Water **2015**, *7*, 747-779; doi:10.3390/w7020747

OPEN ACCESS

water

ISSN 2073-4441

www.mdpi.com/journal/water

Article

Improvement of Hydrological Simulations by Applying Daily Precipitation Interpolation Schemes in Meso-Scale Catchments

Mateusz Szcześniak [1,†,*] and Mikołaj Piniewski [1,2,†]

[1] Department of Hydraulic Engineering, Warsaw University of Life Sciences—SGGW, ul. Nowoursynowska 159, Warsaw 02-776, Poland; E-Mail: M.Piniewski@levis.sggw.pl

[2] Potsdam Institute for Climate Impact Research, P.O. Box 601203, Potsdam 14412, Germany

[†] These authors contributed equally to this work.

[*] Author to whom correspondence should be addressed; E-Mail: M.Szczesniak@levis.sggw.pl; Tel.: +48-22-593-5267.

Academic Editor: Miklas Scholz

Received: 14 November 2014 / Accepted: 2 February 2015 / Published: 12 February 2015

Abstract: Ground-based precipitation data are still the dominant input type for hydrological models. Spatial variability in precipitation can be represented by spatially interpolating gauge data using various techniques. In this study, the effect of daily precipitation interpolation methods on discharge simulations using the semi-distributed SWAT (Soil and Water Assessment Tool) model over a 30-year period is examined. The study was carried out in 11 meso-scale (119–3935 km^2) sub-catchments lying in the Sulejów reservoir catchment in central Poland. Four methods were tested: the default SWAT method (Def) based on the Nearest Neighbour technique, Thiessen Polygons (TP), Inverse Distance Weighted (IDW) and Ordinary Kriging (OK). =The evaluation of methods was performed using a semi-automated calibration program SUFI-2 (Sequential Uncertainty Fitting Procedure Version 2) with two objective functions: Nash-Sutcliffe Efficiency (NSE) and the adjusted R^2 coefficient (bR^2). The results show that: (1) the most complex OK method outperformed other methods in terms of NSE; and (2) OK, IDW, and TP outperformed Def in terms of bR^2. The median difference in daily/monthly NSE between OK and Def/TP/IDW calculated across all catchments ranged between 0.05 and 0.15, while the median difference between TP/IDW/OK and Def ranged between 0.05 and 0.07. The differences between pairs of interpolation methods were, however, spatially variable and a part of this variability was attributed to catchment properties: catchments characterised by low station density and low

coefficient of variation of daily flows experienced more pronounced improvement resulting from using interpolation methods. Methods providing higher precipitation estimates often resulted in a better model performance. The implication from this study is that appropriate consideration of spatial precipitation variability (often neglected by model users) that can be achieved using relatively simple interpolation methods can significantly improve the reliability of model simulations.

Keywords: spatial interpolation; rainfall; SWAT; watershed model; Sulejów reservoir; Pilica; temporal scale; spatial scale; station density

1. Introduction

The growing needs in the field of hydrological modelling necessitate the continual improvement of existing hydrological models. One of the most commonly used hydrological models nowadays is the semi-distributed Soil and Water Assessment Tool (SWAT; [1]). Gassman [2] reported that the model has been applied on almost every continent and that SWAT or SWAT model spinoff applications have been the subject of around 1075 peer-reviewed articles in nearly 216 different journals.

Temporally and spatially variable precipitation is the major driving force in all (semi-)distributed hydrological models, including SWAT. Its temporal variability is fundamental for hydrological modelling and has been discussed many times elsewhere. Likewise, there is a large and still growing body of literature on the role and effect of the spatial distribution of precipitation in hydrological modelling [3–8]. In general, there are several types of potential data sources: (1) station data from precipitation gauges; (2) reanalysis data and (3) radar data. Gauge data are traditionally the most widely used data type for hydrological modelling. However, they pose multiple problems such as gauge undercatch [9] and the high costs of supporting dense networks of gauges which are crucial for reliable areal precipitation estimates [10], in particular during intensive and spatially variable rainfall events causing flash floods [11]. Reanalysis data products such as the WATCH Forcing Data (Water and Global Change, "WFD"; [12]) are promising in that they usually cover long time periods (100 years in the case of the WFD), but their spatial resolution is not sufficient for modelling of small and medium-sized catchments. Finally, (high-resolution) radar data do not have drawbacks of gauge and reanalysis data and are a valuable asset for the hydrological modelling community [5,13–15]. Nevertheless, their use often implies using gauge data as well in order to find the optimal parameters of the transformation equation [16]. Furthermore, their accessibility in many countries is low, and is, in particular, not sufficient for simulation periods covering a few decades, e.g., a climate normal 30-year period.

Hence, in many cases, the primary source of precipitation data in hydrological modelling, in particular for the climate normal period in small and medium-sized catchments, is still ground-based measurements. Since distributed models require spatially variable precipitation, there is a need for spatial interpolation of gauge data. The use of different interpolation methods may result in significant differences from the actual spatial distribution of precipitation.

Several authors have compared different spatial interpolation methods of precipitation data for hydrological modelling. Most notable are: the review paper of Ly *et al.* [17] and works of

Hartkamp *et al.* [18], Goovaerts [19], Zhang and Srinivasan [20], Tabios and Salas [21] which compare a number of different methods, of which the most frequent are Thiessen Polygons, Inverse Distance Weighted and Ordinary Kriging. Evaluation of methods is typically done by applying an independent or cross-validation [17], which makes it possible to estimate and compare prediction errors. In general, more complex methods give better results, e.g., Ordinary Kriging outperforms Inverse Distance Weighted which surpasses the Thiessen Polygons method. The results also showed that different methods give similar average areal precipitation values, but often differ in extreme values.

An alternative approach in evaluating different interpolation schemes is to validate them in hydrological models [17], but this has been done less frequently. The main advantage of this approach is that it evaluates area-integrated precipitation rather than point measurements, which makes it more fit-for-purpose in terms of hydrological modelling. Technically, this approach can be applied by preparing a number of different precipitation input alternatives obtained using different interpolation schemes, then executing (and often calibrating) the model for each alternative and comparing/evaluating the results. One of the first works of this type was that of Haberlandt *et al.* [22], who applied the SLURP model in the Mackenzie River Basin in Canada. They concluded that more complex interpolation techniques and the use of combined precipitation data help to improve discharge simulations. Furthermore, the relative size of the simulation units was the factor that explained these improvements. Masih *et al.* [23] applied SWAT in the Karkheh basin in Iran, concluding that the IDEW (Inverse Distance Elevation Weighted) method outperformed the default method mainly in small sub-catchments in the range 600–1600 km^2. For larger catchments (above 5000 km^2), no significant difference in performance between studied methods was reported.

By default, precipitation input data in SWAT are processed by a rather simple, Nearest Neighbour-based method, in which each sub-basin is assigned data from the nearest stations. The distance from the station to a certain sub-basin is calculated based on the location of its centroid. Since this is the default method, it is presumably used by most SWAT users, even though it is expected that the model performance is likely to be affected in this case. Our hypothesis is that applying spatial interpolation of precipitation prior to reading input data in SWAT should improve its performance in discharge simulation.

The main goal of this paper is to verify whether selected spatial interpolation techniques can improve the performance of the SWAT model in predicting daily and monthly flows over a climate normal period in a set of meso-scale catchments of different size. The term meso-scale catchments refers to catchments whose order of magnitude lies between 10 and 10^3 km^2 [24]. The secondary goal is to analyse the effect of certain catchment properties in order to explain the between-catchment differences in results. Even though there exists a body of literature on this topic (e.g., [22,23,25–28]), our study brings certain improvements in methodological design, by combining simultaneously the following features:

- We perform interpolation of daily precipitation data for a climate normal period (in contrast to many other studies that used a monthly or annual time step, e.g., [19,29–33], or a daily time step for a much shorter period, e.g., [20]). Long simulation periods are recommended for model application for climate change impact assessment [34];
- We include both conventional interpolation methods such as Thiessen Polygons or Inverse Distance Weighted and geostatistical methods such as kriging (in contrast to Masih *et al.* [23] and Hwang *et al.* [26], who did not include geostatistical methods in their comprehensive studies);

- We focus on the effect of different interpolation methods on hydrology (in contrast to some interesting papers that limit their attention to the comparison of performance of interpolators, e.g., [17,20,35]) and evaluate this effect using a semi-automated SUFI-2 (Sequential Uncertainty Fitting) algorithmin 11 catchments spanning in size between 119 and 3935 km^2. Taking advantage of the relatively large number of studied catchments (compared to other studies that usually focused on one catchment), we investigate the influence of certain catchment characteristics on evaluation results, which to our knowledge has not been done to date;
- We present the results of our analysis for both daily and monthly time step aggregations; Many similar studies limited their attention to either only monthly (e.g., [22,27]) or only daily (e.g., [23,28]) time steps.

2. Materials and Methods

2.1. Study Area

The study was conducted in 11 sub-catchments of the Sulejów reservoir catchment (hereafter referred to as the SRC), *i.e.*, a part of the Pilica catchment, situated upstream of the dam on Sulejów reservoir, built in 1974. Sulejów reservoir is located in central Poland and its total drainage area consists of catchments of two main inflowing rivers: the Pilica and the Luciąża rivers and direct sub-catchment with several smaller reaches (Figure 1). This is a large reservoir (95 × 10^6 m^3 of total capacity) but its effect on river flow is outside the scope of this paper, because all catchments studied in this paper (*i.e.*, sub-catchments of the SRC) are situated upstream of it. The SWAT model is however set up for the whole SRC, whose total area equals 4928 km^2, of which the Pilica catchment contributes nearly 80% and the Luciąża catchment nearly 16% [36].

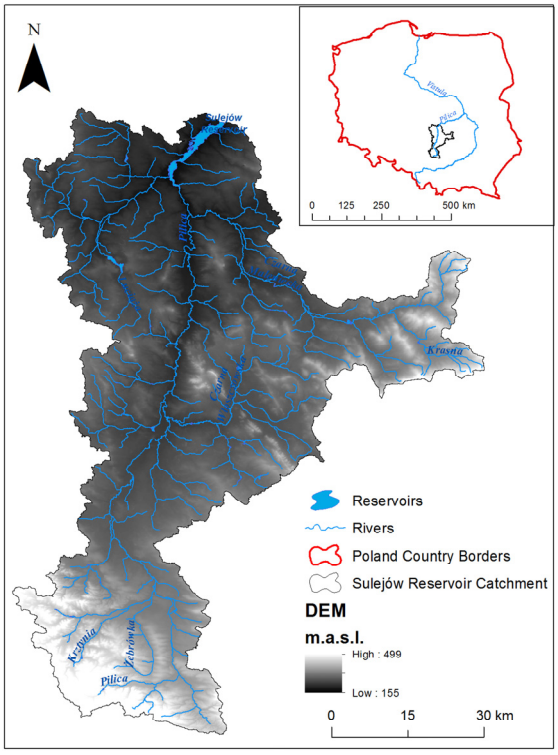

Figure 1. Study area: Sulejów Reservoir Catchment (SRC) and its location in Poland.

Elevation in the SRC varies from 154 m a.s.l. in lowland areas in the North to 499 m a.s.l. in the highland areas in the South. Land cover in the catchment is distributed as follows: 44.4% arable land, 38.6% forest areas, 12.3% grasslands, 4.7% urban and the rest occupied by other land covers types (data according to Corine Land Cover 2006). Loamy sands and sands predominate in the soil cover. Climate is typical for central Poland with a mean annual temperature of *ca.* 7.5 °C, and mean January/July temperature equal to −4 and 18 °C, respectively. Mean annual precipitation is *ca.* 600 mm, with the highest totals in June/July and the lowest ones in January.

As for the hydrologic regime, floods occur usually due to snowmelt but quite often also due to rainfall events in summer, which usually is a low flow period. There is relatively low pressure on water resources: very small water consumption by industry, households, irrigation and locally moderate consumption for fish farming; the largest reservoir Cieszanowice constructed in 1998 on the Luciąża river has the total capacity of 7.3×10^6 m^3. Such a quasi-natural character of the SRC makes it a suitable study area for hydrological modelling.

2.2. SWAT Model

2.2.1. General Features

SWAT is a public domain, river basin scale model developed to quantify the impact of land management practices in large, complex river basins [1]. SWAT2009 rev. 591 model version [37,38] under ArcSWAT 2012.10_0.1 [39], an ArcGIS-ArcView extension and graphical user input interface for SWAT, was used in this study. SWAT is a physically based, semi-distributed, continuous time model that simulates the movement of water, sediment, nutrients, pesticides and bacteria on a catchment scale. It can operate on sub-daily, daily, monthly and yearly time step. The SWAT model river basin can be divided into sub-basins based on the Digital Elevation Model (DEM) or predefined stream network layer and a threshold that defines the minimum drainage area required to form the origin of a stream. The basic unit of discretization in SWAT is so-called "hydrological response unit" (HRU) which is a unique compound of soil, land use and slope overlay and one of the most important features of SWAT is the fact that HRUs are lumped within sub-basins. Each HRU runoff is simulated individually and then aggregated to the sub-basin level. In order to acquire the total runoff for the river basin, runoff from HRUs is routed through the stream network to the main outlet.

2.2.2. Model Setup

The GIS input data required to build the SWAT model setup include Digital Elevation Model (DEM), land cover map, and soil map. The following input data were used to develop this setup:

- DEM based on the ASTER satellite data, 1:25,000 topographic map and Regional Water Management Authority (RZGW) water cadastre.
- Land cover data derived from reclassified Corine Land Cover 2006 (CLC2006) available from General Directorate of Environmental Protection (GDOŚ).
- Soil map composed of a 1:100,000 digital map from Institute of Soil Science and Plant Cultivation (IUNG) and 1:25,000 soil map available from Regional Directorate of State Forests (RDLP).

The ArcSWAT interface was used to delineate the catchment into 272 sub-basins. In the next step, the overlay of land cover and soil maps resulted in 3401 Hydrologic Response Units (HRU).

The meteorological data required by SWAT (precipitation, solar radiation, relative humidity, wind speed, maximum and minimum temperatures) were acquired from the Institute of Meteorology and Water Management-National Research Institute (IMGW-PIB). Meteorological data (except precipitation which will be described in the next section) were available from six synoptic (highest level) and 11 climate stations. This data was interpolated by the Thiessen Polygon method (which will also be described in the next section) and assigned to all sub-basins.

2.3. Precipitation Data and Interpolation Methods

The total number of stations with precipitation observations obtained from IMGW-PIB was 49 but not all of them had the full set of data. In three cases, when two stations were located next to each other (within 3 km distance) and had complementary periods of data availability, they were "merged" into a single one with the same coordinates as the station with a longer period of record. Spatial distribution of the stations is presented in Figure 2, whereas periods of data availability can be found in Table S1.

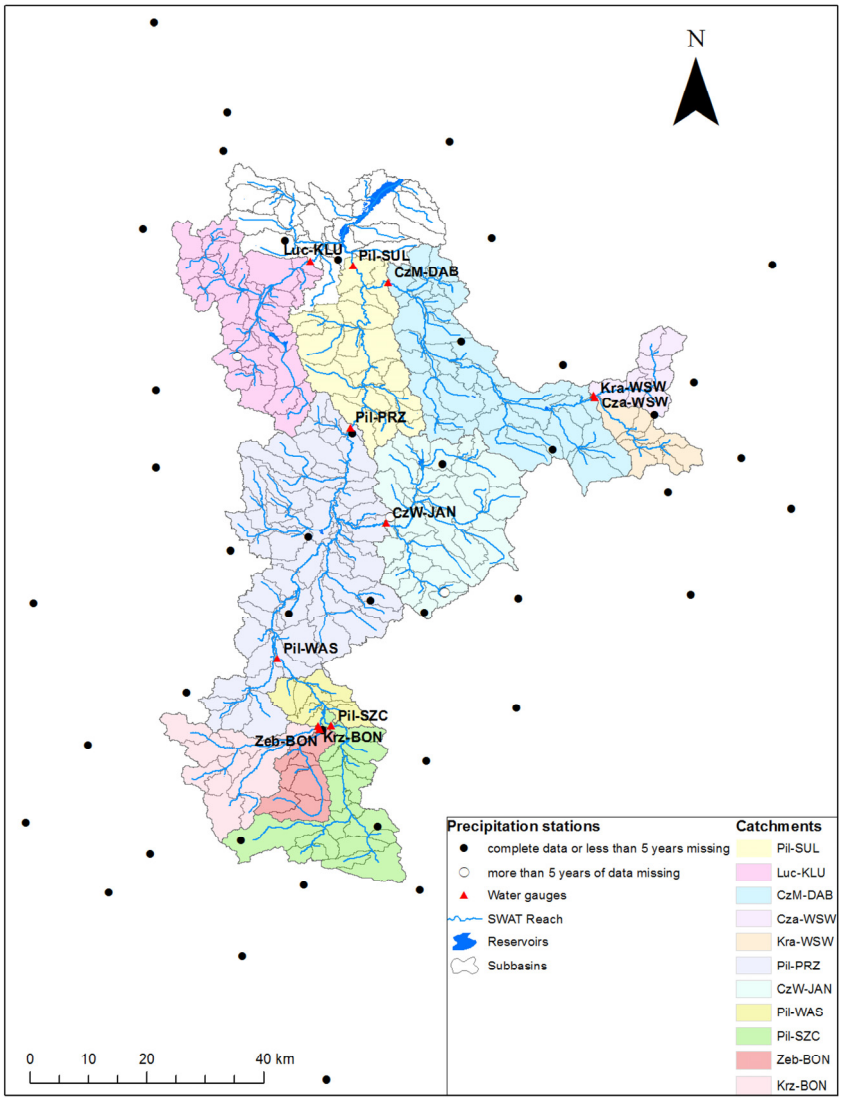

Figure 2. Location of flow gauges and precipitation stations in the Sulejów Reservoir catchment.

In selection of interpolation methods evaluated in this paper, we have taken the model user perspective. Selection was based on the popularity of different methods in research papers on precipitation interpolation. Obviously, the method based on the Nearest Neighbour technique that is used as default in SWAT was tested in first instance. Three other tested methods were: Thiessen Polygons, Inverse Distance Weighted and Ordinary Kriging. The literature review showed that the number of applications of these three methods largely exceeded respective numbers for other methods.

The first method tested in this paper was the default SWAT method (hereafter referred to as "Def") shortly mentioned in the Introduction). It is a version of the least sophisticated Nearest Neighbour ("NN") technique. NN applies the values to unknown points such that:

$$P(x_0) = P(x_k)\ for\ d(x_0, x_k) = \min_i d(x_0, x_i) \quad (1)$$

where $P(x_0)$ is interpolated value in the point x_0, $P(x_k)$ represents the measured value in the point x_k and $d(x_0, x_k)$ denotes the Euclidean distance between the points x_0 and x_k.

In the Def method, observed values are applied to the centroids of each sub-basin as in the NN method. Therefore, inputs from the nearest station are used throughout the whole sub-basin. An important consideration about this method is that it uses SWAT built-in Weather Generator ("WXGEN"; [40]) that produces synthetic daily inputs based on user-provided long-term weather statistics in order to fill in the missing values if they are present in the time series (*cf.* Table S1 and Figure 2). Precipitation statistics for the Weather Generator were calculated for all stations and loaded into SWAT.wgn files.

The Def method will be used as a point of reference throughout this paper. It is assumed that this is the method that is used by majority of SWAT users. All other interpolation methods described in this paper were executed externally of ArcSWAT so the influence of using them will be evaluated mainly in comparison with the Def method.

The method number 2 was Thiessen Polygons ("TP"), developed by Thiessen [41]. This method uses the same principle as NN, however, instead of assigning values to sub-basin centroids (as in Def), for each sub-basin a weighted average is calculated from values belonging to polygons intersecting a given sub-basin.

The method number 3 was Inverse Distance Weighted ("IDW"). In this method, values are assigned to the unknown points based on the calculated weighted average of the known point values. The weights of the known values are inversely proportional to distance from the known point to the estimated point. General equation for Inverse Distance Weighted is [42]:

$$P_x = \sum_{i=1}^{n} w_i P_i \quad (2)$$

where P_x represents interpolated precipitation value in the point *x*, w_i is a weight assigned to the station and P is observed precipitation on the station *i*. The weights are calculated as follows:

$$w_i = \frac{1}{d_i^k} \quad (3)$$

where d_i is the distance between points *x* and *i* and *k* is an exponent. In this paper, the *k* value was set to 2 that is the most widely used exponent value for IDW applications (e.g., [20,26,27,33]).

IDW spatial interpolation was performed in ArcMap 10.1 with the use of Geostatistical Analyst Tool with the daily time step. Because there were almost 11,000 days to calculate, the model was developed in the ModelBuilder. Values interpolated by the IDW method were saved as rasters (spatial resolution 2 km) and then they were averaged in the sub-basin boundaries using Zonal Statistics as Table tool.

The last method (number 4) of spatial interpolation was Ordinary Kriging ("OK") which belongs to the category of advanced geostatistical techniques that offer the unbiased estimation of the variable at an unobserved location from observations of the random field at nearby locations. The OK is one of the most commonly applied Kriging techniques that is characterized by an unknown and constant trend. The OK estimates the unknown values at a given location as a weighted linear combination of neighbouring observations [19].

The main goal in the OK method is to define the spatial correlation of the analysed process for the best estimation of the output surface. In this study, we used a commonly applied semivariance, which represents the spatial variation set against the distance or separation of input sample points [43]. The empirical semivariance $y(h)$ is computed from the input data, as follows: [19]:

$$y(h) = \frac{1}{2N}\sum_{i=1}^{N}[Z(x_i) - Z(x_i + h)]^2 \quad (4)$$

where, N is the number of possible pairs of points, $Z(x_i)$ is the value in output location and $Z(x_i + h))$ is the value in location moved by vector h. Next, a theoretical, continuous function (curve) needs to be fitted to the empirical semivariance. Finally, predictions at unmeasured locations are made using a similar formula as for IDW (the measured values closest to the unmeasured locations have the most influence). However, kriging weights come from a semivariogram and the spatial arrangement of measured values that are nearby.

Spatial interpolation by the OK method was performed in ArcMap Geostatistical Analyst tool using a similar approach as in the case of the IDW interpolation. As in the previous case, the ModelBuilder feature was applied and values interpolated by the OK method were saved as rasters (spatial resolution 2 km) and then were averaged in the sub-basin boundaries using Zonal Statistics as Table tool.

The Gaussian model was used as the primary semivariance model. Additionally, the Rational Quadratic model was used in days when the Gaussian model generated negative values of interpolation results. This kind of semivariance model change approach is similar to the one that was used by Ly *et al.* [44]. Negative values of interpolation results may come from negative weights assigned to very high precipitation values on distant stations. The change of the semivariance model in case of negative interpolation values was also applied to the model of interpolation developed in ModelBuilder.

Because finding the best parameters of semivariance is time-consuming, complex and hard to automatize—especially for the 30-year period with daily time step—an automatic estimation option was used in ArcGIS Geostatistical Analyst. The following parameters were automatically optimised: range of influence, sill and nugget. Other parameters that were set prior to interpolation were: measurement error set to 20%, as average uncertainty of precipitation measuring instruments given by Larson and Peck [45] and lag size that was set to 9500 m after applying the Average Nearest Neighbour tool as suggested in ArcGIS help [46]. Furthermore, OK in contrast to other used methods allows to calculate prediction standard errors and visualise them on maps, which is a useful tool in validating the interpolation results and uncertainty of prediction.

It should be noted, that there is one aspect in which the TP, IDW and OK methods differ from the Def method: they do not use the Weather Generator function built in SWAT. In case of missing values on a given day for a given station, estimated values are calculated only from known, observed data, thus stations with missing data are neglected in interpolation.

2.4. Precipitation Station Density Factor

As shown in Figure 2, the density of precipitation stations is variable over the area, which could potentially affect the results. We applied a Kernel Density ("KD") function in ArcGIS [43,47] in order to calculate a smooth surface being a proxy of precipitation stations density. The KD function calculates the density of features in a neighbourhood around those features and expresses the result as magnitude per unit area. The KD function has two important parameters: Population field that enables to count some features more heavily than others, and search radius that defines the size of the circular neighbourhood. We used the number of available years for each station as the Population field and set the search radius to 20,000 m. Kernel density function generates a smooth surface (Figure 3A), whereas our interest is in mean values across each catchment upstream of a given flow gauging station. Hence, Figure 3B shows the calculated catchment-averaged values (hereafter denoted as KD), which can be viewed as a proxy for station density within a catchment. This index will be used as one of the catchment properties in an attempt to explain the differences in results between different catchments.

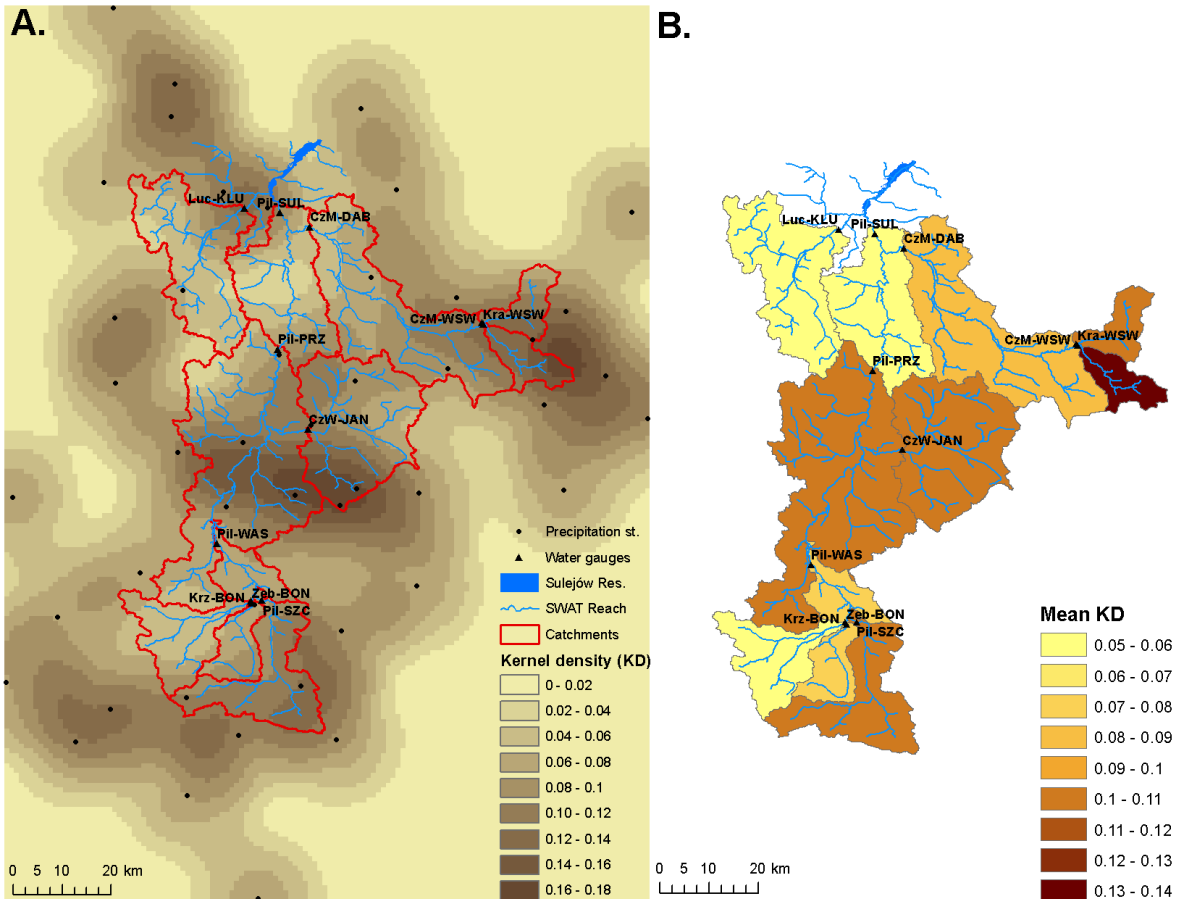

Figure 3. A surface of kernel density function (**A**) applied for estimation of station density within each catchment (**B**) expressed in units (number of stations) per km^2.

2.5. Strategy for Evaluation of Hydrological Simulations

Four separate model setups of the SRC were prepared using four different precipitation data input alternatives and holding all other parameters and inputs unchanged. We will hereafter refer to each of the four setups (input/method alternatives) as "scenarios" with the following abbreviations: (1) Def (Default SWAT method); (2) TP (Thiessen Polygons); (3) IDW (Inverse Distance Weighted) and (4) OK (Ordinary Kriging). In each case the model was run for a time period 1982–2011 with two first years treated as a warm-up period.

The main goal of this analysis was the assessment of the impact of different precipitation interpolation methods on the SWAT model performance in discharge simulation across different temporal and spatial scales. It is well known that hydrological models require calibration and validation prior to real application. Calibration of the SWAT model consists of adjustment of parameters (chosen in sensitivity analysis) within predefined ranges to the point when simulated and observed discharge values are satisfactory. This process can be very complex and time-consuming, especially in catchments with numerous water gauges [48]. Nowadays, with increasing computational power, calibration is usually executed using various automatic techniques. The most widely used SWAT calibration program is SWAT-CUP (Soil and Water Assessment Tool-Calibration and Uncertainty Programs) [49]. However, calibration itself was not the goal of this paper; it was rather a way of setting up a semi-automated testing model created using different interpolation methods against the observed discharge data.

2.5.1. SWAT-CUP and SUFI-2

SWAT-CUP is a freeware program which allows to use several different algorithms for optimization of the SWAT model. SWAT-CUP allows sensitivity analysis, calibration, validation and uncertainty analysis [49]. In this paper we applied SWAT-CUP version 2009 4.3 and selected an optimization algorithm SUFI-2 (Sequential Uncertainty Fitting Procedure Version 2), which is a kind of inverse modelling program that contains elements of calibration and uncertainty analysis [50]. Parameters and their initial ranges applied in SUFI-2 (Table 1) were chosen based on the previous applications of the SWAT model in Polish conditions [48,51], and on sensitivity analysis performed in the SRC.

SWAT-CUP enables testing various objective functions in calibration. We used two different objective functions in order to evaluate different interpolation methods. The first one was Nash–Sutcliffe model efficiency coefficient ("NSE"). NSE can range from $-\infty$ to 1, where 1 is the optimal value. Values between 0 and 1 are acceptable and below 0 provide unsatisfactory results of simulation. Further thresholds for NSE in different contexts were provided e.g., by Moriasi *et al.* [52]. NSE is calculated as follows:

$$NSE = 1 - \frac{\sum_i (Q_m - Q_s)_i^2}{\sum_i (Q_{m,i} - \bar{Q}_m)^2} \quad (5)$$

where Q_m is the mean of observed discharges, and Q_s is simulated discharge.

The second objective function was bR^2 which is the coefficient of determination R^2 multiplied by the coefficient of the regression line b. This modified coefficient of determination allows accounting for the discrepancy in the magnitude of two signals (depicted by b) as well as their dynamics (depicted by R^2) [50]. It can range between 0 and 1, where 1 is the optimal value. Equation for bR^2 is [50]:

$$bR^2 = \begin{cases} |b|R^2 \ if \ |b| \le 1 \\ |b|^{-1}R^2 \ \ if \ |b| > 1 \end{cases} \quad (6)$$

where b is coefficient of the regression line, coefficient of determination is calculated as:

$$R^2 = \frac{\left[\sum_i (Q_{m,i} - \bar{Q}_m)(Q_{s,i} - \bar{Q}_s)\right]^2}{\sum_i (Q_{m,i} - \bar{Q}_m)^2 \sum_i (Q_{s,i} - \bar{Q}_s)^2} \quad (7)$$

The optimal simulations found using both objective functions were also evaluated in terms of the root mean square error (RMSE) and percent bias (PBIAS), which measures the average tendency of the modelled data to be larger or smaller than their observed counterparts. Positive values indicate model underestimation bias, and negative values indicate model overestimation bias [53].

Table 1. Parameter definitions and initial ranges used in SUFI-2.

Name	Lower Limit	Upper Limit	Definition
ESCO.hru [2]	0.7	1	Soil evaporation compensation factor
EPCO.hru [2]	0	1	Plant uptake compensation factor
SOL_Z().sol [1]	−0.4	0.4	Depth from soil surface to the bottom of layer
SOL_AWC().sol [1]	−0.4	0.4	Available water capacity of the soil layer
SOL_BD().sol [1]	−0.4	0.4	Moist bulk density
SOL_K().sol [1]	−0.9	2	Saturated hydraulic conductivity
HRU_SLP.hru [1]	−0.3	0.3	Average slope steepness
ALPHA_BF.gw [2]	−0.9	2	Baseflow alpha factor
GW_DELAY.gw [2]	50	400	Groundwater delay time
GWQMN.gw [2]	0	1000	Threshold depth of water in the shallow aquifer required for return flow to occur
GW_REVAP.gw [2]	0.02	0.2	Groundwater "revap" coefficient
RCHRG_DP.gw [2]	0	0.3	Deep aquifer percolation fraction
CN2.mgt [1]	−0.15	0.15	Initial SCS (Soil Conservation Service) runoff curve nr for moisture condition II
SURLAG.bsn [2]	0.3	3	Surface runoff lag coefficient
SLSUBBSN.hru [1]	−0.3	0.3	Average slope length (m)
CH_N2.rte [2]	0.01	0.1	Manning's "*n*" value for the main channel
CH_N1.sub [2]	0.01	0.1	Manning's "*n*" value for the tributary channel (-)
SMTMP.bsn [2]	−2	2	Snow melt base temperature
TIMP.bsn [2]	0	1	Snow pack temperature lag factor
SNOCOVMX.bsn [2]	0	40	Minimum snow water content that corresponds to 100% snow cover

Notes: [1] parameter multiplied by 1 + *r*, where *r* is a number between lower and upper limits; [2] parameter replaced by the new value from the range.

Additionally, two uncertainty coefficients (default in SUFI-2) were used: p-factor and r-factor. *p*-factor denotes the percentage of observations covered in 95% range of uncertainty (95PPU), while *r*-factor indicates the thickness of the 95PPU band divided by the standard deviation of the measured data. In theory, *p*-factor ranges from 0%–100% and *r*-factor from 0 to ∞. The optimal situation is when p-factor is equal to 1 (100%) and *r*-factor to 0, [49]. It is important to note that while NSE and bR^2 refer to one single parameter set that produces the best value of the objective function, *p*- and *r*-factors refer to the whole body of simulations resulting from the 95PPU, not one simulation.

2.5.2. The Observed Data and Catchment Properties

Discharge data (m^3/s) required for model testing in SUFI-2 were used in two temporal aggregations: daily and monthly, most widely used in hydrological modelling. They were obtained from 11 IMGW-PIB flow gauges for the period of 28 hydrological years from 1984–2011 (the hydrological year in Poland begins on 1 November). Names and codes of the gauges, corresponding rivers and available data periods as well as some basic catchment characteristics are presented in Table 2, whereas their location is shown in Figure 2.

Table 2. List of IMGW-PIB flow gauges in the Sulejów reservoir catchment used in this study with data availability and selected catchment properties.

No	Gauge Name	River Name	Code	A (km^2)	Period of Available Data	Years	Flow (m^3/s)		q_m ($m^3/s/km^2$)	c_v (-)
							Mean	St. Dev.		
1	Sulejów	Pilica	Pil-SUL	3934	11/1/1983–10/31/2011	28	21.9	14.9	5.5	0.68
2	Przedbórz	Pilica	Pil-PRZ	2491	11/1/1983–10/31/2011	28	13.5	9.6	5.3	0.71
3	Wąsosz	Pilica	Pil-WAS	974	11/1/2005–10/31/2011	6	5.7	4.7	6.5	0.82
4	Szczeko-ciny	Pilica	Pil-SZC	360	11/1/1983–10/31/2009	26	1.8	1.3	5.0	0.69
5	Kłudzice	Luciąża	Luc-KLU	507	11/1/1983–10/31/2011	28	2.6	2.5	5.2	0.95
6	Dąbrowa	Czarna Maleniecka	CzM-DAB	946	11/1/1983–10/31/2008; 11/1/2009–10/21/2011	27	5.6	5.6	5.8	1.00
7	Wąsosz-Stara Wieś	Czarna Maleniecka	Cza-WSW	119	11/1/1991–10/31/2003	12	0.91	1.2	7.6	1.09
8	Wąsosz-Stara Wieś	Krasna	Kra-WSW	120	11/1/1991–10/31/2011	20	0.77	1.2	6.3	1.51
9	Janusze-wice	Czarna Włoszczowska	CzW-JAN	598	11/1/1983–10/31/2011	28	3.3	4.0	5.5	1.20
10	Bonowice	Żebrówka	Zeb-BON	128	11/1/1983–10/31/2009	26	0.51	0.47	6.5	0.82
11	Bonowice	Krztynia	Krz-BON	256	11/1/1990–10/31/2009	19	1.3	0.68	5.0	0.54

Note: A is the upstream catchment area, q_m is the area-specific runoff and c_v is coefficient of variation in daily flows. St. Dev. is abbreviation of standard deviation

Seven out of 11 gauges have data record longer than 26 years, whereas the other four have shorter availability. No division into calibration and validation periods was made, but instead the whole period of data availability for each station was used for evaluation of methods in SUFI-2. It is also worth noting that some of the analysed catchments are nested, hence only seven out of 11 gauges designate hydrologically independent (non-nested) catchments. Their upstream areas (A) vary from 119–3935 km^2, mean area-specific runoff (q_m) from 4.1–7.6 $m^3/s{\cdot}km^2$ and the coefficient of variation of daily flows (c_v) from 0.54 to 1.51. The two last hydrological characteristics are indices that accumulate the complex interplay of physiographic (e.g., elevation, slope, land cover, soil permeability), climatic (e.g., evapotranspiration, precipitation) and water management (e.g., abstractions, reservoirs) properties. They will be used along with KD index illustrated in Figure 3 in an attempt to explain the differences in results (e.g., objective function values) between different catchments.

2.5.3. Study Design

Eight SUFI-2 projects were created: four for each interpolation method and two for each type of observed flow data temporal aggregation (daily or monthly). For each of these projects, identical input settings were defined. A single iteration of the main SUFI-2 program was composed of 800 SWAT model simulations with different parameter values sampled (in SUFI-2 pre-processing program) across the parameter space using Latin hypercube sampling [49]. In all eight SWAT-CUP projects, after execution of the SUFI-2 program, we executed SUFI-2 post-processing program for each type of objective function (NSE and bR^2) and for each of the 11 catchments (gauging stations) separately. More precisely, when the observed data input file contains data from multiple gauging stations, SUFI-2 enables assigning weights to different stations and thus calculating weighted objective function values:

$$OF(\omega_1, \ldots, \omega_n) = \sum_{i=1}^{n} \omega_i \cdot OF_i \quad (8)$$

where OF_i is the value of objective function for catchment (gauging station) i, OF is a weighted objective function, ω_i is a non-negative weight coefficient ($\sum \omega_i = 1$) and n is the number of catchments (gauging stations; here 11). We have applied 11 different sets of weights for each objective function type: $(1,0 \ldots,0), (0,1,0, \ldots,0), \ldots, (0,0, \ldots,0,1)$, *i.e.*, in consecutive post-processing runs we were assigning the weight equal to 1 for one particular station and 0 for all other stations. This approach allowed us to conduct SUFI-2 procedure for all gauging stations at the same time in a single SWAT-CUP project instead of creating 11 projects and executing SUFI-2 11 times for each studied catchment separately.

SUFI-2 is an iterative procedure in that after each iteration new parameter ranges are suggested and normally the program is executed again with these new ranges with a goal of improving the goodness-of-fit or uncertainty measures. We have also followed this guidance and executed SUFI-2 with new parameter ranges for each interpolation method—temporal aggregation—objective function combination, but since the results appeared to be negligibly different from the set of results coming from the first iteration, we do not present them in this paper. The advantage of this is that another iteration with new parameter ranges could lead to a problem of model calibration compensating for possibly a wrong input. This issue will be further referred to in Discussion.

As indicated before, we treat the Default method as the point of reference for evaluating three interpolation methods: TP, IDW and OK. Let $OF_{X,t}$ denote the value of objective function (NSE or bR^2) for the interpolation method X (Def, TP, IDW or OK) and temporal aggregation t (d for day or m for month). In the first step, since the sample size is small ($n = 11$) and the population cannot be assumed to be normally distributed, we applied a Wilcoxon signed-rank test that is a nonparametric analogue to the paired *t*-test [54]. Eleven catchments and three pairs of interpolation methods (Def *vs.* TP, Def *vs.* IDW and Def *vs.* OK) served as nominal variables, while $OF_{X,t}$ served as measurement variable. The null hypotheses is that the median difference in $OF_{X,t}$ between two given interpolation methods is zero. We applied this test at two significance levels: $p = 0.05$ and $p = 0.1$. Additionally, the $OF_{X,t}$ values will be illustrated as box plots showing the median, interquartile range and minimum/maximum values.

In the second step, we wanted to examine the relationship between catchment properties and the differences in $OF_{X,t}$ between any two methods X and Y hereafter denoted as $\Delta OF_{X,Y,t}$, *i.e.*,

$$\Delta OF_{X,Y,t} = OF_{X,t} - OF_{Y,t} \quad (9)$$

The positive values denote an improvement in model behaviour with method X over the method Y. Pearson r correlation coefficients between catchment properties and $\Delta OF_{X,Y,t}$ and respective significance levels were calculated in each case.

3. Results and Discussion

3.1. Evaluation of Interpolation Results

Spatial variability in mean annual precipitation is visible, but is not very high in the SRC. As shown in Figure 4A, the highest difference between any two catchments for Def equals 61 mm (for three other methods the difference is also *ca.* 60 mm). Clearly, spatial differences at shorter temporal scales usually grow larger. However, the most interesting is the difference (bias) in mean annual precipitation between any two methods X and Y, hereafter denoted as $\Delta P_{X,Y}$:

$$\Delta P_{X,Y} = P_X - P_Y \quad (10)$$

Figure 4B shows the values of $\Delta P_{X,Def}$ calculated for all 11 catchments. In 10 out of 11 cases, there is a positive bias for each method and in one case (KrzBON) it exceeds 60 mm. In seven out of 11 cases, the bias is in the range (−20 mm, 20 mm). The differences between interpolation methods themselves are very small compared to the differences between them and the Def method. Only in three out of 33 cases $\Delta P_{X,Def}$ exceeds 10 mm in terms of the absolute values. There is no general pattern for the sign of the difference, however it is more frequent that OK and IDW have higher values than TP. The analysis of interpolated maps for certain events with high precipitation showed that the differences can be considerably higher than for a long term period showed in Figure 4. The IDW method produced the smoothest pattern, whereas the Default method the sharpest pattern. Which is in agreement with observations of Zhang and Srinivasan [20].

We have also evaluated correlation between station density indicator KD and precipitation bias indicator $\Delta P_{X,Def}$. For each method, we have found a significant (at significance level $p = 0.05$) negative correlation between these two variables, with Pearson correlation coefficients in the range (−0.68, −0.61). This shows that for higher station densities there was a little difference in mean annual precipitation between interpolation methods and the Def method, whereas for lower densities the interpolated mean annual precipitation was higher than precipitation according to the Def method.

Each interpolation method carries a certain amount of uncertainty, but the only method that enables quantifying this uncertainty is the OK. Hence, Figure 5 presents raw precipitation data measured at gauge stations, interpolated values and prediction standard errors ("SE") for four selected wet days with different precipitation patterns. The only common pattern valid for four different days is increasing value of SE in the outward direction from the SRC catchment (*i.e.*, in places with no gauges, where we deal with extrapolation instead of interpolation). It can be seen that on days 9 July 1994 and 7 November 1998 SE is nearly spatially constant, but also significant (*ca.* 4 mm on 9 July 1994 and 6 mm on November 1998). In contrast, on 4 January 1983 mean SE values are much smaller, not exceeding 0.6 mm in any point of the catchment. The most interesting pattern can be observed for the day 17 April 1989, when there is a clear relationship between SE values and gauge density (*cf.* Figure 3A). In the

buffer zone of *ca.* 2 km around each station, SE did not usually exceed 0.8 mm, while in some areas in the catchment where distance to the nearest station exceeded 15 km, SE values increased up to 3.8 mm. This analysis was made for a small number of events, but it shows a large spectrum of possibilities in how prediction errors may look on days with different precipitation patterns.

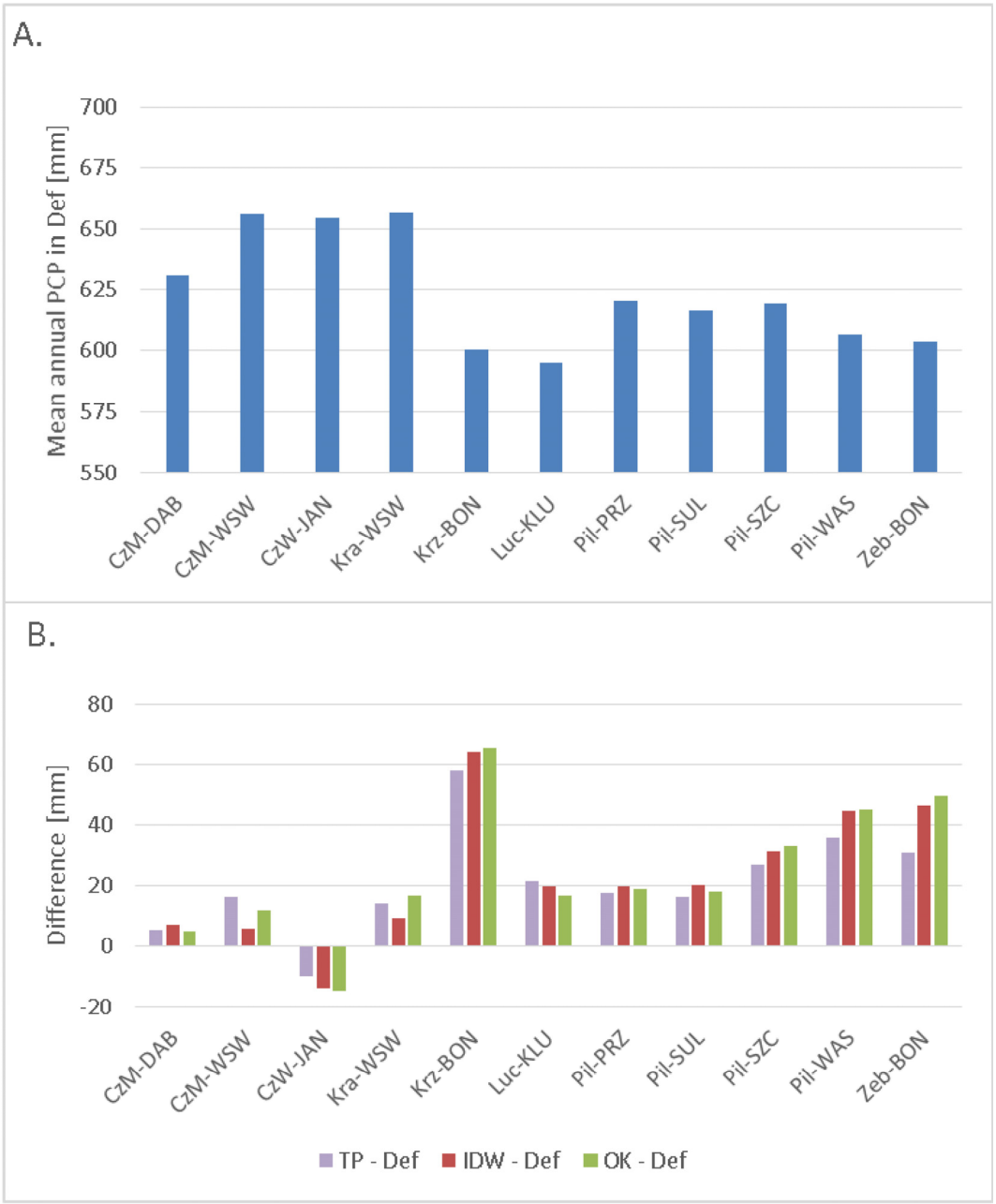

Figure 4. Mean annual precipitation for the period 1982–2011 calculated for 11 studied catchments for the Default method (**A**); The difference in mean annual precipitation between three interpolation methods (TP, IDW, OK) and Def (**B**).

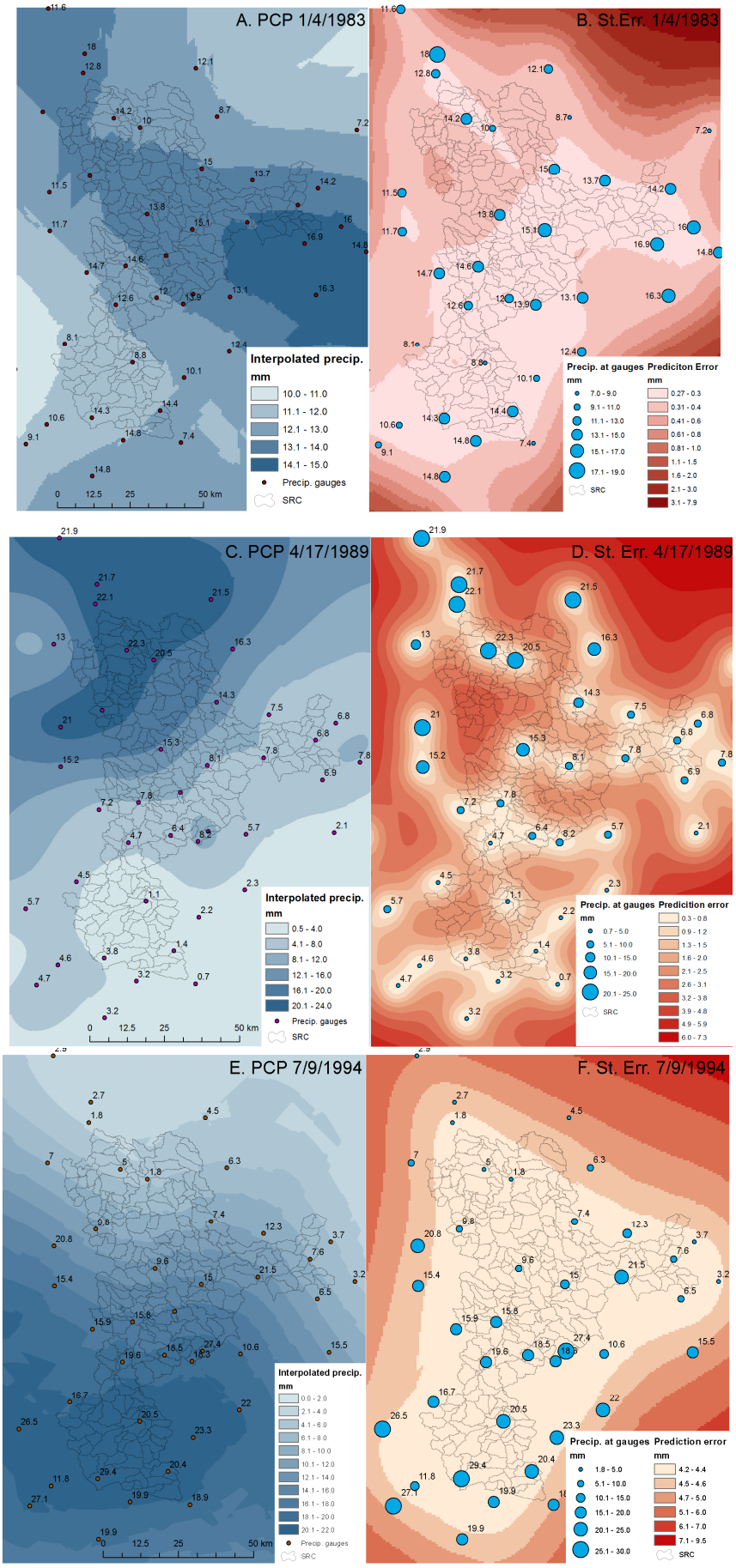

Figure 5. *Cont*

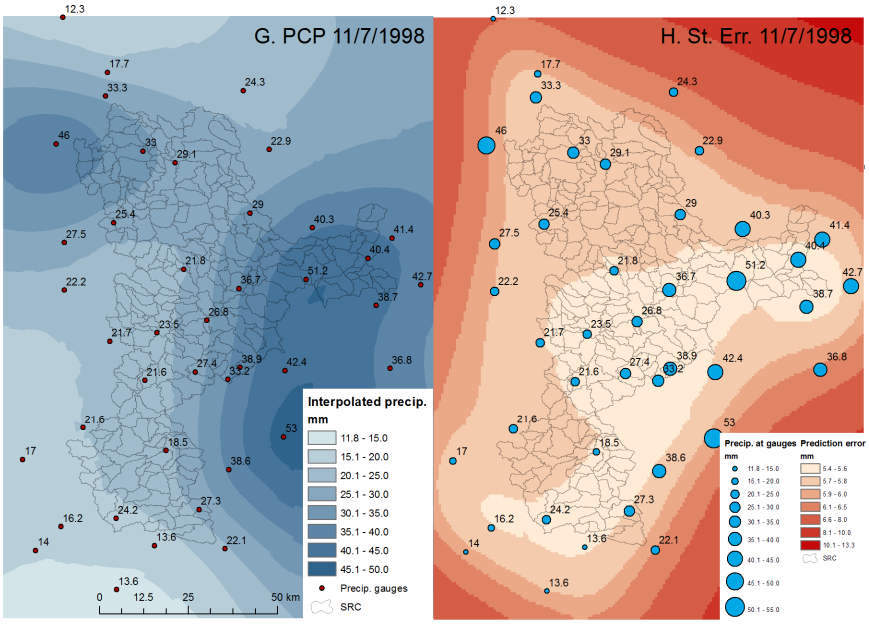

Figure 5. Spatial distribution of precipitation on four selected wet days (4 January 1983; 17 April 1989; 9 July 1994; 7 November 1998) estimated using Ordinary Kriging (**A**,**C**,**E**,**G**) and respective prediction standard errors for the same days (**B**,**D**,**F**,**H**).

3.2. Effect of Interpolation Methods on Model Performance

3.2.1. Statistical Summary for All Catchments

Figure 6 illustrates the $OF_{X,t}$ values obtained for best parameter combinations for different interpolation scenarios and different temporal aggregations. The $OF_{X,t}$ values are depicted as box plots, calculated across 11 analysed gauging stations. A general observation coming from this figure is that precipitation interpolation methods improve, to a variable extent, simulations of daily and monthly discharge across a range of scales. More specifically, OK outperformed all three other methods in terms of daily and monthly NSE (Figure 6a,c). The median difference of $\Delta OF_{OK,Y,d}$ was in this case equal to 0.06, 0.05 and 0.05 for Y being Def, TP or IDW, respectively. The median difference of $\Delta OF_{OK,Y,m}$ was equal to 0.08, 0.05 and 0.15, respectively. In the case of bR^2 all three interpolation methods led to a significant improvement over the default method, however the difference between each of them individually was very small (Figure 6b,d). The median difference of $\Delta OF_{X,Def,d}$ was equal to 0.05 for X being TP, IDW or OK. The median difference of $\Delta OF_{X,Def,m}$ was equal to 0.07 for X being TP or OK and 0.06 for X being IDW. A comparison of results between daily and monthly aggregations leads to a conclusion that the effect is broadly similar and its magnitude is only a bit stronger for monthly aggregation.

Figure 7 shows the box plots of PBIAS calculated for the optimal simulations found using NSE or bR^2 as the objective functions. It can be noted that the positive values of PBIAS largely predominate, which shows that the model generally underestimated the discharge. This underestimation was always higher for bR^2 than for NSE. For NSE, it was significantly lower for the monthly temporal aggregation than for the daily aggregation. For bR^2 no such trend was observed. As regards comparison between methods, the results were quite variable. For monthly NSE, clearly OK produced the values of PBIAS closest to zero as compared to all other methods. For daily NSE, OK was characterised by lower variability of PBIAS

than other methods, particularly lower than IDW. For monthly bR^2, the PBIAS box plot statistics were slightly lower for Def than for other methods, whereas for daily bR^2 no clear trend could be found, even though the variability for IDW and OK was slightly higher than for Def and TP.

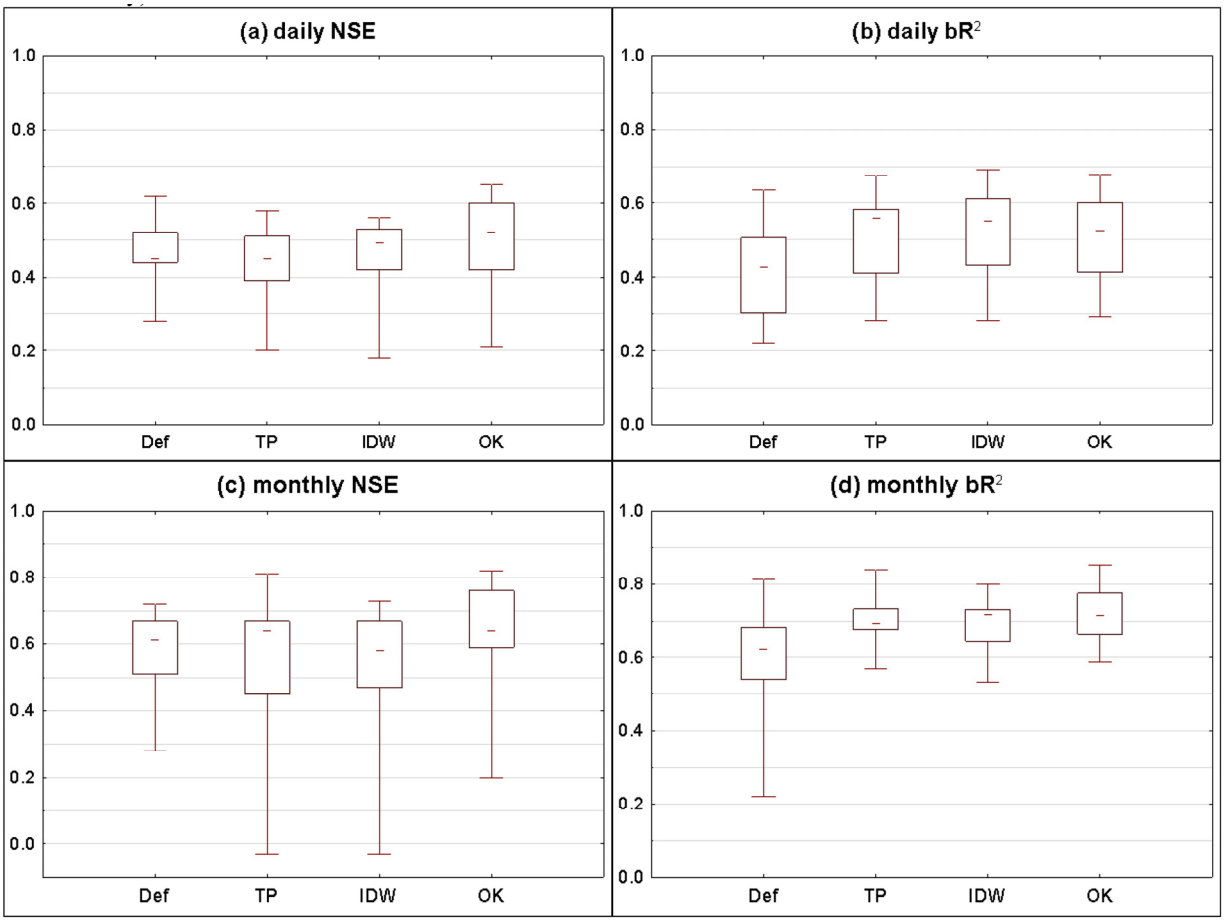

Figure 6. Box plots of selected objective functions across all 11 flow gauging stations for different interpolation methods (Def—Default version, TP—Thiessen Polygons, IDW—Inverse Distance Weighted, OK—Ordinary Kriging) and different temporal aggregations (**a**,**b**: daily; **c**,**d**: monthly).

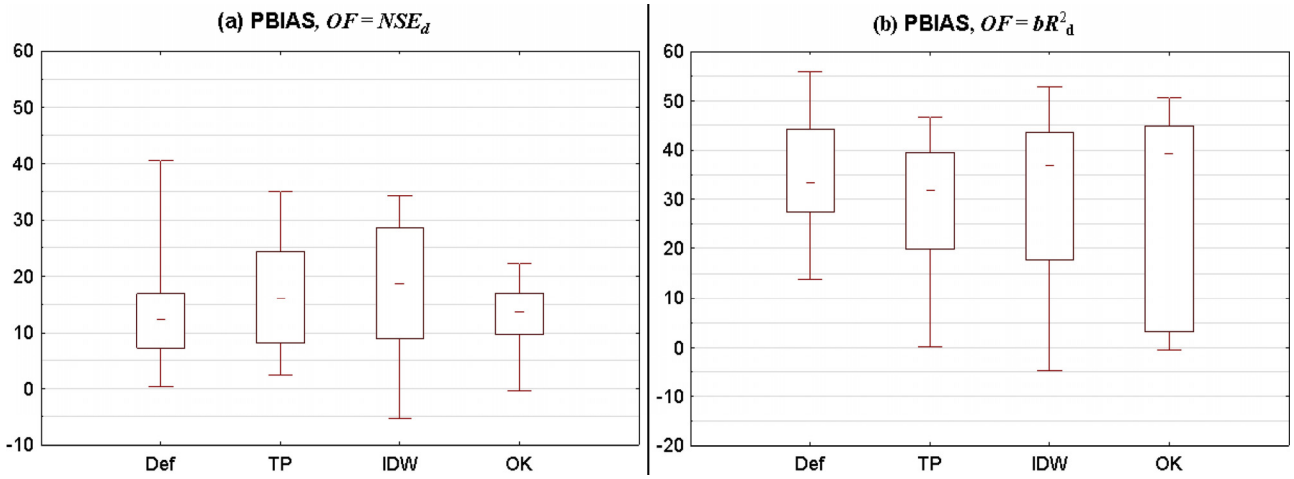

Figure 7. *Cont.*

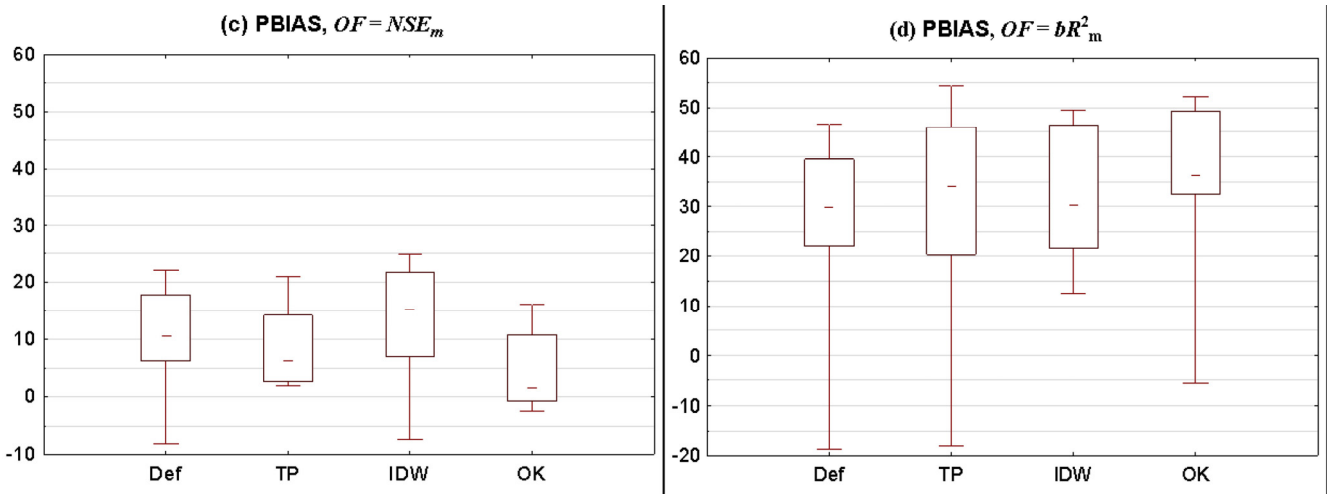

Figure 7. Box plots of PBIAS across all 11 flow gauging stations for different interpolation methods (Def—Default version, TP—Thiessen Polygons, IDW—Inverse Distance Weighted, OK—Ordinary Kriging) and different objective function/temporal aggregation combinations (**a**—NSE_d; **b**—bR_d^2; **c**—NSE_m; **d**—bR_m^2).

We have also evaluated the values of RMSE for each of the optimal solutions found using NSE or bR^2 as objective functions. In general, the values of RMSE were always lower for NSE than for bR^2, and always lower for the monthly temporal aggregation than for daily aggregation. When it comes to the comparison between methods, since RMSE is not dimensionless, it cannot be compared between different catchments. Hence, we have calculated percent changes in RMSE for each pair of methods (Figure 8). As with PBIAS, the results are different for different objective functions. For NSE, RMSE for OK is significantly lower than for all other methods for both temporal aggregations, which is in agreement with observations from Figure 6a,c. For daily bR^2 all interpolation methods are characterised by lower RMSE than Def and OK has slightly lower RMSE than IDW. For monthly bR^2 the results are highly variable between catchments and no clear relationship can be distinguished.

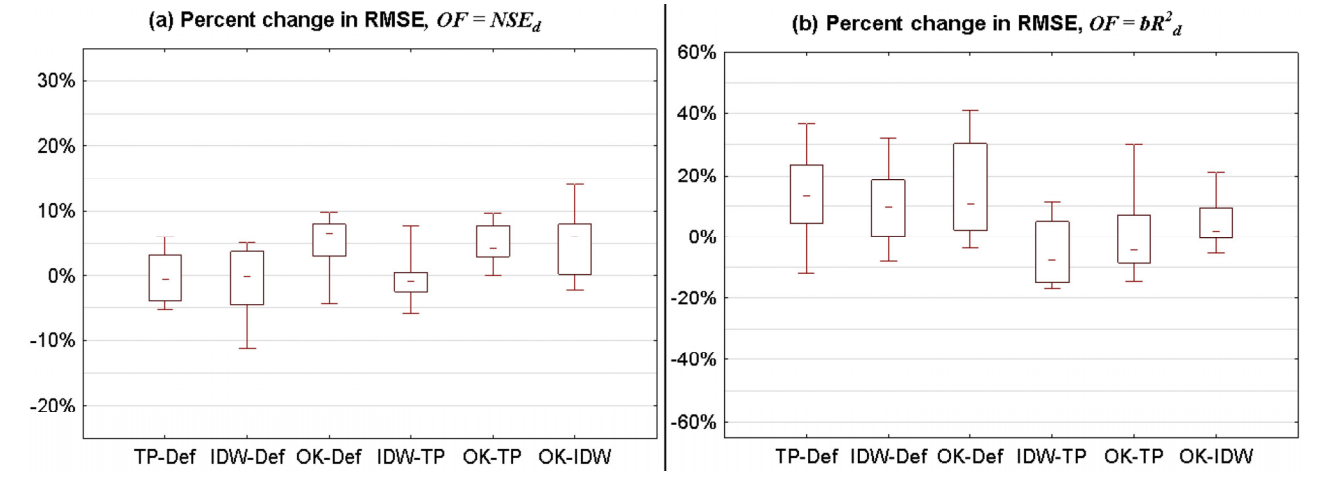

Figure 8. *Cont.*

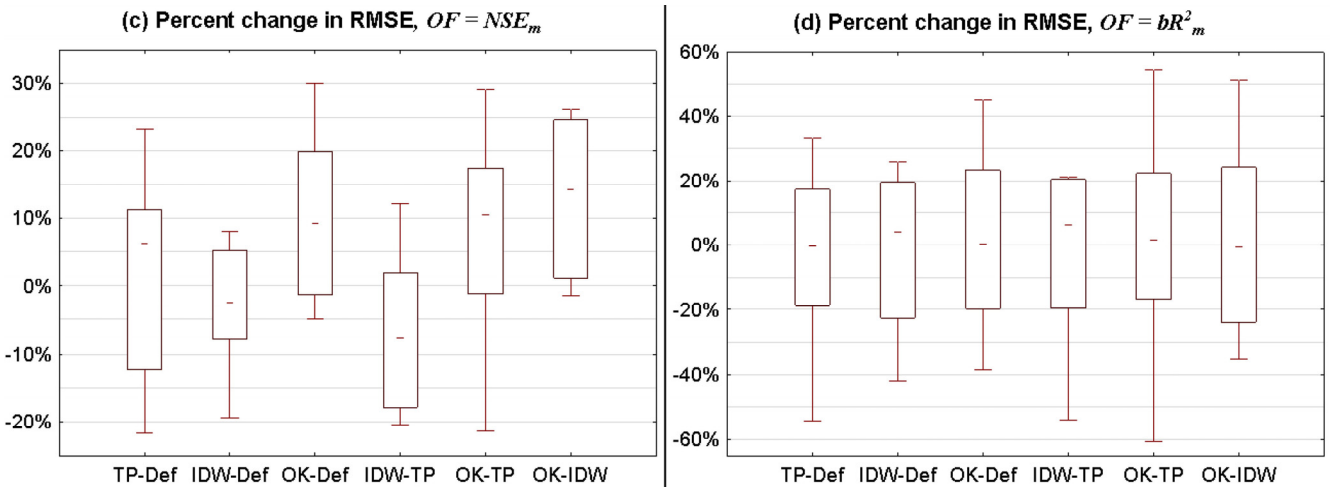

Figure 8. Box plots of percent changes in RMSE across all 11 flow gauging stations for different interpolation methods (Def—Default version, TP—Thiessen Polygons, IDW—Inverse Distance Weighted, OK—Ordinary Kriging) and different objective function/temporal aggregation combinations (**a**—NSE_d; **b**—bR_d^2; **c**—NSE_m; **d**—bR_m^2).

Another dimension of analysis is provided by studying the uncertainty coefficients p and r-factor (Figure 9). These two factors are strongly related to each other and should be interpreted jointly, *i.e.*, as a multi-objective problem. One can say that method A is better than method B only if

(1) *p*-factor for method A is higher than *p*-factor for method B and *r*-factor for method A is not higher than r-factor for method B.
(2) *p*-factor for method A is not lower than *p*-factor for method B and r-factor for method A is lower than *r*-factor for method B.

Mathematically, it can be said that a solution for method A is Pareto optimal. Following this rule, the OK method outperforms the IDW method for both daily and monthly aggregations, whereas it outperforms the TP method but only for monthly aggregation. For these three cases, the aforementioned rule holds true for the mean objective function values across all catchments as well as for most of the box plots characteristics shown in Figure 9. However, the *p*- and *r*-factors do not provide sufficient evidence to evaluate the three interpolation methods in relation to the Default method. It can be noted that *r*-factor is significantly lower for Def than for all other methods for each temporal aggregation, which shows that the width of 95PPU uncertainty band is lower in this case, but in most cases *p*-factor is also lower, which shows that a fewer percent of observed values fall into the uncertainty band. This phenomenon can be explained by lower precipitation for the Def method than for other methods, which leads to lower river flows and hence thinner uncertainty bands. Comparing the results between daily and monthly aggregation one can say that both p- and r-factors were slightly lower for the latter ones.

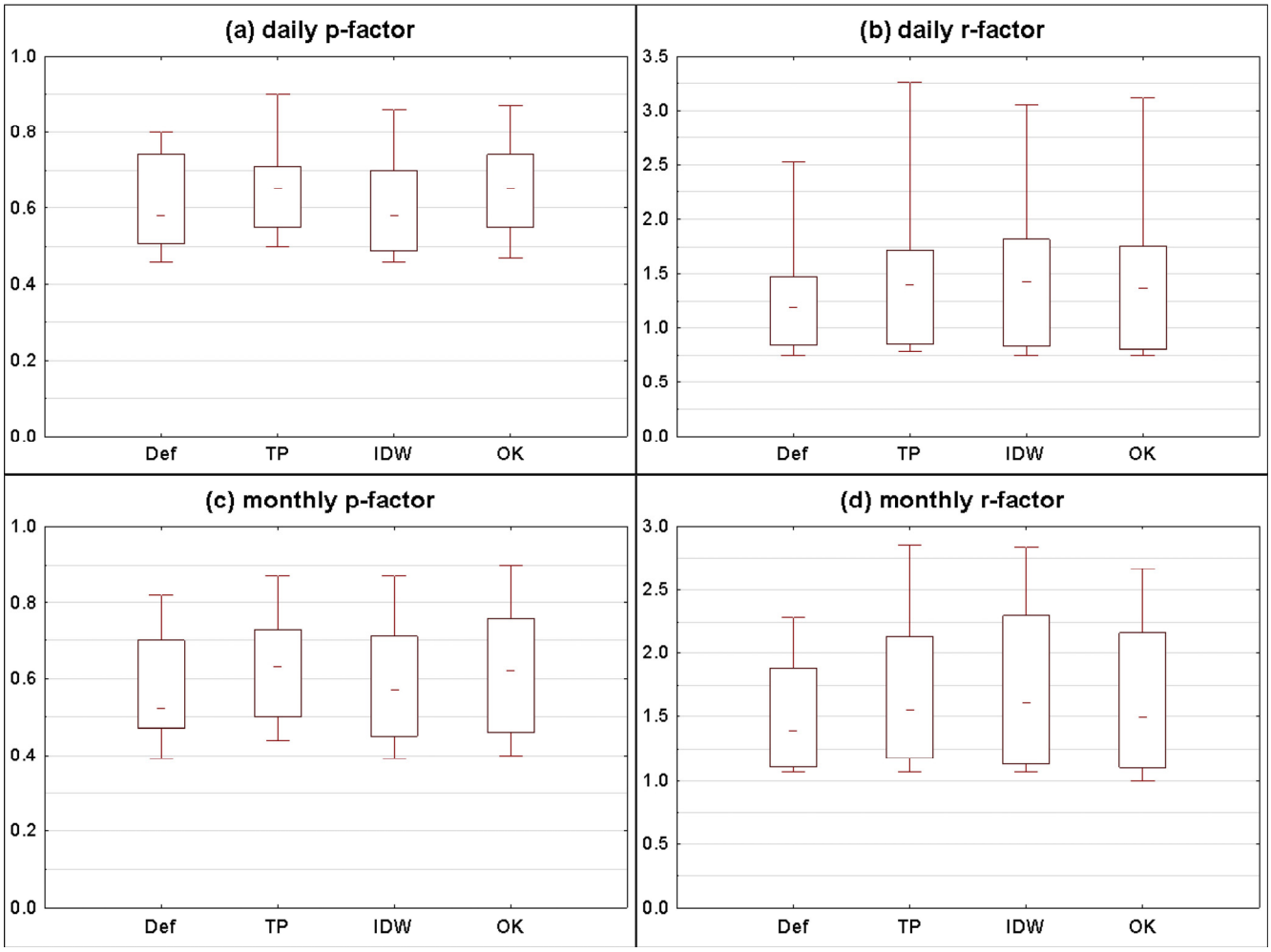

Figure 9. Box plots of uncertainty coefficients across all 11 flow gauging stations for different interpolation methods (Def—Default version, TP—Thiessen Polygons, IDW—Inverse Distance Weighted, OK—Ordinary Kriging) and different temporal aggregations (**a**,**b**: daily; **c**,**d**: monthly).

3.2.2. Relationship between the Objective Functions and Catchment Characteristics

The previous section outlined the differences between studied interpolation scenarios on the most general level. In this section, we analyse the results for individual catchments and try to explain the differences in results between catchments and between interpolation methods by examining relationships between $\Delta OF_{X,Y,t}$ (*cf.* Equation (9)) and a set of catchment descriptors. The following five descriptors were studied: upstream catchment area (A), mean area-specific runoff (q_{m}), coefficient of variation of daily/monthly flows (c_{v}) (*cf.* Table 2), mean kernel density (KD) as a proxy of precipitation station density (*cf.* Figure 3), and the difference in mean annual precipitation between different combinations of methods ($\Delta P_{X,Y}$, *cf.* Equation (10)). Table 3 presents Pearson correlation matrix for all these variables apart from q_m for which no significant correlation was found. $\Delta P_{X,Y}$ and KD were two catchment descriptors with the highest number of significant correlations with the output variables. KD was the most frequently negatively correlated with $\Delta bR^2_{X,Y,t}$. $\Delta P_{X,Y}$ had significant positive correlations with $\Delta bR^2_{X,Y,t}$, whereas for $\Delta NSE_{X,Y,t}$ one significant correlation was positive, while two were negative.

Table 3. Pearson correlation matrix between selected catchment properties (precipitation station density indicator KD (km^{-2}), upstream catchment area A, (km^2), coefficient of variation of daily/monthly flows c_v (-) and the difference in mean annual precipitation between methods X and Y $\Delta P_{X,Y}$ (mm) and the differences in objective functions $\Delta OF_{X,Y,t}$.

Catchment Properties	$TP-Def$	$IDW-Def$	$OK-Def$	$IDW-TP$	$OK-TP$	$OK-IDW$
			$\Delta NSE_{X,Y,d}$			
KD	−0.15	−0.36	−0.14	−0.36	0.00	0.28
A	−0.25	−0.49	0.14	−0.44	0.60 ‡	0.73 ‡
c_v	0.19	0.28	−0.10	0.18	−0.46	−0.44
$\Delta P_{X,Y}$	0.03	0.10	−0.04	0.14	0.07	−0.88 ‡
			$\Delta bR^2_{X,Y,d}$			
KD	−0.54 †	−0.56 †	−0.60 †	−0.07	−0.32	0.17
A	−0.13	−0.13	−0.16	−0.09	−0.22	0.08
c_v	−0.58 †	−0.60 ‡	−0.59 †	0.14	−0.05	0.20
$\Delta P_{X,Y}$	0.81 ‡	0.78 ‡	0.79 ‡	0.19	0.51	0.24
			$\Delta NSE_{X,Y,m}$			
KD	0.20	−0.29	−0.43	−0.69 ‡	−0.50	−0.10
A	0.27	0.03	0.35	−0.40	−0.04	0.34
c_v	0.51	0.49	−0.27	−0.23	−0.71 ‡	−0.88 ‡
$\Delta P_{X,Y}$	−0.60 †	−0.25	0.32	0.34	0.74 ‡	−0.37
			$\Delta bR^2_{X,Y,m}$			
KD	−0.54 †	−0.59 †	−0.70 ‡	−0.52 †	−0.80 ‡	−0.56 †
A	−0.27	−0.29	−0.26	−0.20	−0.13	−0.03
c_v	−0.25	−0.16	−0.33	0.26	−0.38	−0.55 †
$\Delta P_{X,Y}$	0.63 ‡	0.53 †	0.70 ‡	0.32	0.64 ‡	0.46

Notes: † Significant at significance level $p = 0.1$. ‡ Significant at significance level $p = 0.05$.

We have selected all cases from Table 3 with significant correlations (at significance level $p = 0.05$) and showed the respective relationships as scatter plots in Figure 10. First of all, it should be noted that in most plots the values of $\Delta OF_{X,Y,t}$ are positive for the majority of points, which shows that method X has higher objective function value than method Y. More specifically, the outcomes should be discussed in four groups related to different objective functions and temporal aggregations:

(1) $\Delta NSE_{X,Y,d}$ (Figure 10A,B): OK is superior over IDW and TP in catchments with larger drainage areas; OK is superior over IDW in catchments with small mean precipitation difference between these two methods.

(2) $\Delta bR^2_{X,Y,d}$ (Figure 10C,D): IDW is superior over Def in catchments with lower daily c_v (more stable flow regime); TP, IDW and OK are superior over Def in catchments with high positive difference in mean precipitation.

(3) $\Delta NSE_{X,Y,m}$ (Figure 10E–G): TP is superior over IDW in catchments with higher station densities; OK is superior over IDW and TP in catchments with lower monthly c_v (more stable flow regime); OK is superior over TP in catchments for which the difference in mean precipitation between OK and TP is positive.

(4) $\Delta bR^2_{X,Y,m}$ (Figure 10H,I): OK is superior over Def (more apparently) and TP (less apparently) in catchments with low station density. TP and OK are superior over Def in catchments with high positive difference in mean precipitation.

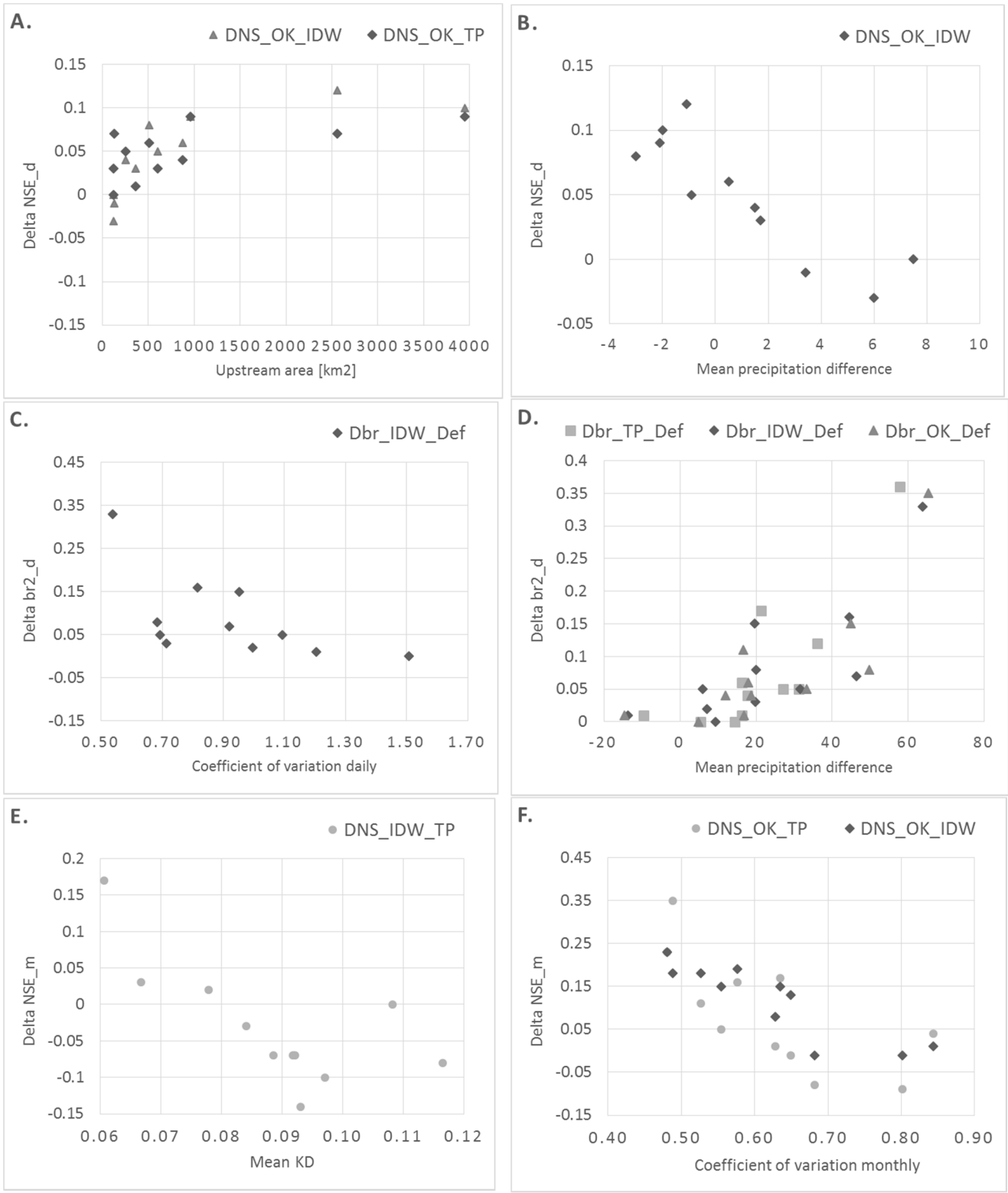

Figure 10. *Cont.*

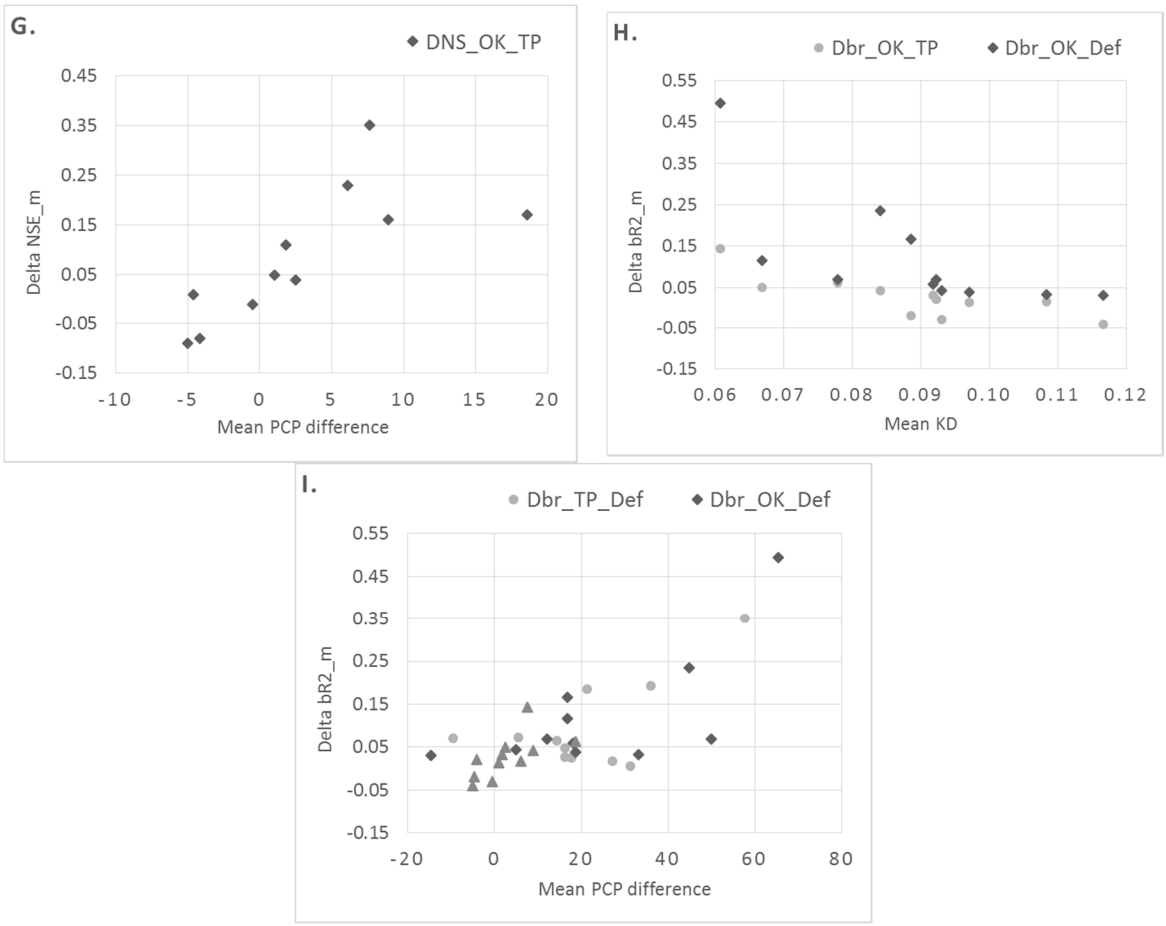

Figure 10. Scatter plots of $\Delta OF_{X,Y,t}$ and various catchment descriptors for relationships with significant correlation (at significance level $p = 0.05$) from Table 3. Note: DNS_X_Y = $\Delta NS_{X,Y}$, Dbr_X_Y = $\Delta bR^2_{X,Y}$.

3.3. Discussion

Table 4 shows basic information and data extracted from peer-reviewed papers on the impact of precipitation interpolation methods on flow simulation identified through a literature review. For a better context, our study was placed in the first row. In this summary, we present only those studies that fulfil the following criteria: (1) apply at least two different precipitation spatial interpolation schemes to daily precipitation for a period of at least three years; (2) force semi-distributed or distributed continuous time hydrological models with different input alternatives and test their efficiency in flow simulation. We have excluded studies using event-scale models running usually at sub-daily time step for prediction of individual flood event hydrographs (e.g., [55]). Some of the studies included in Table 4 presented also results for the lumped model version in parallel to a semi-distributed version [25], others reported results for output variables other than only flow (e.g., sediment, nutrient loadings [27]) or used other precipitation products apart from interpolation (e.g., GCM—General Circulation Model and reanalysis products: [22]). In all these cases, we have neglected this additional information that was not relevant for comparison with our study. The results from Hwang *et al.* [26] were divided into two separate records as two contrasting catchments were modelled separately in this study. The results from Wagner *et al.* [28] were however kept in one record because one model setup covered two neighbouring catchments.

Table 4. Literature review of studies evaluating precipitation interpolation methods for hydrological modelling.

ID	Publication Code and Material	Catchment (Country)	Area (km^2)	Number of Precipitation Gauges	Station Density (Stations/1000 km^2)	Model Name	Simulation Period (Years)	Number of Flow Gauges	Analysis Time Step	Interpolation Methods	Evaluation Criterion	Main Conclusion
1	This paper; Figures 6 and 8	Pilica (PL)	4,928	46	9.3	SWAT	30	11	d, m	Def (NN), TP,IDW, OK	NSE, bR^2	OK, IDW, TP outperformed Def for bR^2 for both daily and monthly time step; OK slightly better than others for NSE (daily and monthly)
2	Haberlandt1998 [22]; Figure 7A	Mackenzie (CA)	1,800,000	81	0.05	SLURP	16	29	m	NN, OK	Relative standard error	OK superior over NN but mainly in smaller subbasins (below 50,000 km^2)
3	Hwang2012 [26]; Table 7, Figures 17 and 18	Animas (CO, USA)	1,792	37	20.6	PRMS (distributed)	26	1	d, s, a	IDW, MLR, CMLR, LWP	RMSE, NSE, Flow statistics	All methods similar in terms of NSE and RMSE; all methods provide accurate timing of flood events but the magnitude is underestimated
4	Hwang2012 [26]; Tables 6 and 7, Figures 17 and 18	Alapaha (GA, USA)	3,626	28	7.7	PRMS (distributed)	22	1	d, s, a	IDW, MLR, CMLR, LWP	RMSE, NSE, Flow statistics	LWP and MLR superior over CMLR in terms of NSE and RMSE; all methods provide accurate timing of flood events but the magnitude is underestimated
5	Masih2011 [23] Table 3, Figures 5–7	Karkheh (IR)	4,2620	41	0.96	SWAT	15	15	d, m	Def (NN), IDEW	R^2, NSE, Flow statistics	Little difference between two methods for R^2, but IDEW superior over Def for NSE, especially in smaller subbasins (below 2500 km^2)

Table 4. *Cont.*

ID	Publication Code and Material	Catchment (Country)	Area (km^2)	Number of Precipitation Gauges	Station Density (Stations/1000 km^2)	Model Name	Simulation Period (Years)	Number of Flow Gauges	Analysis Time Step	Interpolation Methods	Evaluation Criterion	Main Conclusion
6	Ruelland2008 [25]; Table 5, Figure 10	Bani (ML, CI, BF)	100,000	13	0.13	Hydrostrahler	6	7	10d	TP, IDW, Spline, OK	NSE, VE, PE	The best results in terms of selected criteria were obtained for IDW, intermediate for TP and OK and the worst for Spline; all methods underestimated flood peaks
7	Shen2013 [27]; Tables 2 and 3, Figure 3a,b	Daning (CN)	4,426	19	4.3	SWAT	7	3	m	Def (NN), TP, IDW, Dis-Kriging, CoKriging	NSE, flow statistics	All methods showed an improvement over the Default method in terms of NSE (the highest for CoKriging); all methods underestimate most of flow characteristics
8	Wagner2012 [28]; Tables 4 and 5, Figure 8	Mula and Mutha (IN)	2,036	16	7.9	SWAT	21	4	d	$RIDW_x$, $RIDW_{trmm}$, RK_x, RK_{trmm}	NSE, PBIAS, flow statistics	RIDWTrmm and RKTrmm outperform RIDWX and RKX in terms of NSE and PBIAS; RKX overestimates runoff and does not reproduce right timing of floods in contrast to RKTrmm

Notes: Country codes: PL—Poland, CA—Canada, CO—Colorado, GA—Georgia, IR—Iran, ML—Mali, CI—Cote d'Ivoire, BF—Burkina Faso, CN—China, IN—India). Analysis time step: d—daily, m—monthly, s—seasonal, a—annual. Objective functions: NSE—Nash-Sutcliffe Efficiency, R^2—coefficient of determination, RMSE—root mean squared error; VE—volume error, PE—relative peak error; PBIAS—percentage bias. Methods: Def—Default, NN—Nearest Neighbour, TP—Thiessen Polygons, IDW—Inverse Distance Weighted, OK—Ordinary Kriging, MLR—Multiple Linear Regression, CMLR—Climatological Multiple Linear Regression LWP—Locally Weighted Polynomial Regression, IDEW—Inverse Distance and Elevation Weighting, $RIDW_x$, $RIDW_{trmm}$—Regression Inverse Distance, RK_x, RK_{trmm}—Regression Kriging methods.

All studies reported in Table 4 apart from ours were located in Asia (three cases), North America (three cases) or Africa (one case). Out of eight cases, five have drainage areas in the range 1792–4928 km^2, whereas three others have drainage areas in the range 42,620–1,800,000 km^2. The first group, containing our study, can thus be classified as meso-scale applications, and the second one as macro-scale applications. We used these values as well as the numbers of precipitation stations used for interpolation in order to estimate the values of mean station density indicator for each case: these varied considerably between studies, from 0.05 stations per 1000 km^2 [22] to 20.6 stations per km^2 [26]. In this respect, our study (9.3 stations per km^2) belongs to the group of higher station densities together with [26–28]. SWAT was the most frequently used hydrological model (four cases), whereas other models were: semi-distributed SLURP (Semi-distributed Land Use-based Runoff Processes), Hydrostrahler, and fully-distributed PRMS (Precipitation Runoff Modeling System). The length of simulation period was the longest in our study (30 years) and exceeded almost by a factor of two the mean value across all other studies. Since calibrated parameter values are very sensitive to climatic conditions and those calibrated for dry and short periods might not be suitable for simulating the opposite conditions [34,56], we conclude that study designs for this type of assessments should contain longer simulation periods (20–30 years) rather than only several years. The number of flow gauging stations varied between 1 and 29. Our study used 11 stations, however when referring the number of stations to catchment unit area, our study had the largest station density. Temporal aggregations were diversified, however daily and monthly aggregation used in our study were the most frequent. The range of applied interpolation schemes was also wide and included also regression-based methods [26,28] and more sophisticated than OK geostatistical methods (Dis-kriging, Cokriging; [27]). Only three studies including ours have applied the model default (usually NN) method that served as a reference for other more complex methods. The assessment criteria for method evaluation were non-uniform as well. However, all studies apart from Haberlandt *et al.* [22] used NSE as one of objective functions. In five cases, various flow statistics (e.g., mean flow and extreme flows) were reported and compared to measured values. None of the studies apart from ours used bR^2.

Looking at the main conclusions from these studies, one should notice that only in one case (Animas catchment in Colorado, USA [26]) only relatively minor differences between discharge simulated using various input alternatives was observed. In all other cases, there existed methods that were assessed as superior over other methods. Shen *et al.* [27] and Masih *et al.* [23] reported that all investigated interpolation methods were superior (in some sense) over the SWAT default method, which is in agreement with our results. Masih *et al.* [23] and Haberlandt *et al.* [22] showed that this improvement can be observed mainly in smaller catchments (*i.e.*, below 2500 km^2 and 50,000 km^2, respectively. Figure 10 A for daily NSE shows a relationship in the opposite direction: OK outperforms TP and IDW mainly in catchments with larger drainage areas. This discrepancy can be explained by a considerable scale and station density differences between our study and aforementioned studies (*cf.* Table 4), *i.e.*, the relationships found by Haberlandt *et al.* [22] and Masih *et al.* [23] are not necessarily valid for smaller spatial scales and/or for catchments with significantly higher station densities.

None of investigated papers did, however, make an attempt to study the effect of selected catchment properties on the evaluation of methods (apart from the catchment area discussed above). Negative correlation between station density indicator KD and some of the $\Delta OF_{X,Y,t}$ (Table 3, Figure 10E,H) imply that in catchments characterised by higher station densities spatial interpolation of precipitation does not bring much benefit. A similar conclusion was reported in the study of Masih *et al.* [23] in Iran, who

showed that mainly in smaller catchments with low station densities an the IDEW method outperformed the NN method. An interesting observation comes from the comparison of results for two catchments: Krz-BON (Krztynia upstream of Bonowice) and Pil-SZC (Pilica upstream of Szczekociny). They are neighbouring catchments, with quite similar area, physiographic properties and flow regime (*cf.* Table 2 and Figure 2) but differ largely in terms of station density (*cf.* Figure 3). We therefore assume that the fact that Pil-SZC has two precipitation stations inside its borders and Krz-BON has none explains the huge difference in results between these catchments.

The effect of c_{v} on $\Delta OF_{X,Y,t}$ is perhaps less intuitive and requires more in-depth analysis. Table 3 and Figure 10C,F show that catchments with lower daily/monthly flow variability are more prone to improvements brought by interpolation methods than catchments with higher variability. This is particularly visible for evaluation with monthly NSE (Figure 10F), for which the correlation between c_{v} and $\Delta NSE_{OK,IDW,m}$ is very high (−0.88). The improvement of using OK over using IDW (and TP) is thus particularly high for catchments with more stable flow regime. The ruling mechanism lying behind is not so clear, though. More stable flow regime (low c_{v}) usually means higher baseflow and smaller frequency and magnitude of flood events. Since applying interpolation schemes is assumed to affect mainly the flood events and since the objective functions used are very sensitive to highest flows, perhaps an improvement in simulation of flood events brought by interpolation schemes is more likely to have a measurable effect in catchments with more stable flow than in catchments with highly variable flow.

The highest number of significant correlations between catchment descriptors and $\Delta OF_{X,Y,t}$ was found for $\Delta P_{X,Y}$ (Table 3, Figure 10 B,D,G,I). Significant correlations were found for all four possible combinations of objective functions and temporal aggregations. With one exception (Figure 10B for $\Delta P_{OK,IDW}$) all correlations were positive and the values of $\Delta P_{X,Y}$ were also predominantly positive. It should also be noted that there is a significant negative correlation between KD and $\Delta P_{X,Y}$ (*cf.* Section 3.1). This is in agreement with the results of Hwang *et al.* [26], who reported that the differences in estimated precipitation were significantly larger in the Alapaha basin than in the Animas basin, while station density was 2.7 times higher in the latter basin (Table 4). In summary, our results show that in many cases, when precipitation input derived by method X is higher than that derived by method Y, it leads to a higher model efficiency for method X than for method Y. For any catchment, true areal precipitation is unknown and each interpolation method provides only its approximation. However the results may suggest that better model performance is associated with more correct approximation, while worse results with underestimated precipitation.

One of the problems associated with this type of input data evaluation is compensation for possibly wrong input by model calibration. This problem was described in more detail and exemplified with a set of meaningful cases by Heistermann and Kneis [57]. Wagner *et al.* [28] partly overcome this issue by not calibrating the model to observed rainfall-runoff events but using parameter values based on regional knowledge, literature review or just set to default. Most of the authors listed in Table 4 [23,25,26] did perform formal model calibration for each of evaluated interpolation schemes, so their results might be biased by the "compensation" issue. In contrast, Haberlandt *et al.* [22] and Shen *et al.* [27] performed calibration for only one method and then applied the identified parameter set for other methods. Such an approach carries the risk of advantaging the method for which calibration was performed, though. Hence, the approach used in this paper lies somewhere in between: we did apply semi-automated calibration program, SUFI-2, but used only one iteration per interpolation method and kept the same parameter

ranges throughout all iterations. It probably does not eliminate the parameter compensation problem totally, but we believe that it is partly eliminated due to the fact that further SUFI-2 iterations with adjusted parameter ranges would have inevitably led to more pronounced parameter compensation issues.

Moriasi *et al.* [52] suggested the thresholds for the goodness-of-fit criteria for hydrological models, indicating that the values of PBIAS higher than 25% in terms of absolute values yield the model unsatisfactory. In our study, PBIAS was not at all used as an objective function, but it was calculated for all parameter sets providing the best fit with observed values, for all combinations of objective functions and temporal aggregations (Figure 7). PBIAS was highly dependent on the type of objective function and temporal aggregation, e.g., for NSE_d (Figure 7C) all values shown in the box plots were lower than 25% in terms of the absolute values. In contrast, for bR_d^2 (Figure 7B) for all interpolation methods majority of catchments had unsatisfactory model performance in terms of PBIAS according to criteria of Moriasi *et al.* [52]. However, as emphasised by Moriasi *et al.* [52], the criteria should always be adjusted to the project goals, and in our case model calibration and application was not the goal, but it was testing of different input alternatives. Furthermore, other studies with a similar scope (e.g., Masih *et al.* [23], Wagner *et al.* [28]) also reported very high PBIAS values for certain catchments.

We have also evaluated the best fit parameter sets using RMSE, in a similar manner as for PBIAS. For the objective functions NSE_d, NSE_m and bR_d^2, the results for RMSE (Figure 8A–C) supported the results showed in Figure 6A–C. In contrast, for bR_m^2, even though Figure 6D showed that IDW, TP and OK outperformed Def, the relative changes in RMSE shown in Figure 8D do not confirm this. The reason is that there were a few catchments for which a method X outperformed a method Y in terms of bR_m^2, but at the same time method Y outperformed method X in terms of RMSE and PBIAS. This is a well-known, inherent problem of multi-objective calibrations: moving from one solution to another results in the improvement of one objective function (in this case bR_m^2) while causing a deterioration in the value of at least one other objective function (in this case RMSE or PBIAS) [58].

4. Conclusions

In this study, the effect of daily precipitation interpolation methods on the semi-distributed SWAT model efficiency of daily and monthly discharge simulation over a 30-year period was examined in 11 meso-scale catchments in central Poland. The results showed that the most complex OK method outperformed other methods in terms of NSE, whereas OK, IDW and TP outperformed Def in terms of bR^2, regardless of temporal aggregation. The difference between these three interpolation methods and Def was, however, spatially variable and a part of this variability was attributed to catchment properties: catchments characterised by low station density, low coefficient of variation of daily/monthly flows and a higher interpolated precipitation estimation experienced more pronounced improvement as a result of using interpolation methods. The implication of this study is that appropriate consideration of spatial precipitation variability (often neglected by model users) that can be achieved applying interpolation methods can significantly improve the reliability of model simulations across different scales. Ordinary Kriging should be considered as the optimal method, however we recommend testing various methods, as the results tend to be catchment-specific. From the practical point of view, this study identified certain circumstances (sparse precipitation gauge networks, catchments with stable flow regimes, higher

precipitation estimation) under which one can expect a larger improvement in model efficiency criteria by applying precipitation interpolation schemes.

Acknowledgments

We would like to acknowledge the Institute of Meteorology and Water Management—National Research Institute (IMGW-PIB) for providing hydrometeorological data. This study was a side activity within the EKOROB LIFE08 ENV/PL/000519 project coordinated by Katarzyna Izydorczyk from ERCE PAN (European Regional Centre for Ecohydrology of the Polish Academy of Sciences) and Wojciech Frątczak from ECRE PAN and RZGW Warszawa who helped in providing certain model inputs. Our colleagues from SGGW: Ignacy Kardel, Marek Giełczewski and Paweł Marcinkowski contributed to development of model setup. The second author benefitted from the START 2014 stipend from the Foundation for Polish Science and from the Humboldt Research Fellowship for Postdoctoral Researchers from the Alexander von Humboldt Foundation during the manuscript preparation phase.

Author Contributions

Contribution of Mateusz Szcześniak and M.P. to this manuscript is assessed as equal. Mikołaj Piniewski developed the methodological framework. Mikołaj Piniewski and Mateusz Szcześniak contributed to development of model setup. Mateusz Szcześniak performed all interpolation calculations. Both Mateusz Szcześniak and Mikołaj Piniewski contributed to calculations in SWAT-CUP. Mikołaj Piniewski performed statistical analyses. Mateusz Szcześniak prepared an early draft of the manuscript based on his M.Sc. thesis. Mikołaj Piniewski extended and prepared the final version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Arnold, J.G.; Srinivasan, R.; Muttiah, R.S.; Williams, J.R. Large area hydrologic modeling and assessment—Part 1: Model development. *J. Am. Water Resour. Assoc.* **1998**, *34*, 73–89.
2. Gassman, P.W. Worldwide use of SWAT: 2012 update. In Proceedings of the 2012 International SWAT Conference, India Habitat Centre, Lodhi Road, New Delhi, India, 18–20 July 2012.
3. Smith, M.B.; Korena, V.I.; Zhanga, Z.; Reeda, S.M.; Panb, J.J.; Moreda, F. Runoff response to spatial variability in precipitation: an analysis of observed data. *J. Hydrol.* **2004**, *298*, 267–286.
4. Tetzlaff, D.; Uhlenbrook, S. Significance of spatial variability in precipitation for process-oriented modelling: Results from two nested catchments using radar and ground station data. *Hydrol. Earth Syst. Sci.* **2005**, *9*, 29–41.
5. Price, K.; Purucker, S.T.; Kraemer, S.R.; Babendreier, J.E.; Knightes, C.D. Comparison of radar and gauge precipitation data in watershed models across varying spatial and temporal scales. *Hydrol. Process.* **2014**, *28*, 3505–3520.
6. Caracciolo, D.; Arnone, E.; Noto, L. Influence of Spatial Precipitation Sampling on Hydrological Response at the Catchment Scale. *J. Hydrol. Eng.* **2014**, *19*, 544–553.

7. Chintalapudi, S.; Sharif, H.O.; Xie, H. Sensitivity of Distributed Hydrologic Simulations to Ground and Satellite Based Rainfall Products. *Water* **2014**, *6*, 1221–1245.
8. Di Piazza, A.; Lo Conti, F.; Noto, L.V.; la Loggia, G. Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 396–408.
9. Yang, D.; Goodison, B.E.; Metcalfe, J.R.; Louie, P.; Leavesley, G.; Emerson, D.; Hanson, C.L.; Golubev, V.S.; Elomaa, E.; Gunther, T.; *et al.* Quantification of precipitation measurement discontinuity induced by wind shields on national gauges. *Water Resour. Res.* **1999**, *35*, 491–508.
10. Mishra, A.K. Effect of rain gauge density over the accuracy of rainfall: A case study over Bangalore, India. *SpringerPlus* **2013**, *2*, 1–7.
11. Majewski, W. Urban flash flood in Gdańsk, 2001. *Ann. Wars. Univ. Life Sci. SGGW Land Reclam.* **2008**, *39*, 129–137.
12. Weedon, G.P.; Gomes, S.; Viterbo, P.; Shuttleworth, W.J.; Blyth, E.; Österle, H.; Adam, J.C.; Bellouin, N.; Boucher, O.; Best, M. Creation of the WATCH Forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *J. Hydrometerol.* **2011**, *12*, 823–848.
13. Kalin, L.; Hantush, M. Hydrologic Modeling of an Eastern Pennsylvania Watershed with NEXRAD and Rain Gauge Data. *J. Hydrol. Eng.* **2006**, *11*, 555–569.
14. Sun, X.; Mein, R.G.; Keenan, T.D.; Elliott, J.F. Flood estimation using radar and raingauge data. *J. Hydrol.* **2000**, *239*, 4–18.
15. Sharma, S.; Isik, S.; Srivastava, P.; Kalin, L. Deriving spatially distributed precipitation data using the artificial neural network and multilinear regression models. *J. Hydrol. Eng.* **2012**, *18*, 194–205.
16. Morin, E.; Krajewski, W.F.; Goodrich, D.C.; Gao, X.; Sorooshian, S. Estimating Rainfall Intensities from Weather Radar Data: The Scale-Dependency Problem. *J. Hydrometeor.* **2003**, *4*, 782–797.
17. Ly, S.; Charles, C.; Degré, A. Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale. A review. *Biotechnol. Agron. Soc. Environ.* **2013**, *17*, 392–406.
18. Hartkamp, A.D.; de Beurs, K.; Stein, A.; White, J.W. *Interpolation Techniques for Climate Variables*; CIMMYT: Mexico D.F., Mexico, 1999.
19. Goovaerts, P. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.* **2000**, *228*, 113–129.
20. Zhang, X.; Srinivasan, R. GIS-based spatial precipitation estimation: A comparison of geostatistical approaches. *J. Am. Water Resour. Assoc.* **2009**, *45*, 894–906.
21. Tabios, G.Q.; Salas, J.D. A comparative analysis of techniques for spatial interpolation of precipitation. *Water Resour. Bull.* **1985**, *21*, 365–380.
22. Haberlandt, U.; Kite, G.W. Estimation of daily space-time precipitation series for macroscale hydrological modelling. *Hydrol. Process.* **1998**, *12*, 1419–1432.
23. Masih, I.; Maskey, S.; Uhlenbrook, S.; Smakhtin, V. Assessing the Impact of Areal Precipitation Input on Streamflow Simulations Using the SWAT Model. *J. Am. Water Resour. Assoc.* **2011**, *47*, 179–195.
24. Blöschl, G. Scale and scaling in hydrology. In *Wiener Mitteilungen, Wasser–Abwasser–Gewässer*; Technical University of Vienna: Vienna, Austria, 1996.

25. Ruelland, D.; Ardoin-Bardin, S.; Billen, G.; Servat, E. Sensitivity of a lumped and semi-distributed hydrological model to several methods of rainfall interpolation on a large basin in West Africa. *J. Hydrol.* **2008**, *361*, 96–117.
26. Hwang, Y.; Clark, M.; Rajagopalan, B.; Leavesley, G. Spatial interpolation schemes of daily precipitation for hydrologic modeling. *Stoch. Environ. Res. Risk Assess.* **2012**, *26*, 295–320.
27. Shen, Z.; Chen, L.; Liao, Q.; Liu, R.; Hong, Q. Impact of spatial rainfall variability on hydrology and nonpoint source pollution modeling. *J. Hydrol.* **2012**, *472*, 205–215.
28. Wagner, P.D.; Fiener, P.; Wilken, F.; Kumar, S.; Schneider, K. Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. *J. Hydrol.* **2012**, *465*, 388–400.
29. Boer, E.P.J.; de Beurs, K.M.; Hartkamp, A.D. Kriging and thin plate splines for mapping climate Variables. *Int. J. Appl. Earth Obs. Geoinf.* **2001**, *3*, 146–154.
30. Todini, E.; Pellegrini, F.; Mazzetti, C. Influence of parameter estimation uncertainty in Kriging: Part 2. Test and case study applications. *Hydrol. Earth Syst. Sci.* **2001**, *5*, 225–232.
31. Marqunez, J.; Lastra, J.; Garcia, P. Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis. *J. Hydrol.* **2003**, *270*, 1–11.
32. Vicent-Serrano, S.M.; Saz-Sanchez, M.A.; Cuadrat, J.M. Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): Application to annual precipitation and temperature. *Clim. Res.* **2003**, *24*, 161–180.
33. Lloyd, C.D. Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. *J. Hydrol.* **2005**, *308*, 128–150.
34. Vaze, J.; Post, D.A.; Chiew, F.H.S.; Perraud, J.M.; Viney, N.R.; Teng, J. Climate non-stationarity—Validity of calibrated rainfall–runoff models for use in climate change studies. *J. Hydrol.* **2010**, *394*, 447–457.
35. Hattermann, F.F.; Wattenbach, M.; Krysanova, V.; Wechsung, F. Runoff simulations on the macroscale with the ecohydrological model SWIM in the Elbe catchment–validation and uncertainty analysis. *Hydrol. Process.* **2005**, *19*, 693–714.
36. Wagner, I.; Zalewski, M. Effect of hydrological patterns of tributaries on biotic processes in a lowland reservoir—Consequences for restoration. *Ecol. Eng.* **2000**, *16*, 79–90.
37. Neitsch, S.; Arnold, J.; Kiniry, J.; Williams, J. *Soil and Water Assessment Tool Theoretical Documentation Version 2009*; Texas Water Resources Institute Technical Report No. 406; Texas A & M University: Temple, TX, USA, 2011.
38. Arnold, J.; Kiniry, J.; Srinivasan, R.; Williams, J.; Haney, E.; Neitsch, S. *Soil and Water Assessment Tool Input/Output File Documentation Version 2009*; Texas Water Resources Institute Technical Report No. 365; Texas A & M University: Temple, TX, USA, 2011.
39. Winchell, M.; Srinivasan, R.; Luzio, M.D.; Arnold, J. *ArcSWAT Interface for SWAT2009. User's Guide*; Technical Report; Blackland Research and Extension Center; Grassland, Soil and Water Research Laboratory, Texas A & M University: Temple, TX, USA, 2010.
40. Sharpley, A.N.; Williams, J.R. *Erosion/Productivity Impact Calculator, 1. Model Documentation*; USDA-ARS Technical Bulletin Vol. 1768; Department of Agriculture: Washington, DC, USA; 1990; pp. 3–9.
41. Thiessen, A.H. Precipitation averages for large areas. *Monthly Weather Rev*. **1911**, *39*, 1082–1084.

42. Meijerink, A.M.J.; de Brouwer, H.A.M.; Mannaerts, C.M.; Valenzuela, C.R. *Introduction to Use of Geographic Information Systems for Practical Hydrology*; International Institute for Aerospace Survey and Earth Sciences: Enschede, The Netherlands, 1994.
43. Isaaks, E.H.; Srivastava, R.M. *Applied Geostatistics*; Oxford University Press: New York, NY, USA, 1998.
44. Ly, S.; Charles, C.; Degre, A. Geostatistical interpolation of daily rainfall at catchment scale: The use of several variogram models in the Ourthe and Ambleve catchments, Belgium. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 2259–2274.
45. Larson, L.L.; Peck, E.L. Accuracy of precipitation measurements for hydrologic modeling. *Water Resour. Res.* **1974**, *10*, 857–863.
46. ArcGIS Desktop 10 Help. Available online: http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//0031000000m3000000 (accessed on 9 February 2015).
47. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Chapman and Hall: New York, NY, USA, 1986.
48. Piniewski, M.; Kardel, I.; Giełczewski, M.; Marcinkowski, P.; Okruszko, T. Climate change and agricultural development: Adapting Polish agriculture to reduce future nutrient loads in a coastal watershed. *Ambio* **2014**, *43*, 644–660.
49. Abbaspour, K. *SWAT-CUP2: SWAT Calibration and Uncertainty Programs—A User Manual*; Department of Systems Analysis, Integrated Assessment and Modelling(SIAM), Eawag, Swiss Federal Institute of Aquatic Science and Technology: Duebendorf, Switzerland, 2009.
50. Abbaspour, K.; Johnson, C.A.; van Genuchten, M.T. Estimating Uncertain Flow and Transport Parameters Using a Sequential Uncertainty Fitting Procedure. *Vadose Zone J.* **2004**, *3*, 1340–1352.
51. Piniewski, M.; Okruszko, T. Multi-site calibration and validation of the hydrological component of SWAT in a large lowland catchment. In *Modelling of Hydrological Processes in the Narew Catchment*; Świątek, D., Okruszko, T., Eds.; Springer: Berlin, Germany, 2011; pp. 15–41.
52. Moriasi, D.N.; Arnold, J.G.; van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Tran. ASABE* **2007**, *50*, 885–900.
53. Gupta, H.V.; Sorooshian, S.; Yapo, P.O. Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *J. Hydrol. Eng.* **1999**, *4*, 135–143.
54. Randles, R.H. Wilcoxon signed rank test. *Encycl. Stat. Sci.* **1988**, *9*, 613–616.
55. Tobin, C.; Nicotina, L.; Parlange, M.B.; Berne, A.; Rinaldo, A. Improved interpolation of meteorological forcings for hydrologic applications in a Swiss Alpine region. *J. Hydrol.* **2011**, *401*, 77–89.
56. Zhu, Q.; Zhang, X.; Ma, C.; Gao, X.; Xu, Y. Investigating the uncertainty and transferability of parameters in SWAT model under climate change. *Hydrol. Sci. J.* **2014**, in press.
57. Heistermann, M.; Kneis, D. Benchmarking quantitative precipitation estimation by conceptual rainfall-runoff modeling. *Water Resour. Res*. **2011**, *47*, doi:10.1029/2010WR009153.
58. Yapo, P.O.; Gupta, H.V.; Sorooshian, S. Multi-objective global optimization for hydrologic models. *J. Hydrol.* **1998**, *204*, 83–97.