Hydrology and Earth System Sciences

Open Access

Evaluation of areal precipitation estimates based on downscaled reanalysis and station data by hydrological modelling

D. Duethmann[1], J. Zimmer[1,*], A. Gafurov[1], A. Güntner[1], D. Kriegel[1], B. Merz[1], and S. Vorogushyn[1]

[1]GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany
[*]now at: Engineering Consultancy Load and Energy Management, Leipzig, Germany

Correspondence to: D. Duethmann (doris.duethmann@gfz-potsdam.de)

Abstract. In data sparse mountainous regions it is difficult to derive areal precipitation estimates. In addition, their evaluation by cross validation can be misleading if the precipitation gauges are not in representative locations in the catchment. This study aims at the evaluation of precipitation estimates in data sparse mountainous catchments. In particular, it is first tested whether monthly precipitation fields from downscaled reanalysis data can be used for interpolating gauge observations. Secondly, precipitation estimates from this and other methods are evaluated by comparing simulated and observed discharge, which has the advantage that the data are evaluated at the catchment scale. This approach is extended here in order to differentiate between errors in the overall bias and the temporal dynamics, and by taking into account different sources of uncertainties. The study area includes six headwater catchments of the Karadarya Basin in Central Asia. Generally the precipitation estimate based on monthly precipitation fields from downscaled reanalysis data showed an acceptable performance, comparable to another interpolation method using monthly precipitation fields from multi-linear regression against topographical variables. Poor performance was observed in only one catchment, probably due to mountain ridges not resolved in the model orography of the regional climate model. Using two performance criteria for the evaluation by hydrological modelling allowed a more informed differentiation between the precipitation data and showed that the precipitation data sets mostly differed in their overall bias, while the performance with respect to the temporal dynamics was similar. Our precipitation estimates in these catchments are considerably higher than those from continental- or global-scale gridded data sets. The study demonstrates large uncertainties in areal precipitation esti-

mates in these data sparse mountainous catchments. In such regions with only very few precipitation gauges but high spatial variability of precipitation, important information for evaluating precipitation estimates may be gained by hydrological modelling and a comparison to observed discharge.

1 Introduction

In data sparse mountain regions it is challenging to derive areal precipitation estimates. At the same time, evaluating different spatial interpolation approaches also is difficult, as cross validation may lead to wrong conclusions if large fractions of the catchment are underrepresented by precipitation gauges (Heistermann and Kneis, 2011). Large uncertainties in areal precipitation estimates are generally due to measurement errors and the scale difference between the point measurements and the areal estimate. This is amplified in mountainous regions, where, despite the high spatial variability of precipitation, the gauge network often has a low density with an unequal distribution towards lower and less exposed locations (Frei and Schär, 1998).

Orography affects the spatial pattern and the amount of precipitation through various processes (e.g. Houze, 1993, for an overview). Despite complex relations between orography and precipitation, in general, these processes often result in an increase of precipitation with elevation, particularly on windward slopes, and lower precipitation on the leeward side of a mountain range (rain shadow effect). For the spatial interpolation of precipitation in mountainous areas, methods which consider the orography are therefore often advantageous over methods neglecting the relation with the terrain

(Goovaerts, 2000; Hevesi et al., 1992; Martinez-Cob, 1996; Phillips et al., 1992; Tobin et al., 2011). Exceptions from this occur when the correlation between precipitation and elevation is low, or in regions where the station density is so high that the relation between precipitation and topography is already represented by the observations (Haberlandt et al., 2005; Ly et al., 2011). Elevation may be taken into account using geostatistical methods like modified residual kriging, external drift kriging or co-kriging with elevation (Garen et al., 1994; Goovaerts, 2000; Hevesi et al., 1992; Lloyd, 2005; Martinez-Cob, 1996; Phillips et al., 1992; Tobin et al., 2011), or using multi-linear or polynomial regression against various topographical variables (Basist et al., 1994; Brown and Comrie, 2002; Cheval et al., 2003; Daly et al., 1994; Goodale et al., 1998; Hay et al., 1998; Johansson and Chen, 2003, 2005; Marquinez et al., 2003; Ninyerola et al., 2000, 2007; Perry and Hollis, 2005; Prudhomme and Reed, 1998; Sun et al., 2008).

These statistical approaches require that the spatial variability of precipitation is captured by the observed precipitation, including, e.g. the relationship to topographic variables. In sparsely gauged areas with a more complex topography this may not be possible. In this case precipitation from reanalysis data downscaled by a regional climate model (RCM) could be a helpful source for deriving the spatial variability of precipitation within the catchment. Such data become increasingly available (van der Linden and Mitchell, 2009). As an RCM considers the interactions between the orography and the wind field for simulating precipitation, it should be able to represent orographic precipitation and rain shadowing effects in a suitable and physically based way. Only few studies started to work in this direction. Haberlandt and Kite (1998), for example, used daily precipitation output from the NCAR reanalysis (without downscaling) for the geostatistical interpolation of station-based precipitation time series, and recently Tobin et al. (2011) interpolated precipitation data from gauge observations by external drift kriging with precipitation fields from event accumulated COSMO7 reanalysis data as trend variable. As generally the performance of downscaled reanalysis data is lower on shorter time steps (Hurkmans et al., 2008), we propose to combine monthly accumulated spatial fields from downscaled reanalysis data with daily station data for the estimation of areal precipitation in data sparse regions. We compare this interpolation method with the direct use of downscaled reanalysis precipitation data, precipitation estimates based on multi-linear regression against topographical variables, data interpolated by inverse distance and with the gauge based daily gridded precipitation data set APHRODITE (Yatagai et al., 2012).

Traditionally, different precipitation data sets are evaluated based on measured values from precipitation gauges, e.g. by cross validation. This may however lead to wrong conclusions if the precipitation gauges are not in representative locations of the catchment. In such situations, the comparison of observed runoff with simulated runoff from a hydrological model driven by the different precipitation data sets can be a more suitable method for the evaluation of areal precipitation estimates (Heistermann and Kneis, 2011; Stisen and Sandholt, 2010). This has the advantage that the scale problem between point measurements and areal estimates is eliminated, as discharge measurements represent an integrated response from the entire catchment. Under average flow conditions, discharge measurements are also usually afflicted with smaller measurement errors than precipitation measurements, especially if these contain a large fraction of snow measurements. On the other hand, it has to be considered that this approach also introduces other uncertainties related to model uncertainties, errors in the catchment runoff from unknown subsurface inflow/outflow and unknown abstractions or flow diversions.

There are basically two approaches for the assessment of different precipitation estimates using hydrological modelling. The model may either be recalibrated for each precipitation data set (Yilmaz et al., 2005; Stisen and Sandholt, 2010; Bitew and Gebremichael, 2011a, b; Behrangi et al., 2011; Artan et al., 2007) or applied within a Monte Carlo framework (Heistermann and Kneis, 2011; Gourley and Vieux, 2005). The different precipitation data sets are then typically assessed using model performance measures for the simulated discharge. In addition, the bias of the precipitation data set may be evaluated using the bias in the simulated discharge. This approach, however, has some drawbacks. It does not allow for directly quantifying the bias of the precipitation estimate, which due to non-linearities of the system is usually different from the streamflow bias. Additionally, if a precipitation estimate has a large bias and one wants to evaluate its performance with respect to the temporal dynamics, it is not advisable to directly use it as input to a hydrological model, as the whole system may function in a different mode. Scaling all precipitation estimates to a reference precipitation data set allows evaluating the precipitation estimates independent of their biases (Stisen and Sandholt, 2010), but has the disadvantage that a reference data set needs to be identified, which may itself also be afflicted with an unknown bias. In this study we therefore extended this approach by adding a precipitation bias factor to the calibration parameters in order to evaluate the bias within the calibration framework. This also brings the advantage that uncertainties in the estimated bias can be assessed within the calibration framework.

The aim of this study is the evaluation of precipitation estimates in data sparse mountainous regions. It is first tested whether spatial precipitation fields from downscaled reanalysis data can be used to interpolate station observations. Second, the approach for comparing and evaluating areal precipitation estimates by hydrological modelling is further developed to separately consider the performance of different precipitation data sets with respect to their overall bias and

their temporal dynamics, and to account for different sources of uncertainties.

With respect to the case study region – the Karadarya catchment in Central Asia – estimating and assessing the precipitation input contributes to a better understanding of the hydrology in such a sparsely gauged region, and is a prerequisite for reliable hydrological modelling. The region strongly depends on water resources from mountain catchments for irrigation, hydropower generation and for water inflow to the Aral Sea (e.g. Siegfried and Bernauer, 2007). The question of possible climate change effects on water availability therefore is highly relevant in this area and there is a demand in setting up hydrological models for approaching this task.

2 Study area

The Karadarya catchment is a mountainous catchment in Kyrgyzstan, Central Asia. The confluence of the Karadarya and Naryn River in Uzbekistan forms the Syrdarya, the second largest tributary to the Aral Sea. The study area upstream of the Andijan Reservoir has an area of $13\,000\,\mathrm{km^2}$. The catchment is bordered by the Fergana Range in the northeast and by the Alay Range in the south, where elevations reach up to 4753 m (Fig. 1). Dominant land cover types are grasslands (59 %) and croplands (23 %), followed by smaller fractions of shrub land (5 %), woody vegetation (5 %), and glaciated areas (1 %). Mean annual precipitation, based on the 1961–1990 time series at the precipitation stations, ranges from 350 to $1050\,\mathrm{mm\,a^{-1}}$. The precipitation regime shows a maximum in spring and a second smaller maximum in autumn.

The focus of our study is on six headwater subcatchments, for which discharge data are available and which are assumed to be only marginally influenced by water management. The location of these six subcatchments is shown in Fig. 1 and important characteristics are listed in Table 1. For most of these subcatchments the mean annual runoff over the period 1961–1990 has values of $400\,\mathrm{mm\,a^{-1}}$ to $600\,\mathrm{mm\,a^{-1}}$, outliers are Ak-Tash with nearly $800\,\mathrm{mm\,a^{-1}}$ and Gulcha in the south with less than $300\,\mathrm{mm\,a^{-1}}$. The discharge regime is strongly seasonal with maximum discharges during the snowmelt season in spring and early summer. In accordance with increasing average elevation, maximum monthly discharges occur in April in Tosoi and Donguztoo, in May in Salamalik, and in June in Gulcha, Cholma and Ak-Tash.

Table 1. Area, glaciation, elevation range and mean annual runoff of the studied subcatchments of the Karadarya Basin.

| | Area | Glacier | Elevation (m) | | | Runoff |
	$(\mathrm{km^2})$	(%)	Min.	Max.	Mean	$(\mathrm{mm\,a^{-1}})$*
Tosoi	216	0.0	1253	3165	2001	432
Donguztoo	166	0.0	1271	3502	1999	505
Salamalik	1180	0.5	1288	4381	2592	585
Ak-Tash	907	2.3	1728	4752	3121	778
Cholma	3840	1.9	1352	4753	3117	410
Gulcha	2010	0.7	1557	4623	3013	267

* Mean annual runoff over the period 1961–1990.

3 Methods and data

3.1 Precipitation data and interpolation approaches

3.1.1 Downscaled reanalysis data

A relatively good performance of global reanalysis data in Central Asia was shown by Schär et al. (2004) and Schiemann et al. (2008). This was attributed to the fact that weather systems typically move into the region from the west and reanalysis data for Central Asia therefore benefit from the denser observation network in Europe and the Middle East, which partly compensates the sparse data coverage in the region. In order to resolve orographic precipitation and rain shadowing effects at smaller scales, it is necessary to downscale the reanalysis data by a regional climate model.

In this study, data from the ERA-40 reanalysis (Uppala et al., 2005) with a horizontal resolution of $1°$ are downscaled to a 12 km grid using the RCM Weather Research and Forecasting Model (WRF, Skamarock et al., 2008) for the time period 1959–1990. This study period was chosen due to the availability of precipitation data, which strongly declines after 1990. A two-way nesting approach is applied, with the first nest at a horizontal resolution of 36 km covering a region between 35 to $47°$ N and 62 to $83°$ E and the second nest at a resolution of 12 km covering an area between 38 to $45°$ N and 65 to $80°$ E. The model is run with daily restarts in order to keep it close to the ERA-40 boundary and initial conditions; the simulation time for each day is 30 h, of which the first 6 h are used for model initialisation and discarded. Figure 1 shows the elevation as represented in the regional climate model compared to the elevation from SRTM (Shuttle Radar Topography Mission) (Jarvis et al., 2008). The general features of the topography are captured well, but due to the much coarser resolution the highest model elevations are much lower than the actual peaks and narrow mountain ranges, for example southwest of the Karadarya catchment, are not resolved.
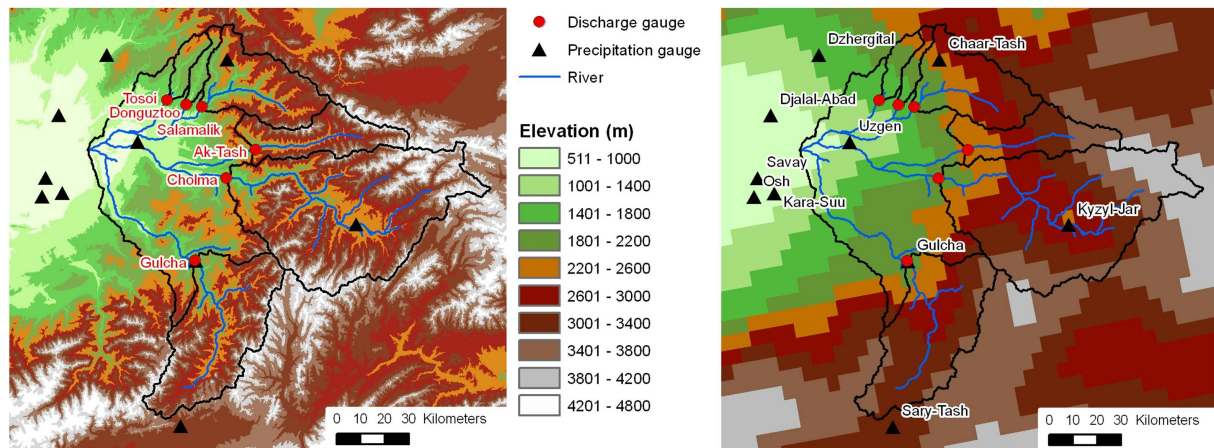
Fig. 1. The Karadarya Basin upstream of the Andijan reservoir. Left: SRTM elevation, right: elevation in the WRF model. Shown are the headwater subcatchments where the hydrological model is applied (black outlines), and their corresponding discharge gauges (red dots), as well as the precipitation gauges (black triangles).

3.1.2 Precipitation station data

For 10 gauges within or close to the Karadarya catchment daily precipitation data for the time period 1959–1990 are retrieved from the National Climatic Data Center (NCDC, 2005) and complemented by data from the National Hydrometeorological Services of Uzbekistan and Kyrgyzstan. Precipitation measurements are affected by systematic errors due to evaporation, wetting and wind-losses. Precipitation undercatch of the Tretyakov gauge, which is the common gauge in this region, due to wind losses is corrected using the approach of Yang et al. (1995). These regression equations (Eqs. 4–7 in Yang et al., 1995) give the catch ratio of the Tretyakov gauge in comparison to the double fence intercomparison reference and were derived through the World Meteorological Organization Solid Precipitation Measurement Intercomparison. Measured temperature and wind data, which are required as inputs for this approach, are not available for all gauges. Therefore, temperature data are derived from the WRF downscaled ERA-40 data and, after consulting the WRF output and the available measurement data, an average wind speed of $2\,\mathrm{m\,s}^{-1}$ is assumed. The undercatch correction results, on average, in an increase of the measured values by 10 %.

3.1.3 Interpolation of station data by inverse distance weighting

In addition to the more sophisticated methods described in the following sections, the precipitation data are also interpolated using a simple inverse distance weighting (IDW) approach (Shepard, 1968). In this method precipitation for a location j is estimated as weighted mean of the gauge observations at surrounding stations. The weights are determined based on the inverse of the distance between location j and the gauge locations, raised to the power of b. The method

is applied in a standard way (e.g. Goovaerts, 2000; Lloyd, 2005) using an inverse distance power of two, and with the distance calculated as Euclidean distance in a two dimensional plane:

$$P_j = \frac{\sum_{i=1}^{n} (w_i \cdot P_i)}{\sum_{i=1}^{n} (w_i)}, \quad w_i = d_{ij}^b, \quad b = -2, \tag{1}$$

with P_j: estimated precipitation at location j; P_i: observed value at gauge i; d_{ij}: horizontal distance between i and j; and n: number of gauges.

3.1.4 Interpolation of station data using spatial fields from downscaled reanalysis data

The approach developed here interpolates daily time series of station data using spatial fields from downscaled reanalysis data. The WRF-ERA-40 precipitation data are first aggregated to monthly maps. For the generation of daily precipitation maps, a scaling factor at the station locations is calculated by dividing the daily gauge observation at location i by the mean monthly precipitation of the WRF-ERA-40 data at location i:

$$F_i = \frac{P_i}{M_i}, \tag{2}$$

with F_i: scaling factor at station location i; and M_i: mean monthly precipitation of the WRF-ERA-40 data at location i.

In order to avoid abnormally large values when dividing by very small numbers, stations where the mean monthly precipitation is less than $1\,\mathrm{mm\,month}^{-1}$ are excluded from the calculation of the scaling factor for that month. The calculated factor is next interpolated to all locations j on a

1 km × 1 km grid using the inverse squared distance weighting method:

$$F_j = \frac{\sum\limits_{i=1}^{n} \left(d_{ij}^{-2} \cdot F_i \right)}{\sum\limits_{i=1}^{n} \left(d_{ij}^{-2} \right)}, \tag{3}$$

with F_j: scaling factor interpolated to location j.

Multiplication of the interpolated scaling factor map with the mean monthly WRF-ERA-40 data mapped to a 1 km × 1 km grid then results in the daily precipitation map:

$$P_j = F_j \cdot M_j, \tag{4}$$

with M_j: mean monthly WRF-ERA-40 data at location j.

Two different variants of this method are tested: (i) in the variant WRFadj-all the monthly maps are calculated as means over the whole period 1960–1990, i.e. for the interpolation of station data in January 1960 a map of the mean monthly precipitation over all Januaries is used; (ii) in the variant WRFadj-ind the monthly maps are calculated for each year individually, i.e. for the interpolation of station data in January 1960 a map of the monthly precipitation of January 1960 is used.

3.1.5 Interpolation of station data using monthly fields derived by multi-linear regression

Due to the topography and the main wind direction from the west, precipitation in the catchment generally increases with increasing elevation and decreases to the south and east. Precipitation is therefore also interpolated by multiple linear regression against elevation, x and y. Since the correlations between precipitation and these three variables are higher for monthly than for daily data, the multi-linear regression is performed on monthly data. We apply the stepwise backwards approach (e.g. Backhaus et al., 2003), setting the p value of an F statistic for exclusion and inclusion to 0.1 and 0.05 respectively. This means that in the initial model all variables (elevation, x and y) are included. At each step the explanatory power of the current model is compared with incrementally smaller and larger models. This stepwise backward approach can lead to different variables being included in the final regression equation than the stepwise forward approach. An initial analysis showed lower standard errors and lower root mean squared errors for the stepwise backward approach, which was thus selected for this study.

After calculating the monthly regression maps, daily precipitation maps are calculated in the following way: for each day scaling factors between the daily precipitation and the monthly regression at the station locations are calculated and interpolated to a 1 km × 1 km grid using IDW (see Eqs. 2 and 3, but M_i is here replaced by the monthly regression at location i).

The interpolated scaling factors are multiplied with the monthly map derived by multi-linear regression to generate the daily precipitation fields (see Eq. 4, but M_j here denotes the mean monthly regression value at location j). Again two variants of this method are applied using (i) monthly means over a month in all years (MLR-all) and (ii) monthly means of individual years (MLR-ind).

3.1.6 Gridded precipitation data

APHRODITE (Yatagai et al., 2012) is a daily gridded precipitation data set at a resolution of 0.25° covering Asia, the former Soviet Union and the Middle East. It is based on gauge observations from the Global Telecommunication System, precompiled data sets like from the Global Historical Climatology Network, the National Climate Data Center, the Food and Agriculture Organization of the United Nations, and others, as well as additional data from national hydrometeorological services. The spatial interpolation scheme takes into account the effect of mountain ranges by giving a high weight to gauges on slopes inclined to the target location and a low weight to gauges on the leeward side behind a mountain ridge.

Other, globally available precipitation data are only assessed with respect to their spatial distribution and subcatchment mean values and not included in the evaluation by hydrological modelling. We use three different data sets based on interpolated station data: the Global Precipitation Climatology Centre (GPCC) full data reanalysis version 6 (Schneider et al., 2011), the University of Delaware (UDEL) precipitation data set version 2.01 (Legates and Willmott, 1990), and the University of East Anglia Climate Research Unit (CRU) TS 3.10.01 (Mitchell and Jones, 2005). These data are all available as monthly time series with a spatial resolution of 0.5°. Furthermore we also inspected the precipitation data from the ERA-40 reanalysis (Uppala et al., 2005) at their original resolution, which is a spectral resolution of T159, regridded to a regular geographic coordinate system of 1°. For an overview, Table 2 lists all precipitation data sets used in this study.

3.2 Point based evaluation of the precipitation data

In the first step, the precipitation data sets are evaluated by comparison to observed station data. The precipitation data generated by downscaling the ERA-40 reanalysis data with WRF are directly compared to observed station time series. For this, WRF data from the pixel which contains the station location are extracted. There are limitations to such a comparison between point observations and pixel-based data, as gauge observations cannot be considered as ground truth for a 12 km × 12 km WRF pixel area, and due to errors in the undercatch correction or in the observation data themselves. However, a first indication of the performance of the WRF precipitation data for the Karadarya catchment is provided.

The interpolated precipitation data sets are evaluated by cross validation. In this method only a part of the stations is

Table 2. Overview of the precipitation data sets used in this study.

Abbreviation	Description
WRF	Precipitation from the ERA-40 reanalysis data downscaled using WRF to a resolution of 12 km.
WRFadj-ind	Station data interpolated using monthly precipitation maps of WRF, monthly maps from individual years.
WRFadj-all	Station data interpolated using monthly precipitation maps of WRF, monthly maps averaged over all years.
MLR-ind	Station data interpolated using monthly precipitation maps from multi-linear regression, monthly maps from individual years.
MLR-all	Station data interpolated using monthly precipitation maps from multi-linear regression, monthly maps averaged over all years.
IDW	Station data interpolated using the inverse squared distance weighting method.
APHRODITE V1003R1	Gridded observation based daily precipitation data set with a resolution of 0.25° (Yatagai et al., 2012).
GPCC v6	Gridded observation based monthly precipitation data set with a resolution of 0.5° (Schneider et al., 2011).
CRU TS 3.10.01	Gridded observation based monthly precipitation data set with a resolution of 0.5° (Mitchell and Jones, 2005).
UDEL 2.01	Gridded observation based monthly precipitation data set with a resolution of 0.5° (Legates and Willmott, 1990).
ERA-40	Precipitation data from the ERA-40 reanalysis data at a resolution of 1° (Uppala et al., 2005).

used for the interpolation, and the others are employed for the evaluation of the interpolated values at these locations. As the error statistics are only calculated at the locations of the stations, the value of such an analysis may be very limited if the gauges are not in representative locations for the catchment (for example in a situation where precipitation increases with elevation, but most stations are located in relatively low elevations). Also, in regions with only few stations, the interpolated fields may be strongly changed if stations with a high weight in the interpolation are left out. In this study we only remove one station from the data set at a time. The interpolated time series at this location is compared to the observed time series and evaluated using bias and mean absolute error of the daily time series.

3.3 Evaluation of areal precipitation estimates based on simulated discharge

3.3.1 Approach

The suitability of different precipitation estimates is tested by comparing observed discharge and discharge simulated by a hydrological model driven with the different precipitation estimates. Running a hydrological model with a different precipitation data set than the one it has been calibrated with usually results in lower model performance, and is therefore not a suitable approach for the comparison of precipitation data sets. Generally there are two possibilities to evaluate different precipitation data sets by hydrological modelling: calibrating the model for each precipitation data set, and Monte Carlo simulations using various parameter values between defined bounds.

The Monte Carlo approach is for example applied by Gourley and Vieux (2005) and Heistermann and Kneis (2011). In this approach Monte Carlo simulations are carried out for each precipitation data set and the selected goodness of fit measure is calculated for each simulation. For each precipitation data set one then evaluates the mean goodness of fit over the whole or subsets of the Monte Carlo ensemble and ranks the precipitation data sets according to this value. An advantage of this approach compared to the cali-

bration approach is that it easily allows evaluating the model for various subsets of the data, e.g. only for high or low flows. However, in some cases, particularly when parameters have a linear influence on the fraction of rainfall generating runoff and the precipitation estimates do not have random errors but a systematic bias, the Monte Carlo approach may lead to wrong conclusions. Heistermann and Kneis (2011) give the following example: assume a very simple linear hypothetic catchment with $Q = \psi \cdot P$, where Q represents the runoff, ψ the runoff coefficient with values between 0 and 1, and P the precipitation. Monte Carlo simulations with uniform sampling over the runoff coefficient are performed for a precipitation data set without bias and a second precipitation data set characterised by a constant bias. In the next step the root mean squared error (RMSE) between simulated and observed discharge is evaluated for each simulation. It can be shown (see Heistermann and Kneis, 2011) that in a system with the true value of the runoff coefficient $\psi_{\text{true}} < 0.7$ ($\psi_{\text{true}} > 0.7$) the mean RMSE of a precipitation data set with a negative (positive) bias is lower than for the unbiased precipitation data set so that the biased precipitation data set would be classified as the better one. While very obvious ill-posed settings may be avoided by careful analysis of the model, less obvious cases may not always be avoided from the outset.

One solution to the problem of false rankings is to evaluate not all but only a percentage of the best Monte Carlo simulations. Heistermann and Kneis (2011) showed that reducing the number of the evaluated best performing Monte Carlo runs reduces the number of false rankings, though it also decreases the discriminatory power between different precipitation data sets. This can be seen as a transition to the model calibration approach.

Calibrating the model for each precipitation data set may have the disadvantage that model parameters can partly compensate for inadequacies of the precipitation data sets, which might result in different precipitation data sets being hardly distinguishable with respect to the simulated hydrograph. On the other hand, the approach is less prone to false rankings, as ill-posed settings would rather result in indistinguishable

precipitation data sets so that setting the parameter bounds and analysing the model for possible ill-posed settings becomes a less sensitive issue. Another advantage is that by using an optimisation algorithm instead of random Monte Carlo runs, usually better model performances are achieved with a lower number of simulations. As it is seen as more important to avoid false rankings than to discriminate between already similar precipitation data sets, the model calibration approach is selected for this study.

In order to gain more information on different aspects of the performance of a precipitation data set, a precipitation bias factor is introduced as additional calibration parameter. The precipitation estimate is then evaluated with respect to the bias based on the precipitation bias factor, and evaluated with respect to the temporal dynamics based on the objective function used for model calibration.

Three different sources of uncertainties are considered in this study. Uncertainties in the precipitation bias factor (as part of the parameter uncertainties) need to be considered, because the precipitation factor of the best optimised parameter set might differ from other equally good performing parameter sets. The model calibration is then repeated for different time periods in order to evaluate the robustness of the precipitation bias factor and ranking of the objective function value with respect to the selected time period. Finally the robustness of the results with respect to uncertainties in model inputs is investigated using sensitivity analyses.

3.3.2 Description of the hydrological model

The hydrological model WASA (Güntner, 2002; Güntner and Bronstert, 2004) is a semi-distributed daily time step model based on process-oriented and conceptual approaches. It was recently extended for high mountain areas by introducing elevation zones, and a snow and glacier mass balance module based on the temperature index method. The model calculates evaporation from the interception storage and open water bodies with the Penman–Monteith equation (Monteith, 1965), evapotranspiration using the two-layer model of Shuttleworth and Wallace (1985), infiltration with the Green–Ampt approach (Green and Ampt, 1911), the generation of infiltration and saturation excess surface runoff, and percolation through a multiple layer soil store. In the model version applied for this study, surface and subsurface flow between model units within a subcatchment (i.e. lateral flow redistribution) are neglected, and subsurface flow is separated between interflow and groundwater based on a calibration parameter. The simulation of small events during the low flow period is improved by introducing an additional parameter for the fraction of the catchment where rainfall directly leads to runoff, like riparian areas, roads or rock areas connected to a stream.

The spatial discretisation of WASA is originally based on hillslopes with characteristic toposequences (Güntner and Bronstert, 2004; Francke et al., 2008). For this study a much simpler approach based on hydrologic response units defined by elevation bands is selected in order to reduce the computation time and allow for a higher number of model simulations for calibration and uncertainty analysis. For each 200 m elevation band the dominant soil and vegetation cover and the glacier fraction are taken into account.

For the model calibration 11 parameters are selected (Table 3). These affect the snow and glacier melt routine (snowmelt factor, glacier melt factor, melt temperature), the soil hydraulic conductivity for infiltration and percolation (kf_corr_f, k_sat_f), the subsurface runoff (frac2gw, interflow delay factor, groundwater delay factor), the fraction of the catchment area leading to direct runoff (frac_riparian), the occurrence of saturated areas as a function of the current soil moisture state (sat_area_var) and the precipitation input (precipitation bias factor).

3.3.3 Hydrological model set-up

The model is set up for the Karadarya catchment based on the SRTM digital elevation model (Jarvis et al., 2008) for elevations and the delineation of subcatchments. As land cover input the MODIS land cover product with a resolution of 500 m (MCD12Q1; Friedl et al., 2002) is applied using the most frequent land cover class over the time period 2001–2008. Mean monthly leaf area index (LAI) values by elevation zone, subcatchment and land cover class are calculated from the 8 day MODIS LAI product with a resolution of 1 km for 2001–2008 (MOD15A2, Myneni et al., 2002). For the soil data a digitised map from the Kyrgyz Atlas (scale, 1 : 1 500 000; Academy of Science of the Kyrgyz SSR, 1987) is used and missing soil hydraulic parameters are assigned using pedo-transfer functions from the literature. Glacier areas are delineated from a LANDSAT MSS scene (resolution 79 m) in summer 1977 using a combination of automated classification and manual digitising. Daily time series of solar radiation, temperature, temperature lapse rate and humidity are taken from the WRF downscaled ERA-40 data described above. The temperature data are corrected for the difference between the SRTM DEM and the WRF topography using daily lapse rates as simulated by the WRF model. All meteorological input data are aggregated to subcatchment mean values.

3.3.4 Model optimisation and analysis of parameter uncertainties

The model is automatically calibrated against observed discharge using six-year simulation periods (1961–1966, 1967–1972, 1973–1978, 1979–1984, 1985–1990); prior to this the model is initialised using an additional simulation period of two years. In order to consider both high and low flows and to keep the overall bias low, the following objective function

Table 3. Calibration parameters including values for the lower and upper bounds.

Routine	Parameter	Unit	Lower bound	Upper bound
Snow and glacier melt	snowmelt factor	mm °C^{-1} day^{-1}	1	7
	melt temperature	°C	-2	2
	glacier melt factor	mm °C^{-1} day^{-1}	0	7
Infiltration and percolation	kf_corr_f	–	0.01	100
	k_sat_f	–	0.01	100
Subsurface flow	frac2gw	–	0	1
	interflow delay factor	days	10	100
	groundwater delay factor	days	30	400
Generation of direct runoff from areas connected to the stream	frac_riparian	–	0	0.05
Spatial variability of saturated areas within a model unit	sat_area_var	–	0	0.3
Precipitation input	precipitation bias factor	–	0.5	2.0

is applied:

$$\text{Obj.function} = 0.5 \times (\text{NSE} + \text{LogNSE})$$
$$- 2 \times \max(\text{Bias} - 0.05, 0), \tag{5}$$

where NSE is the Nash–Sutcliffe efficiency value, LogNSE is the Nash–Sutcliffe efficiency calculated on logarithmic flows, and Bias is the absolute value of the overall volume bias. The Nash–Sutcliffe efficiency is particularly responsive to errors in high discharge values, while the Nash–Sutcliffe efficiency for logarithmic flows is more sensitive also for errors in low flows so that the average of these two measures results in a more balanced evaluation. The maximum possible value of the objective function is 1, which would indicate perfect agreement between simulated and observed discharge. As the bias in the precipitation estimate is evaluated using the precipitation bias factor, the bias in the simulated discharge should be very low. The objective function is therefore additionally penalised if the bias is greater than 0.05 or 5 % of the observed discharge.

Despite the lack of hard data, two further constraints for the snow and glacier mass balance modules are introduced in order to avoid unrealistic simulations. First, an elevation is defined below which snow is not expected to accumulate over several years. This elevation is derived from LANDSAT images in summer and set to 4200 m for the Karadarya catchment. For each year, the number of elevation zones where simulated snow does not melt away in elevation zones below this elevation is counted, and if it is above a threshold of one per year, the simulation is discarded. For example, for a model evaluation period of six years as in this study, the simulation will still be accepted, if in one year the minimum snow water equivalent is above zero in up to six elevation zones below the threshold elevation. It will also be accepted, if in all six years the minimum simulated snow water equivalent is above zero in only one elevation zone below the threshold elevation, but if the number of elevation zones or years where snow accumulates is higher, the simulation

is discarded. Second, no measured glacier mass balances are available for the Karadarya catchment, but based on measured mass balances in other catchments in Central Asia a wide range of -1000 mm a^{-1} up to $+200$ mm a^{-1} is set as a further constraint.

The model optimisation and parameter uncertainty analysis is performed using the DDS-AU algorithm (dynamically dimensioned search – approximation of uncertainty; Tolson and Shoemaker, 2008). The analysis of parameter uncertainties is particularly important for the investigation of how much the calibrated precipitation bias factor varies between the best and other equally good parameter sets. The DDS-AU algorithm is an informal method (in contrast to formal Bayesian approaches) similar to GLUE (generalised likelihood uncertainty estimation, Beven and Binley, 1992), but instead of simple Monte Carlo simulations, which usually result in a high fraction of runs very far from the objective function maximum, a number of short optimisation runs are started. These short optimisation runs are meant to get into the region of the optimum, but the length of the optimisation run is also short enough that they mostly do not reach the objective function maximum. For each optimisation, 200 of these short DDS runs are started. The number of model evaluations in each DDS run (the length of the DDS run) is set randomly between three and seven times the number of calibration parameters resulting in 33 to 77 model evaluations. In order to assure that at least one very good parameter set is found, one run with 3000 model evaluations is performed. The individual DDS runs are independent from each other. The short optimisation runs with 33 to 77 model evaluations help to approximate the uncertainty bounds, while the long run with 3000 model evaluations is meant to come very close to the global optimum.

3.3.5 Sensitivity to model inputs

Uncertainties in the calibration parameters are directly considered through the selected calibration approach. However, uncertainties in the model inputs may also have an impact. Influences on the precipitation bias factor are particularly expected from uncertainties in those inputs which affect evapotranspiration and thus also the water balance. In order to estimate the magnitude of the effect of these uncertainties, sensitivity analyses are performed. For these sensitivity analyses, we selected those inputs which are expected to strongly affect the calculated evapotranspiration: the climate variables solar radiation and wind velocity; the plant parameters plant height, rooting depth, stomata resistance and the matrix potential values below which transpiration is reduced or ceases; and soil depth (Table 4). Solar radiation, wind velocity and plant height directly influence the potential evapotranspiration. Root depth and soil depth determine the amount of soil water available to plants for transpiration. Minimum stomata resistance influences the potential transpiration rate, and the two matrix potential values determine how this rate changes with decreasing soil water. The model is recalibrated for each variation factor of each of the inputs listed in Table 4 varying one factor at a time. This analysis is performed for all of the six subcatchments and all precipitation data sets, but in order to restrict computing time the analysis is constrained to one time period (1979–1984) and one long DDS run with 3000 model evaluations per subcatchment and precipitation data set (parameter uncertainties resulting from equifinality are not considered).

4 Results and discussion

4.1 Characteristics of the precipitation data sets

4.1.1 Point based evaluation

Comparison of downscaled ERA-40 precipitation data to observations

First, the precipitation data generated by downscaling the ERA-40 reanalysis data with WRF are evaluated relative to observed station time series. As this comparison can deliver information about the performance of the downscaled precipitation data at a few points in the catchment only, it is complemented by visual inspection of the spatial distribution of precipitation (Sect. 4.1.2). Large deviations in terms of volume and/or an unrealistic spatial pattern may indicate a priori that in these areas the downscaled precipitation data are not suitable as input for water balance modelling or as spatial fields for the interpolation of station data.

For mean annual precipitation, the bias of the WRF downscaled ERA-40 data compared to the gauge observations in the study area is in the range of +20 % to −30 % (Table 5). There is no relationship of the bias with elevation.
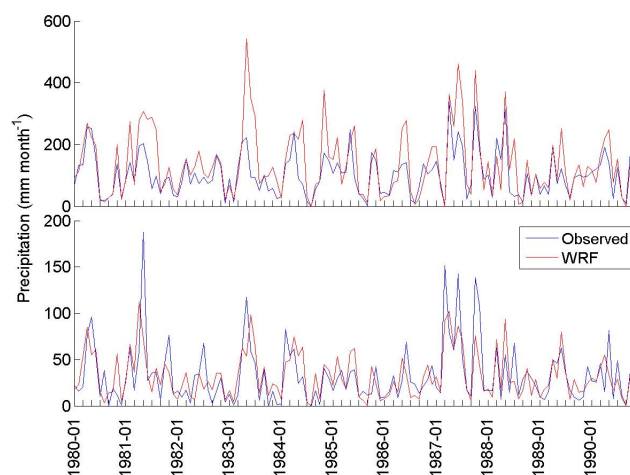

Fig. 2. Comparison of monthly time series over the period 1980–1990 of observed precipitation and WRF downscaled ERA-40 data at the gauge location for Chaar-Tash (top) and Kyzyl-Jar (bottom).

The squared correlation coefficients for daily time series only reach values around 0.3 for the stations at lower elevations and are even lower for the high elevation stations. Monthly precipitation time series from the WRF model at the station locations generally correspond much better to the station data, with squared correlation coefficients around 0.6. Nevertheless, large disagreements may exist for individual months or seasons, for example a strong overestimation in summer 1983 in Chaar-Tash, or a considerable underestimation in June 1981 in Kyzyl-Jar (Fig. 2).

The agreement between gauge and WRF precipitation data is similar to RCM applications in other mountainous regions. For example, Frei et al. (2003) studied the performance of five RCMs (CHRM, HadRM, HIRHAM, REMO, ARPEGE) at a resolution of 0.5° with boundary conditions from ERA15 in the European Alps. The bias of the areal mean of simulated precipitation ranged from +3 % to −23 % in winter and −5 % to −27 % in summer. Suklitsch et al. (2011) evaluated four high resolution (10 km) RCMs (WRF, MM5, REMO, CLM) driven by ERA-40 data over a simulation period of 1 yr and found bias values up to −50 % and +100 % for individual seasons and subregions of the Alps. Higher correlation values between observed and simulated time series at monthly as compared to daily resolution are typical (e.g. Hurkmans et al., 2008); this can be explained by the fact that only a part of the precipitation can be modelled deterministically, and errors from random processes partly average out on a monthly timescale.

Cross validation of interpolated precipitation data sets

The precipitation data sets interpolated from gauge observations using monthly fields from WRF downscaled ERA-40 data (WRFadj-all and WRFadj-ind), multi-linear regression (MLR-all and MLR-ind) and inverse distance weighting are

Table 4. Climate, plant and soil inputs selected for the sensitivity analyses.

Input	Original value	Variation
Solar radiation	WRF downscaled ERA-40 data	Multiply original value by 0.7 and 1.3 (based on differences between the WRF downscaled ERA-40 data and satellite based data from the NASA Surface Radiation Budget (SRB) product, version 3.1, http://eosweb.larc.nasa.gov/sse/; mean seasonal differences over 1984–2001 are 21 % in spring and 12 % in summer).
Wind	Constant value of $2 \, \mathrm{m \, s^{-1}}$	Multiply original value by 0.5 and 2 (based on values of WRF downscaled ERA-40 data and available station data).
Plant height	Varies by land cover; e.g. grassland 30 cm	Multiply original value by 0.25 and 4.
Root depth	Varies by land cover; e.g. grassland 20 cm	Multiply original value by 0.25 and 4.
Soil depth	Varies by soil type between 35–140 cm; mostly 50–100 cm	Multiply original value by 0.5 and 2.
Minimum stomata resistance	Varies by land cover; grassland $126 \, \mathrm{s \, m^{-1}}$ (based on values from Körner, 1994)	Multiply original value by 0.5 and 2 (according to ranges as given in Körner, 1994).
Matrix potential below which transpiration is reduced (minsuction)	-600 hPa (according to values from Feddes and Raats, 2004)	Apply a value of -200 hPa and $-15\,000$ hPa throughout the whole catchment.
Matrix potential below which transpiration is only 1 % of the potential transpiration (maxsuction)	$-15\,000$ hPa	Apply a value of -8000 hPa and $-22\,000$ hPa throughout the whole catchment.

also compared using leave-one-out cross validation. Generally this analysis shows large errors for the stations Chaar-Tash in the north, Kyzyl-Jar in the east and Sary-Tash in the south of the catchment, while for the clustered stations to the west of the catchment, the errors are low (Fig. 3). In particular the methods MLR-all and MLR-ind strongly overestimate at the station Kyzyl-Jar by around 160 % and at Sary-Tash by around 35 % and 50 %, and underestimate at Chaar-Tash by approximately 35 %. By comparison, the method WRFadj-all shows a more balanced performance with bias values of -7 %, $+8$ % and $+40$ %, at these three stations. The performance of the IDW method with respect to bias is between the methods MLR and WRFadj. Regarding the mean absolute error, the methods IDW and WRFadj-all have approximately comparable results with higher errors in Chaar-Tash of the method WRFadj-all and higher errors in Kyzyl-Jar of the method IDW. Both with respect to bias values and mean absolute errors, the performance of WRFadj and MLR with monthly fields averaged over all years ("-all") is similar or slightly better than the versions which use monthly fields for individual years ("-ind").

The low performance of the methods MLR-all and MLR-ind results from the fact that omitting a station from the interpolation changes both the mean monthly fields generated by linear regression and the adjustment factors for the particular day interpolated by IDW. Omitting the stations Kyzyl-Jar or Sary-Tash results in very different monthly regression fields so that then IDW, WRFadj-all and WRFadj-ind clearly outperform the methods MLR-all and MLR-ind. The methods WRFadj-all and WRFadj-ind are similar to the methods MLR-all and MLR-ind, in that also first monthly fields are calculated, and second these are adjusted to daily stations values. However, as the precipitation stations are not used in the calculation of the monthly fields, the methods WRFadj-all and WRFadj-ind show a more robust behaviour, when individual stations are omitted.

4.1.2 Spatial distribution and temporal dynamics of subcatchment mean values for the different precipitation data sets

Despite only relatively small differences at the station locations, the precipitation data sets are very different with

Table 5. Comparison of observed and WRF simulated precipitation at the station locations: bias, daily and monthly squared correlation coefficient calculated over the period 1960–1990.

Station	Elevation (m)	Bias (%)	r^2 month	r^2 day
Chaar-Tash	2748	18	0.54	0.25
Djalal-Abad	971	−16	0.61	0.26
Kyzyl-Jar	2230	−12	0.53	0.19
Sary-Tash	3155	11	0.42	0.13
Uzgen	1014	−31	0.66	0.28
Dzhergital	1198	−25	0.65	0.28
Gulcha	1542	3	0.68	0.27
Savay	753	−1	0.59	0.29
Kara-Suu	866	2	0.63	0.32
Osh	887	9	0.59	0.31

Fig. 3. Bias and mean absolute error calculated from cross validation for the interpolation methods WRFadj-all, WRFadj-ind, MLR-all, MLR-ind and IDW for precipitation stations in or close to the Karadarya catchment.

respect to their spatial distribution (Fig. 4). There are a few agreements, for example all precipitation data sets indicate relatively high precipitation along the mountain range to the north and northeast of the catchment, and relatively low precipitation in the valley close to the station Kyzyl-Jar.

The precipitation data sets estimated by multi-linear regression show a strong increase of precipitation with increasing elevation; additionally precipitation also decreases to the south and to the east (Fig. 4d, e). The WRF downscaled ERA-40 precipitation data indicate spots with very high precipitation values in the southern part of the catchment (Fig. 4c). This is likely to be caused by the coarser topography and the poor representation of one of the mountain ridges in the southwest of the catchment (Fig. 1). Thus in the WRF model the valley in the southern part of the catchment is less sheltered from the wind than in reality, which might cause too high precipitation of the WRF model at this location. The mean annual precipitation maps of the precipitation data set interpolated using monthly maps of the WRF precipitation (Fig. 4a, b) are very similar to the WRF precipitation map, with the main difference that in the former the precipitation at the station locations is closer to the observed values. The spatial distribution of precipitation in the IDW interpolated and the APHRODITE precipitation data sets both markedly differ from the other precipitation data sets in that they both indicate only very little precipitation in the southern and southeastern parts of the catchment (Fig. 4f, g).

Naturally there is much more agreement among the different data sets in terms of the temporal dynamics, as all data sets except for the WRF data originate from station data. The subcatchment mean monthly precipitation data show a bimodal regime with a major peak in April–May and a minor peak in October (Fig. 5). There is a strong agreement between the different precipitation data sets for the three northern subcatchments; only the WRF downscaled ERA-40 precipitation data exhibit a slightly late seasonality with high precipitation also in June–July (Fig. 5a–c). In the three eastern and southern subcatchments, the different precipitation
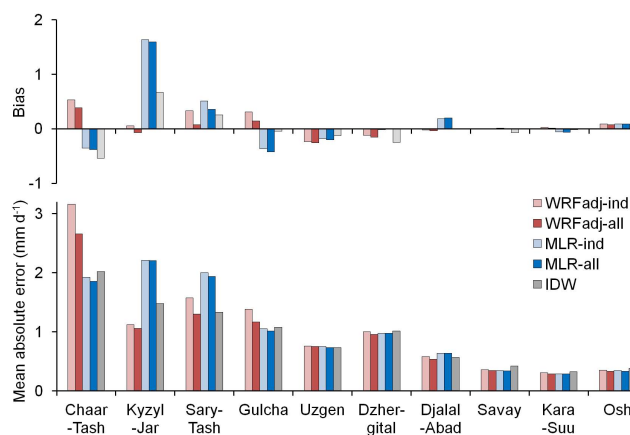
data sets still agree on the general seasonal distribution, but they strongly differ in magnitude.

4.1.3 Comparison to global gridded data sets

Maps of mean annual precipitation from APHRODITE and GPCC show a very similar spatial distribution in the study area, with a distinctive precipitation maximum in the north (Fig. 6). By comparison, the other two gauge based precipitation data sets UDEL and CRU indicate a much lower mean annual precipitation and only show a very weak precipitation increase to the north of the catchment. In contrast to the gauge based data sets, the ERA-40 reanalysis data shows higher precipitation in the southern compared to the northern part of the Karadarya catchment. The precipitation estimates from these global data sources are considerably lower than our estimates based on the methods WRFadj, WRF or MLR (Fig. 4a–e).

The differences between the precipitation data sets applied in this study and global gridded data sets are clearly demonstrated in the values of the subcatchment mean precipitation (Fig. 7). If the precipitation data set MLR-all is used as a reference, UDEL and CRU underestimate precipitation by around −50 % and −60 % in the six subcatchments, for GPCC the underestimation varies from about −15 % in the two northern subcatchments to −50 % in Gulcha and Cholma, and for the ERA-40 data this varies between an underestimation of −70 % for the northern subcatchments to only a very small difference of the values in Cholma.

As the number of stations included in the GPCC and APHRODITE data in this region is higher than in CRU and UDEL, this is likely to be the reason for the differences between APHRODITE and GPCC on the one hand and CRU and UDEL on the other hand. The ERA-40 data are obviously too coarse to derive areal precipitation estimates for
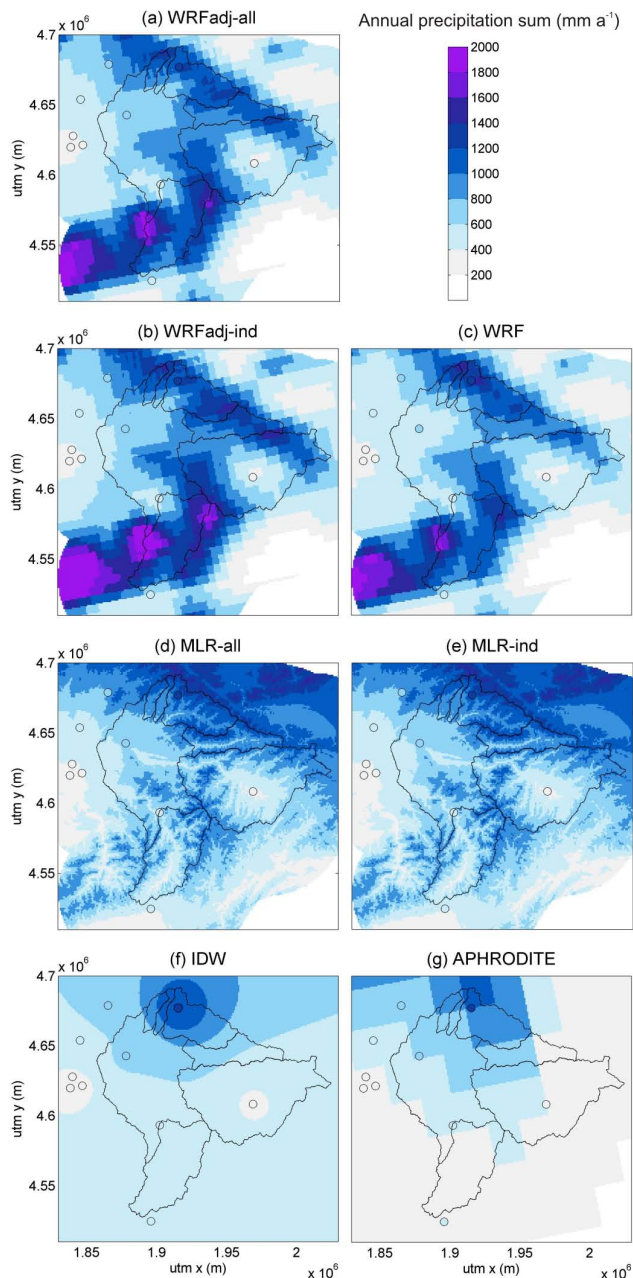
Fig. 4. Estimates of the mean annual precipitation (1960–1990) over the Karadarya catchment using different methods. Circles indicate measured precipitation at stations. Lines indicate subcatchment borders.
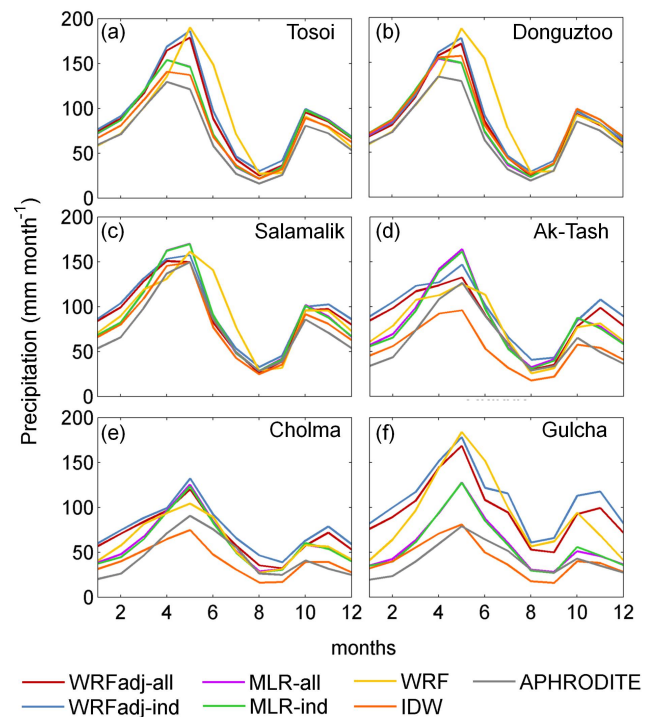

Fig. 5. Monthly subcatchment mean precipitation (1960–1990) for 6 subcatchments of the Karadarya Basin and 7 different precipitation estimates.

4.2 Evaluation of the precipitation data based on simulated discharge

4.2.1 Parameter distributions and correlations between parameters

Parameter distributions for the best 20, 50, 100 and 150 parameter sets are shown as an example for the subcatchment Gulcha, the precipitation data set WRF and the calibration period 1979–1984 (Fig. 8). The general behaviour seen in this example is also typical for the other calibration cases. Most importantly for this study, the precipitation bias factor is confined to a very narrow range, indicating that the problem of identifying the precipitation bias factor is well defined. For many other parameters, good models are achieved nearly over the whole parameter range. For example, the parameters glacier melt factor, k_sat_f, kf_corr_f and sat_area_var are not constrained at all. The remaining parameters are between these two extremes; while the best 150 parameter sets may still include parameters from the whole parameter range, the parameters are confined to a more narrow range if one considers the best 20 or best 50 parameter sets only. As a consequence of this equifinality, i.e. the fact that very different parameter sets result in comparable model performances, unconstrained parameters can for example not be used for catchment characterisation, and it is not possible to transfer individual parameters to catchments with similar
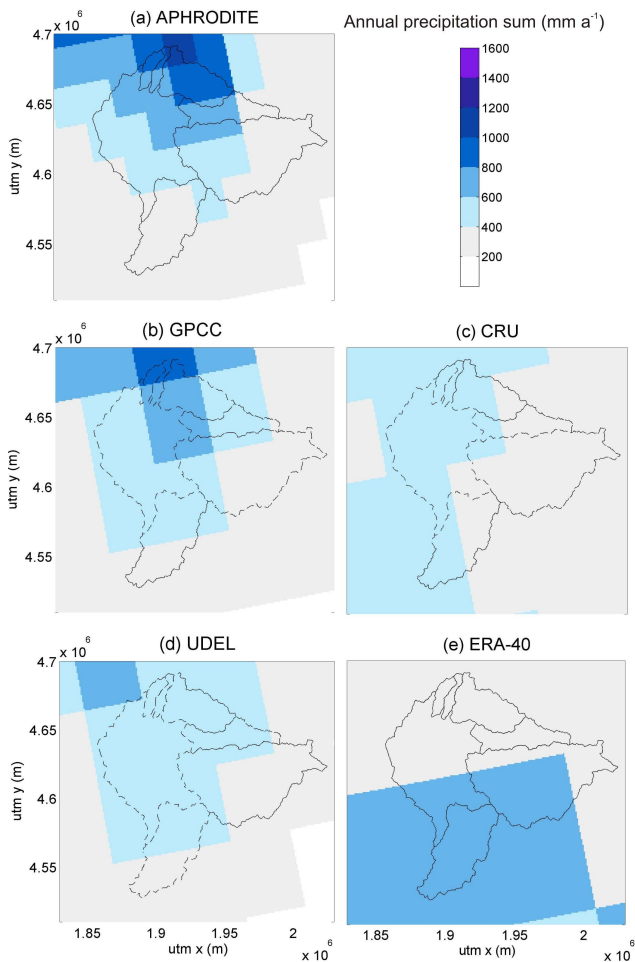
catchments of the size as in this study. The higher precipitation of the ERA-40 cells in the southern part of the catchment is probably caused by the higher elevation of these cells in the ERA-40 model. At this resolution smaller-scale features such as the Fergana range in the north or the valley around Kyzyl-Jar in the west of the catchment cannot be represented.

Fig. 6. Mean annual precipitation (1960–1990) for the Karadarya catchment from APHRODITE and four globally available gridded precipitation data sets GPCC, CRU, UDEL and ERA-40.

Fig. 7. Mean annual precipitation (1960–1990) for 6 subcatchments of the Karadarya Basin for the precipitation data sets MLR-all, APHRODITE and four global precipitation data sets.

reduction in the range of the precipitation bias factor is expected to be comparatively small for the catchments studied here. In other catchments where glacier melt accounts for a higher percentage of total runoff it may not be possible to constrain the precipitation bias factor to a narrow range without glacier mass balance data.

4.2.2 Objective function values and predictive uncertainties resulting from parameter uncertainties

Despite the differences between the precipitation estimates, most of the time they result in rather similar simulated discharge time series. This is also reflected in the objective function values (Fig. 9), which in many cases reach very similar values, both for the best and also for the best 20, best 50 or best 150 parameter sets. Most noticeable exceptions from this are the consistently lower values of the objective function values in the subcatchments Tosoi and Donguztoo for the model driven with WRF precipitation data; lower objective function values for the model driven with WRF precipitation data are also observed in other subcatchments for some time periods, e.g. in Salamalik and Ak-Tash for 1979–1984 and 1985–1990. Additionally, some precipitation products result in lower objective function values at only few gauges and time periods: MLR-ind in Cholma 1973–1978, WRF-ind in Cholma 1979–1984, and WRF-ind in Gulcha 1961–1966.

Lower objective function values of the model driven with WRF precipitation data may be explained by the difficulty of WRF to correctly predict the precipitation amount on a particular day (see comparison to observed gauge precipitation, Sect. 4.1.1). A possible reason why lower objective function values of the models driven with WRF data are predominantly observed in Tosoi and Donguztoo is the smaller size of these subcatchments. This results in less spatial averaging and smoothing of precipitation. Another possible cause

characteristics. This does however not impede the objectives of this study.

Scatter plots of parameter pairs for the best 50 or 150 parameter sets (not shown here) demonstrate that there are hardly any or only very low correlations between the precipitation bias factor and any other parameter. This is in accordance with the relatively narrow ranges of the precipitation bias factor after calibration. In some cases, there is a weak correlation of the bias factor to the glacier melt factor and to frac_riparian. Higher glacier melt increases the total runoff at the expense of a more negative glacier mass balance, and an increase in frac_riparian would result in a higher percentage of direct runoff thus decreasing actual evapotranspiration. The correlation to the glacier melt factor implies that it may be possible to further confine the precipitation bias factor if glacier mass balance data were available to further constrain the glacier melt factor. However, due to the small glacier fraction and relatively high precipitation, glacier melt is only a small fraction of the total annual runoff so that the
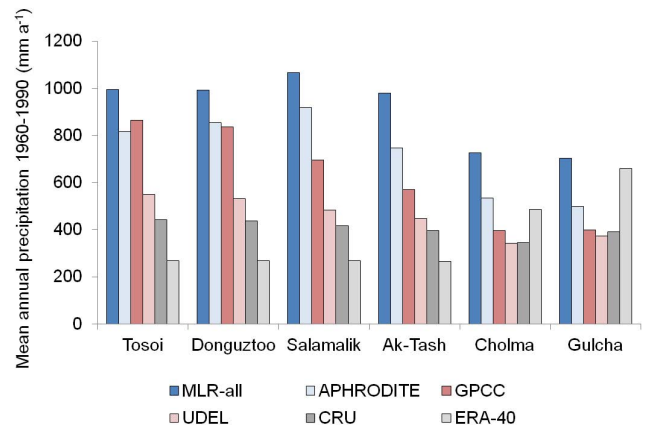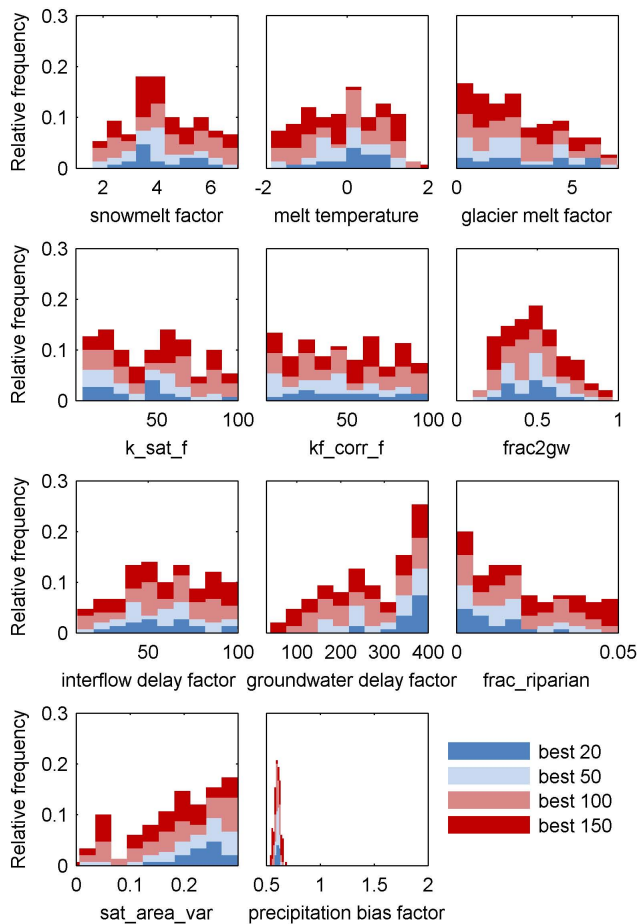
Fig. 8. Histograms of the parameter distributions for the best 20, 50, 100 and 150 parameter sets for the subcatchment Gulcha, precipitation estimate "WRF" and time period 1979–1984.

is the higher percentage of rainfall in total precipitation due to the lower elevation of these two subcatchments. For snowfall the temporal dynamics of precipitation is less important for the temporal dynamics of discharge, as snow accumulates until the melting season.

In this study the uncertainty bands for the simulated discharge are meant to describe only the parameter uncertainties, i.e. the part of the total uncertainty caused by different parameter sets reaching equally good objective function values. If a higher number of parameter sets is included in the analysis, one also includes models with a clearly lower performance. For the subcatchment Salamalik, considerably worse models are included if the analysis is based on the best 100 parameter sets (Fig. 9, row 3). It is therefore decided to focus the evaluation on the best 1, best 20 and best 50 parameter sets. The width of the uncertainty bands based on the 50 best simulations is roughly half of the mean observed flow, and the uncertainty intervals include on average around 70 % of the observed discharge data in Tosoi and Donguztoo, between 50 to 60 % in Cholma, Ak-Tash and Salamalik and

45 % in Gulcha. This shows that parameter uncertainties can only explain some part of the uncertainties and that a relevant part of the uncertainty is also caused by errors in the model structure and in the model input.

4.2.3 Variation of the precipitation bias factor by precipitation estimate, subcatchment and time period

For a well performing precipitation data set, the precipitation bias factor should be close to one, show little variability between different time periods and little variability between the different subcatchments. According to this, the two precipitation data sets based on multi-linear regression seem to be the most suitable precipitation estimates (Fig. 10). The corresponding precipitation factors are very close to one in all subcatchments, except in Ak-Tash, where precipitation is underestimated by 16 to 38 %. Based on the variability between different time periods, the precipitation estimate MLR-all, which uses monthly regression estimates averaged over 1960–1990, should be preferred over MLR-ind, which uses monthly regression estimates from individual years, as the latter shows a higher variability in the subcatchments Gulcha and Cholma. The good performance of the discharge simulations with precipitation data interpolated by MLR-all and MLR-ind despite the low performance in the cross validation is likely due to the fact that the precipitation gauges and thus also the results of the cross validation are not representative for the areal precipitation of the modelled subcatchments. If one is interested in areal precipitation estimates, cross validation can be misleading and an evaluation of different precipitation data sets using simulated discharge should be preferred. On the other hand, cross validation also indicates the dependence of this approach from individual stations with potentially strong changes to the interpolated precipitation if individual stations are removed from the data set.

In contrast to the precipitation data sets interpolated by MLR, the direct use of WRF precipitation results in a precipitation bias which varies both between subcatchments and between time periods. For four of the subcatchments (Tosoi, Donguztoo, Salamalik and Cholma) the bias varies from precipitation underestimation in the early 1960s to overestimation in the 1980s (Fig. 10, and more clearly visible for one subcatchment in Fig. 11), while there is a clear overestimation over all time periods in Gulcha and a clear underestimation over all time periods in Ak-Tash. The decrease of the bias factor indicates an artificial trend in the WRF downscaled ERA-40 precipitation data that is not consistent with the observed discharge data. Such trends in reanalysis data can result from changes in the observing system (Bengtsson et al., 2004). As the downscaled reanalysis data are the only data showing this behaviour and it is known that ERA-40 data have problems with trends, in this case, the problem with the trend in the downscaled ERA-40 data could also have
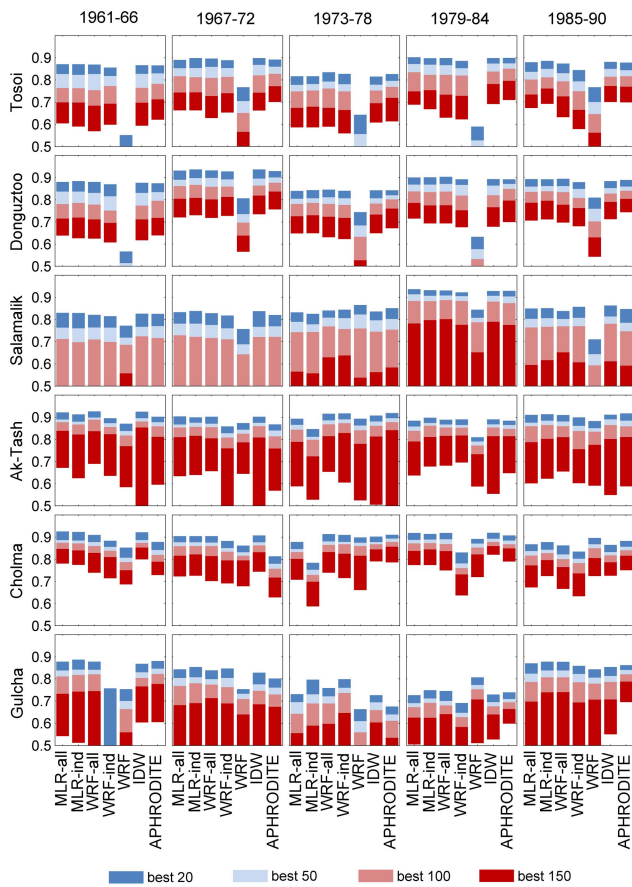
Fig. 9. Variation of the objective function values for different precipitation data sets (bars in each plot), different time periods (columns), and different subcatchments (rows). The colours indicate the range of objective function values for the best 20, 50, 100 and 150 parameter sets.
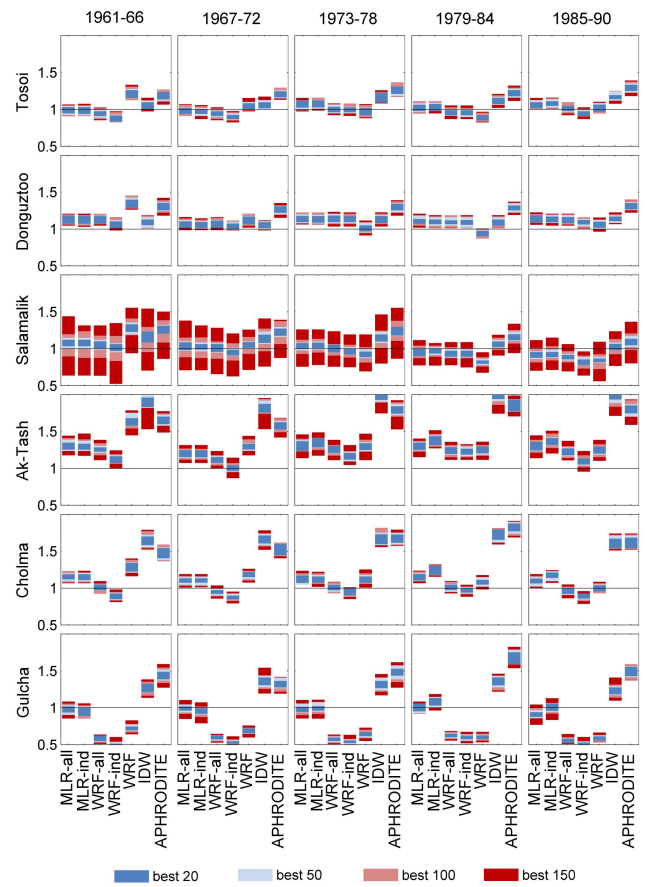

Fig. 10. Variation of the precipitation bias factor for different precipitation data sets (bars in each plot), different time periods (columns), and different subcatchments (rows). The colours indicate the range of the precipitation bias factor for the best 20, 50, 100 and 150 parameter sets.

been detected by simply comparing trends of the precipitation data. However, if two comparable precipitation data sets show different trends, a simulation approach is required to show for which or whether for both precipitation data sets the precipitation bias varies over time, indicating that the trend in the precipitation data is not consistent with the trend in the discharge data. The overestimation for the subcatchment Gulcha is likely to be at least partly caused by a too coarse topography in the WRF model, which results in the valley and mountain ridges to the west of this subcatchment being not well resolved (see Fig. 1).

Using spatial maps of WRF precipitation for the interpolation of gauge observations (WRFadj-all and WRFadj-ind) results in relatively low over- and underestimations for Tosoi, Donguztoo, Salamalik and Cholma, an underestimation of up to 26 % in Ak-Tash, and a stronger overestimation of 36 to 50 % in Gulcha. WRFadj-ind and WRFadj-all result in much less variation of the precipitation factor between time periods than the direct use of the WRF precipitation. However, the overestimation in the southern part of the catchment re-

mains. Due to the lack of precipitation gauges in this part of the catchment, this overestimation cannot be corrected by the combination of the spatial precipitation fields of WRF with observed precipitation time series. The variability of the bias factor between time periods and between subcatchments is similar for the two precipitation estimates WRFadj-all and WRFadj-ind. However, due to the lower objective function values of WRFadj-ind for the period 1979–1984 in Cholma and for the periods 1961–1966 and 1979–1984 in Gulcha, WRFadj-all should be preferred to WRFadj-ind.

Thus, in all subcatchments except Gulcha we find WRFadj and MLR as the most suitable methods. Tobin et al. (2011), who estimated areal precipitation for two Alpine catchments in Switzerland, found that kriging with elevation as external drift variable outperformed kriging with event accumulated precipitation from the COSMO7 downscaled reanalysis data. Due to the differences in the methods and study area, this study is not directly comparable to our study. However, one possible reason for the comparable performance of an interpolation method using downscaled reanalysis data compared
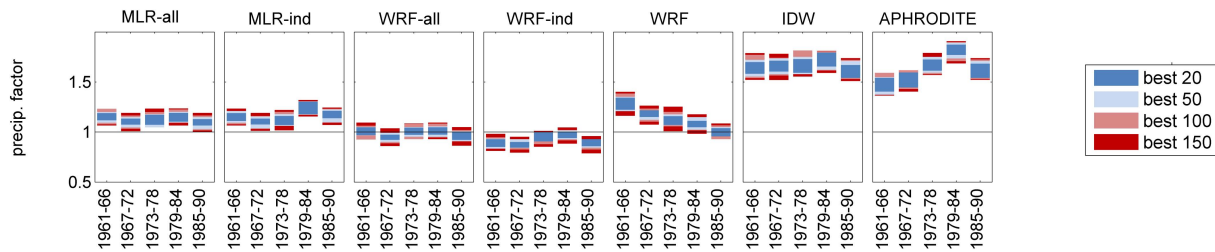
Fig. 11. Variation of the precipitation bias factor by time period for the subcatchment Cholma as an example. The data are the same as also shown in Fig. 10, but they are sorted in a different way in order to better demonstrate the variation by time period (particularly noticeable for the WRF and APHRODITE precipitation data sets).

to other interpolation methods in our study may be the fact that for the basin in Switzerland a larger number of stations was available, which probably allowed a better identification of the observed variability and the precipitation elevation relationship from the data, while methods using simulated precipitation fields from reanalysis data are particularly advantageous in situations where the variability of precipitation cannot be derived from the observed data.

The precipitation estimate based on interpolation of the observed data by IDW results in an underestimation of precipitation in all subcatchments. There is a strong variation of the bias factor between subcatchments, with values indicating around 10 % underestimation in the subcatchments in the northern part (Tosoi, Donguztoo and Salamalik), around 95 and 65 % in Ak-Tash and Cholma located in the west, and about 35 % in Gulcha located in the south of the Karadarya catchment. However, the variation of the precipitation bias factor between the different time periods remains low. The higher bias factors of the subcatchments Ak-Tash, Cholma and Gulcha are probably caused by the fact that the precipitation gauges in this part of the catchment are located in less exposed positions (e.g. Kyzyl-Jar is located within a valley, and Sary-Tash, despite being located at a high elevation, is sheltered by higher mountains) so that the precipitation amount at these stations is not representative for the catchment precipitation. In contrast, there is at least one precipitation gauge at an exposed mountain position in the northern part of the catchment (Chaar-Tash), and thus the measured precipitation amount is more likely to be representative for the areal catchment precipitation in Tosoi, Donguztoo and Salamalik.

The APHRODITE data also underestimate precipitation in all subcatchments. To some extent (probably around 10 %), the underestimation can be explained by the fact that for the APHRODITE data gauge observations are not corrected for undercatch errors before interpolation. In Ak-Tash, Gulcha and Cholma there is a relatively strong variation of the bias factor between the different time periods. This seems not to be due to a change in the number of stations used for the generation of the data set, as with respect to this region and time period the number of stations remains relatively constant.

4.2.4 Sensitivity of results with respect to uncertainties in inputs

In order to check how robust the results are with regard to changes in the inputs (see Sect. 3.3.5), a sensitivity analysis is performed. Varying these inputs and re-calibrating the model has hardly any influence on the objective function values. This shows that the parameters can compensate for input errors. Changes in the precipitation bias factor are shown in Fig. 12. The boxplots summarise the changes in the precipitation bias factor for the seven precipitation data sets and six subcatchments. An increase in the precipitation bias factor of 0.1 in Fig. 12 would for example indicate that a precipitation bias factor of 1.1 would change to 1.2, meaning that the respective precipitation data set underestimates precipitation by 20 % and not by 10 %. The largest uncertainties result from radiation, soil depth, root depth and wind speed. For these inputs the median changes in the precipitation bias factor are between ± 0.03 and ± 0.07, but changes can be up to 0.2 for individual precipitation data sets and subcatchments. Changes in temperature, plant height and stomata resistance have a lower influence with median values of about ± 0.03. Effects of changes in the matrix potential below which transpiration is reduced and in the matrix potential below which transpiration ceases are negligible.

In summary, uncertainties of the precipitation bias factor due to uncertainties in inputs to the evapotranspiration module are less than 0.2 and in most cases even less than 0.1. The combined uncertainties might be higher and different from simply additive – for example the effect of an increase in root depth would be higher if the soil depth was increased at the same time. However, it is unlikely that the default estimates of all of the analysed climate, soil and plant inputs are biased in a way that they would result in the same direction of change of the precipitation factor.

5 Conclusions

This study indicates that spatial fields from downscaled reanalysis data can provide useful information for the spatial interpolation of precipitation data in regions where the spatial
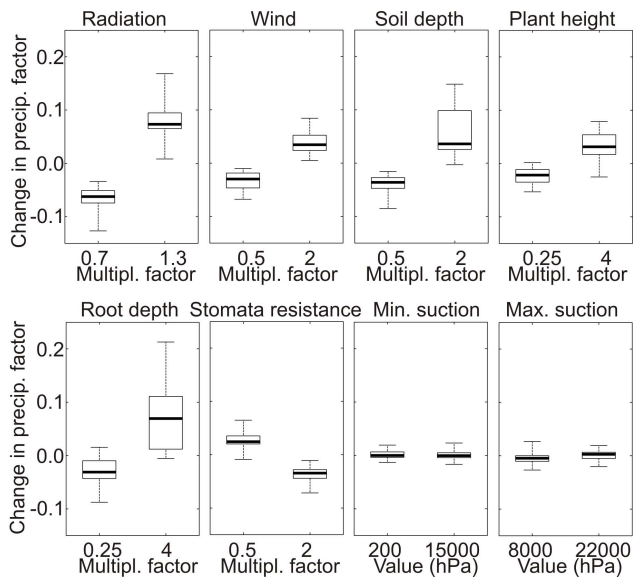
Fig. 12. Sensitivity analysis of the change in the calibrated precipitation bias factor as a result of changes in inputs for the time period 1979–1984. The boxplots show summaries of the results averaged over the six subcatchments and seven precipitation data sets with the thick black line indicating the median, the boxplot area the interquartile range and the whiskers the minimum and maximum change.

variability of precipitation cannot be derived from ground-based observations alone. The method depends on the assumption that the spatial variability is in general correctly represented in the downscaled reanalysis data. While this assumption cannot be fully validated, plausibility tests, like (1) inspecting the simulated precipitation fields for any conspicuous features, (2) checking that the major orographic characteristics of the region are also captured by the model orography and (3) the comparison of simulated and observed precipitation data at locations of available stations, were generally successful for the Karadarya catchment. In the southern part of the catchment, simulated precipitation tends to be overestimated due to mountain ridges to the west of this area that are not represented in the model orography. Compared to the direct use of the WRF downscaled ERA-40 data, hydrological modelling demonstrates a clearly better performance of the precipitation data set WRFadj, both with respect to the goodness of fit and with respect to the stability of the precipitation bias factor over different time periods. The evaluation by hydrological modelling further shows that, except for the subcatchment Gulcha, the method using monthly fields from WRF performs equally well as the best performing method for this area based on monthly fields from multi-linear regression. Additionally, cross validation points out that the method WRFadj behaves more robust against omitting individual stations than the interpolation method based

on monthly fields from multi-linear regression, suggesting that it is particularly suited for data sparse regions.

Using a calibrated precipitation bias factor as an additional performance criterion for the evaluation of precipitation estimates by hydrological modelling allows a more informed differentiation between the precipitation data sets. For our case study, it is for example shown that the main difference between the precipitation data sets based on interpolated station data is in the bias values, while the performance with respect to the time course is rather similar. The evaluation approach was further extended by an assessment of uncertainties resulting from the calibration parameters, from uncertainties in the model inputs, and from different calibration periods. Uncertainties in the calibrated bias factor resulting from parameter uncertainties and from model inputs are not very large and on average both in the order of 0.1, corresponding to a precipitation bias of 10 %. Thus, these uncertainties are often smaller than the differences between different precipitation estimates. The evaluation of the precipitation bias factor for different calibration periods revealed a variation of this factor between time periods for two precipitation data sets, the WRF downscaled ERA-40 data and the APHRODITE data. Ideally, the precipitation input to a hydrological model should have zero bias, but a bias which is largely constant over time could usually be handled for most applications. A variation of the bias factor over time could indicate inconsistencies in gridded precipitation data sets (Mizukami and Smith, 2012). It shows that with these precipitation inputs the observed variability can only be captured by adjusting the precipitation bias factor. The fact that such a variation of the bias factor over time is not necessary for the other precipitation estimates shows that this is caused by the precipitation input and not for example by changes in the catchment or deficits of the model. Currently the bias factor represents a mean value over a calibration period. Future work should also investigate whether variations of this bias factor within this period, for example a seasonal variation, can be identified.

With respect to the headwater catchments of the Karadarya Basin, the different precipitation data sets show very large differences for subcatchment mean precipitation. Based on our evaluation, the precipitation data set MLR-all, which uses monthly fields from multi-linear regression, is judged as the most suitable precipitation input for the studied headwater subcatchments of the Karadarya catchment. It shows good performance with respect to the objective function values – in common with all precipitation data sets based on interpolated station data – low bias and only very small variations of the bias factor between different time periods. Our estimates of the precipitation input to these mountain subcatchments are considerably higher than those from continental- or global-scale gridded data sets. This demonstrates the large uncertainties in these data if they are applied to small to mesoscale mountain catchments. This also has implications for the use of these data for the evaluation or bias correction of regional

climate models applied for climate impact studies. If the focus is on areas with sparsely gauged mountain regions, all precipitation estimates based on gauge observations are afflicted with large uncertainties and an evaluation of the precipitation using hydrological modelling and observed runoff might provide more reliable information.

When evaluating different precipitation estimates by hydrological modelling it should be considered that the approach depends on several assumptions and that it is particularly suited for data sparse regions with large differences in the precipitation estimates. The approach firstly requires valid discharge measurements and assumes that measurement errors, subsurface inflows or outflows and unknown abstractions or flow diversions are only small. Further uncertainties arise from the hydrological modelling, for example from uncertainties in the model inputs and the choice of model structure, particularly those inputs and model components which also influence the water balance. Uncertainties in model structure could be assessed by using an ensemble of hydrological models, which was however beyond the scope of this study. The effect of uncertainties of individual model inputs on the precipitation bias factor in this study is shown to be in the order of up to 10 %; but the uncertainties may be higher when considering combined uncertainties and also taking into account different model structures. In regions where the differences in the precipitation estimates are only small, an evaluation by hydrological modelling may therefore not lead to conclusive results. However, in many data sparse regions, like the catchments investigated in this study, the uncertainties resulting from the hydrological modelling can be assumed to be smaller than the differences between the precipitation data sets, making this approach very well suited for such regions.

References

Academy of Science of the Kyrgyz SSR: Atlas of the Kyrgyz Soviet Socialistic Republic. Volume 1: Natural conditions and resources, State Agency for Cartography and Geodesy, Moscow, 1987.

Artan, G., Gadain, H., Smith, J. L., Asante, K., Bandaragoda, C. J., and Verdin, J. P.: Adequacy of satellite derived rainfall data for stream flow modeling, Nat. Hazards., 43, 167–185, doi:10.1007/s11069-007-9121-6, 2007.

Backhaus, K., Erichson, B., Plinke, W., and Weiber, R.: Multivariate Analysemethoden: Eine anwendungsorientierte Einführung, Springer, 2003.

Basist, A., Bell, G. D., and Meentemeyer, V.: Statistical relationships between topography and precipitation patterns, J. Climate, 7, 1305–1315, doi:10.1175/1520-0442(1994)007<1305:srbtap>2.0.co;2, 1994.

Behrangi, A., Khakbaz, B., Jaw, T. C., AghaKouchak, A., Hsu, K. L., and Sorooshian, S.: Hydrologic evaluation of satellite precipitation products over a mid-size basin, J. Hydrol., 397, 225–237, doi:10.1016/j.jhydrol.2010.11.043, 2011.

Bengtsson, L., Hodges, K. I., and Hagemann, S.: Sensitivity of the ERA-40 reanalysis to the observing system: determination of the global atmospheric circulation from reduced observations, Tellus A, 56, 456–471, doi:10.1111/j.1600-0870.2004.00079.x, 2004.

Beven, K. J. and Binley, A. M.: The future of distributed models: model calibration and uncertainty prediction, Hydrol. Process., 6, 279–298, 1992.

Bitew, M. M. and Gebremichael, M.: Assessment of satellite rainfall products for streamflow simulation in medium watersheds of the Ethiopian highlands, Hydrol. Earth Syst. Sci., 15, 1147–1155, doi:10.5194/hess-15-1147-2011, 2011a.

Bitew, M. M. and Gebremichael, M.: Evaluation of satellite rainfall products through hydrologic simulation in a fully distributed hydrologic model, Water Resour. Res., 47, W06526, doi:10.1029/2010wr009917, 2011b.

Brown, D. P. and Comrie, A. C.: Spatial modeling of winter temperature and precipitation in Arizona and New Mexico, USA, Clim. Res., 22, 115–128, doi:10.3354/cr022115, 2002.

Cheval, S., Baciu, M., and Breza, T.: An investigation into the precipitation conditions in Romania using a GIS-based method, Theor. Appl. Climatol., 76, 77–88, doi:10.1007/s00704-003-0004-9, 2003.

Daly, C., Neilson, R. P., and Phillips, D. L.: A statistical topographic model for mapping climatological precipitation over mountainous terrain, J. Appl. Meteorol., 33, 140–158, doi:10.1175/1520-0450(1994)033<0140:astmfm>2.0.co;2, 1994.

Feddes, R. A. and Raats, P. A. C.: Parameterizing the soil-water-plant root system, Kluwer Academic Publishers, available at: http://edepot.wur.nl/35358 (last access: 17 September 2012), 2004.

Francke, T., Güntner, A., Mamede, G., Müller, E. N., and Bronstert, A.: Automated catena-based discretization of landscapes for the derivation of hydrological modelling units, Int. J. Geogr. Inf. Sci., 22, 111–132, doi:10.1080/13658810701300873, 2008.

Frei, C. and Schär, C.: A precipitation climatology of the Alps from high-resolution rain-gauge observations, Int. J. Climatol., 18, 873–900, doi:10.1002/(sici)1097-0088(19980630)18:8<873::aid-joc255>3.0.co;2-9, 1998.

Frei, C., Christensen, J. H., Déqué, M., Jacob, D., Jones, R. G., and Vidale, P. L.: Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps, J. Geophys. Res.-Atmos., 108, 4124, doi:10.1029/2002JD002287, 2003.

Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., Woodcock, C. E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., and Schaaf, C.: Global land cover mapping from MODIS: algorithms and early results, Remote Sens. Environ., 83, 287–302, doi:10.1016/s0034-4257(02)00078-0, 2002.

Garen, D. C., Johnson, G. L., and Hanson, C. L.: Mean areal precipitation for daily hydrologic modeling in mountainous regions, Water Resour. Bull., 30, 481–491, 1994.

Goodale, C. L., Aber, J. D., and Ollinger, S. V.: Mapping monthly precipitation, temperature, and solar radiation for Ireland with polynomial regression and a digital elevation model, Clim. Res., 10, 35–49, doi:10.3354/cr010035, 1998.

Goovaerts, P.: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, J. Hydrol., 228, 113–129, doi:10.1016/s0022-1694(00)00144-x, 2000.

Gourley, J. J. and Vieux, B. E.: A method for evaluating the accuracy of quantitative precipitation estimates from a hydrologic modeling perspective, J. Hydrometeorol., 6, 115–133, doi:10.1175/jhm408.1, 2005.

Green, W. and Ampt, G.: Studies on soil physics Part I – The flow of air and water through soils, J. Agr. Sci., 4, 1–24, 1911.

Güntner, A.: Large-scale hydrological modelling in the semi-arid North-East of Brazil, Potsdam Institute for Climate Impact Research, Potsdam, Germany, 2002.

Güntner, A. and Bronstert, A.: Representation of landscape variability and lateral redistribution processes for large-scale hydrological modelling in semi-arid areas, J. Hydrol., 297, 136–161, doi:10.1016/j.jhydrol.2004.04.008, 2004.

Haberlandt, U. and Kite, G. W.: Estimation of daily space-time precipitation series for macroscale hydrological modelling, Hydrol. Process., 12, 1419–1432, doi:10.1002/(sici)1099-1085(199807)12:9<1419::aid-hyp645>3.0.co;2-a, 1998.

Haberlandt, U., Schumann, A., and Büttner, U.: Räumliche Niederschlagsschätzung aus Punktmessungen und Radar am Beispiel des Elbehochwassers 2002, Hydrologie und Wasserbewirtschaftung, 49, 56–68, 2005.

Hay, L., Viger, R., and Mccabe, G.: Precipitation interpolation in mountainous regions using multiple linear regression, Hydrology, Water Resources and Ecology in Headwaters, 33–38, 1998.

Heistermann, M. and Kneis, D.: Benchmarking quantitative precipitation estimation by conceptual rainfall-runoff modeling, Water Resour. Res., 47, W06514, doi:10.1029/2010wr009153, 2011.

Hevesi, J. A., Istok, J. D., and Flint, A. L.: Precipitation estimation in mountainous terrain using multivariate geostatistics: I Structural analysis, J. Appl. Meteorol., 31, 661–676, doi:10.1175/1520-0450(1992)031<0661:peimtu>2.0.co;2, 1992.

Houze, R. A. J.: Orographic precipitation, in: Cloud Dynamics, edited by: Houze, R. A. J., Academic Press, San Diego, USA, 1993.

Hurkmans, R., De Moel, H., Aerts, J., and Troch, P. A.: Water balance versus land surface model in the simulation of Rhine River discharges, Water Resour. Res., 44, W01418, doi:10.1029/2007wr006168, 2008.

Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, E.: Holefilled seamless SRTM data, 4th Edn., International Centre for Tropical Agriculture (CIAT), 2008.

Johansson, B. and Chen, D. L.: The influence of wind and topography on precipitation distribution in Sweden: Statistical analysis and modelling, Int. J. Climatol., 23, 1523–1535, doi:10.1002/joc.951, 2003.

Johansson, B. and Chen, D. L.: Estimation of areal precipitation for runoff modelling using wind data: a case study in Sweden, Climate. Res., 29, 53–61, doi:10.3354/cr029053, 2005.

Körner, C.: Leaf diffusive conductances in the major vegetation types of the globe, in: Ecophysiology of Photosynthesis, edited by: Schulze, E.-D. and Caldwell, M. M., Ecological Studies, Springer, Berlin, 463–490, 1994.

Legates, D. R. and Willmott, C. J.: Mean seasonal and spatial variability in gauge-corrected, global precipitation, Int. J. Climatol., 10, 111–127, doi:10.1002/joc.3370100202, 1990.

Lloyd, C. D.: Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain, J. Hydrol., 308, 128–150, doi:10.1016/j.jhydrol.2004.10.026, 2005.

Ly, S., Charles, C., and Degré, 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., 15, 2259–2274, doi:10.5194/hess-15-2259-2011, 2011.

Marquinez, J., Lastra, J., and Garcia, P.: Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis, J. Hydrol., 270, 1–11, doi:10.1016/s0022-1694(02)00110-5, 2003.

Martinez-Cob, A.: Multivariate geostatistical analysis of evapotranspiration and precipitation in mountainous terrain, J. Hydrol., 174, 19–35, doi:10.1016/0022-1694(95)02755-6, 1996.

Mitchell, T. D. and Jones, P. D.: An improved method of constructing a database of monthly climate observations and associated high-resolution grids, Int. J. Climatol., 25, 693–712, doi:10.1002/joc.1181, 2005.

Mizukami, N. and Smith, M. B.: Analysis of inconsistencies in multi-year gridded quantitative precipitation estimate over complex terrain and its impact on hydrologic modeling, J. Hydrol., 428, 129–141, doi:10.1016/j.jhydrol.2012.01.030, 2012.

Monteith, J. L.: Evaporation and environment, Symo. Soc. Exp. Biology, 19, 205–234, 1965.

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M., Morisette, J. T., Votava, P., Nemani, R. R., and Running, S. W.: Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data, Remote Sens. Environ., 83, 214–231, 2002.

NCDC: Daily and sub-daily precipitation for the former USSR (DSI-9813) 1.0st Edn., National Climatic Data Center NESDIS NOAA US Department of Commerce, 2005.

Ninyerola, M., Pons, X., and Roure, J. M.: A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques, Int. J. Climatol., 20, 1823–1841, doi:10.1002/1097-0088(20001130)20:14<1823::aid-joc566>3.0.co;2-b, 2000.

Ninyerola, M., Pons, X., and Roure, J. M.: Monthly precipitation mapping of the Iberian Peninsula using spatial interpolation tools

implemented in a Geographic Information System, Theor. Appl. Climatol., 89, 195–209, doi:10.1007/s00704-006-0264-2, 2007.

Perry, M. and Hollis, D.: The generation of monthly gridded datasets for a range of climatic variables over the UK, Int. J. Climatol., 25, 1041–1054, doi:10.1002/joc.1161, 2005.

Phillips, D. L., Dolph, J., and Marks, D.: A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain, Agr. Forest Meteorol., 58, 119–141, doi:10.1016/0168-1923(92)90114-j, 1992.

Prudhomme, C. and Reed, D. W.: Relationships between extreme daily precipitation and topography in a mountainous region: A case study in Scotland, Int. J. Climatol., 18, 1439–1453, doi:10.1002/(sici)1097-0088(19981115)18:13<1439::aid-joc320>3.0.co;2-7, 1998.

Schär, C., Vasilina, L., Pertziger, F., and Dirren, S.: Seasonal runoff forecasting using precipitation from meteorological data assimilation systems, J. Hydrometeorol., 5, 959–973, 2004.

Schiemann, R., Luthi, D., Vidale, P. L., and Schär, C.: The precipitation climate of Central Asia – intercomparison of observational and numerical data sources in a remote semiarid region, Int. J. Climatol., 28, 295–314, doi:10.1002/joc.1532, 2008.

Schneider, U., Becker, A., Meyer-Christoffer, A., Ziese, M., and Rudolf, B.: Global precipitation analysis products of the GPCC, Deutscher Wetterdienst, Offenbach a. M., Germany, 2011.

Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data, Proceedings of the 1968 ACM National Conference, 517–524, 1968.

Shuttleworth, W. J. and Wallace, J. S.: Evaporation from sparse crops – an energy combination theory, Q. J. Roy. Meteor. Soc., 111, 839–855, doi:10.1256/smsqj.46909, 1985.

Siegfried, T. and Bernauer, T.: Estimating the performance of international regulatory regimes: Methodology and empirical application to international water management in the Naryn/Syr Darya basin, Water Resour. Res., 43, W11406, doi:10.1029/2006wr005738, 2007.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X., Wang, W., and Powers, J. G.: A description of the advanced research WRF Version 3, Technical Report, 124 pp., 2008.

Stisen, S. and Sandholt, I.: Evaluation of remote-sensing-based rainfall products through predictive capability in hydrological runoff modelling, Hydrol. Process., 24, 879–891, doi:10.1002/hyp.7529, 2010.

Suklitsch, M., Gobiet, A., Truhetz, H., Awan, N. K., Gottel, H., and Jacob, D.: Error characteristics of high resolution regional climate models over the Alpine area, Clim. Dynam., 37, 377–390, doi:10.1007/s00382-010-0848-5, 2011.

Sun, R. H., Zhang, B. P., and Tan, J.: A multivariate regression model for predicting precipitation in the Daqing Mountains, Mt. Res. Dev., 28, 318–325, doi:10.1659/mrd.0944, 2008.

Tobin, C., Nicotina, L., Parlange, M. B., Berne, A., and Rinaldo, A.: Improved interpolation of meteorological forcings for hydrologic applications in a Swiss Alpine region, J. Hydrol., 401, 77–89, doi:10.1016/j.jhydrol.2011.02.010, 2011.

Tolson, B. A. and Shoemaker, C. A.: Efficient prediction uncertainty approximation in the calibration of environmental simulation models, Water Resour. Res., 44, W04411, doi:10.1029/2007wr005869, 2008.

Uppala, S. M., Kallberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Van De Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Holm, E., Hoskins, B. J., Isaksen, L., Janssen, P., Jenne, R., McNally, A. P., Mahfouf, J. F., Morcrette, J. J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Q. J. Roy. Meteor. Soc., 131, 2961–3012, doi:10.1256/qj.04.176, 2005.

van der Linden, P. and Mitchell, J. F. B.: ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project, 160 pp., 2009.

Yang, D. Q., Goodison, B. E., Metcalfe, J. R., Golubev, V. S., Elomaa, E., Gunther, T., Bates, R., Pangburn, T., Hanson, C. L., Emerson, D., Copaciu, V., and Miklovic, J.: Accuracy of Tretyakov precipitation gauge: Result of WMO intercomparison, Hydrol. Process., 9, 877–895, doi:10.1002/hyp.3360090805, 1995.

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutom, N., and Kitoh, A.: APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges, B. Am. Meteorol. Soc., 93, 1401–1415, doi:10.1175/BAMS-D-11-00122.1, 2012.

Yilmaz, K. K., Hogue, T. S., Hsu, K. L., Sorooshian, S., Gupta, H. V., and Wagener, T.: Intercomparison of rain gauge, radar, and satellite-based precipitation estimates with emphasis on hydrologic forecasting, J. Hydrometeorol., 6, 497–517, doi:10.1175/jhm431.1, 2005.