya, 2012.
24. Nikolakopoulos, K.G.; Kamaratakis, E.K.; Chrysoulakis, N. SRTM vs ASTER elevation products. Comparison for two regions in Crete, Greece. *Int. J. Remote Sens.* **2006**, *27*, 4819–4838. [\[CrossRef\]](#)
25. MOD13-Gridded Vegetation Indices from 2007. Maintained by the NASA EOSDIS LP DAAC at the USGS EROS, Sioux Falls, SD, USA. 2018. Available online: <https://doi.org/10.5067/MODIS/MOD13A2.006> (accessed on 16 August 2017).
26. Water Research Institute of Ministry of Energy. *Investigation of Changing Environmental Conditions Lake Urmia Basin (Land Use Report)*; Water Research Institute of Ministry of Energy: Navad Urmia, Iran, 2012.
27. FAO/IIASA/ISRIC/ISSCAS/JRC. *Harmonized World Soil Database (Version 1.2)*; FAO: Rome, Italy; IIASA: Laxenburg, Austria, 2012.
28. Remote Sensing Research Center (RSRC). *Water Balance of Lake Urmia Basin in 1995 and 2010*; Sharif University of Technology: Tehran, Iran, 2014.
29. Ghajarnia, N.; Daneshkar, A.P. Verifying precipitation data of TAMAB and meteorology institute in Urmia basin. *Water Soil Conserv. Rep.* **2013**, *4*, 91–109.

30. Wackernagel, H. *Multivariate Geostatistics: An Introduction with Applications*; Springer: Berlin, Germany, 2003; ISBN 978-3-540-44142-7.
31. Lingqing, W.; Dai, L.; Li, L.; Liang, T. Multivariable cokriging prediction and source analysis of potentially toxic elements (Cr, Cu, Cd, Pb, and Zn) in surface sediments from Dongting Lake, China. *Ecol. Indic.* **2018**, *94*, 312–319.
32. Agam, N.; Kustas, W.P.; Anderson, M.C.; Li, F.; Neale, C.M. A Vegetation index based technique for spatial sharpening of thermal imagery. *Remote Sens. Environ.* **2007**, *107*, 545–558. [[CrossRef](#)]
33. Immerzeel, W.W.; Rutten, M.M.; Droogers, P. Spatial downscaling of TMPA-3B42 precipitation using vegetation response on the Iberian Peninsula. *Remote Sens. Environ.* **2009**, *113*, 362–370. [[CrossRef](#)]
34. Barlow, P.M.; Cunningham, W.L.; Zhai, T.; Gray, M.U.S. Geological Survey Groundwater Toolbox, a graphical and mapping interface for analysis of hydrologic data (Version 1.0)—User guide for estimation of base flow, runoff, and groundwater recharge from streamflow data. *USA Geol. Surv. Tech. Methods* **2014**, *3*, 3–4. [[CrossRef](#)]
35. Wahl, K.L.; Wahl, T.L. Determining the flow of Comal Springs at New Braunfels. *Am. Soc. Civ. Eng.* **1995**, *3*, 77–86.
36. Beck, H.E.; de Jeu, R.A.M.; Schellekens, J.; Dijk, A.I.J.M.; van Bruijnzeel, L.A. Improving Curve Number Based Storm Runoff Estimates Using Soil Moisture Proxies. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2010**, *2*, 250–259. [[CrossRef](#)]
37. Huang, M.; Gallichand, J.; Wang, Z.; Goulet, M. A modification to the Soil Conservation Service curve number method for steep slopes in the Loess Plateau of China. *Hydrol. Process.* **2006**, *20*, 579–589. [[CrossRef](#)]
38. Jain, M.K.; Mishra, S.K.; Singh, V.P. Evaluation of AMC-dependent SCS-CN-based models using watershed characteristics. *Water Resour. Manag.* **2006**, *20*, 531–552. [[CrossRef](#)]
39. Mishra, S.K.; Chaudhary, A.; Shrestha, R.K.; Pandey, A.; Lal, M. Experimental verification of the effect of slope and land use on SCS runoff curve number. *Water Resour. Manag.* **2014**, *28*, 3407–3416. [[CrossRef](#)]
40. Mishra, S.K.; Pandey, R.P.; Jain, M.K.; Singh, V.P. A rain duration and modified AMC-dependent SCS-CN procedure for long duration precipitation-runoff events. *Water Resour. Manag.* **2008**, *22*, 861–876. [[CrossRef](#)]
41. Viji, R.; Prasanna, P.R.; Ilango, R. Modified SCS-CN and Green-Ampt Methods in Surface Runoff Modelling for the Kundahpallam Watershed, Nilgiris, Western Ghats, India. *Aquat. Procedia* **2015**, *4*, 677–684. [[CrossRef](#)]
42. Barazzuoli, P.; Izzo, S.; Menicoral, P.; Micheluccini, M.; Salleolini, M. A New Practical Aid to Regional Hydrogeologic. *Environ. Manag.* **2006**, *13*, 15–16.
43. Ghiglieri, G.; Carletti, A.; Pittalis, D. Runoff coefficient and average yearly natural aquifer recharge assessment by physiography-based indirect methods for the island of Sardinia (Italy) and its NW area (Nurra). *J. Hydrol.* **2014**, *519*, 1779–1791. [[CrossRef](#)]
44. Grillone, G.; Baiamonte, G.; D’Asaro, F. Empirical Determination of the Average Annual Runoff. *Am. J. Appl. Sci.* **2014**, *11*, 89. [[CrossRef](#)]
45. Spadoni, M.; Brilli, M.; Giustini, F.; Petitta, M. Using GIS for modelling the impact of current climate trend on the recharge area of the S. Susanna spring (Central Apennines, Italy). *Hydrol. Process.* **2010**, *24*, 50–64. [[CrossRef](#)]
46. Tardi, A.; Vittorini, S. Il calcolo del coefficiente di deflusso su basi fisiografiche: il coefficiente di deflusso calcolato per il bacino dell’Ombrone Grossetano. *Boll. Degli Ing.* **1977**, *4*, 5–14.
47. MATLAB (R2019a). Natick, Massachusetts; The MathWorks Inc.: Natick, MA, USA, 2019.
48. Coleman, T.F.; Yin, Z. *Optimization Toolbox™ User’s Guide R 2015 a*. 780; The MathWorks Inc.: Natick, MA, USA, 2015.

