Hydrological modeling of mountain catchments in Central Asia— approaches for data sparse catchments

Doris Düthmann

Dissertation submitted for the degree of Doctor of Natural Sciences (Dr. rer. nat.) in Hydrology
to the Faculty of Mathematics and Natural Sciences at the University of Potsdam, Germany

Referees

Prof. Dr. Bruno Merz, University of Potsdam
Prof. Dr. Axel Bronstert, University of Potsdam
Prof. Dr. Jan Seibert, University of Zurich

Submitted January 2015

Defended June 2015

Published August 2015

Published online at the
Institutional Repository of the University of Potsdam:
URN urn:nbn:de:kobv:517-opus4-80071
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-80071

Contents

Contents ..... **V**

List of figures ..... **VII**

List of tables ..... **XI**

Abstract ..... **XIII**

Zusammenfassung ..... **XV**

1 Introduction **1**

1.1 Motivation 1

1.2 Water in Central Asia—a scarce resource 1

1.3 Additional data sources and approaches for hydrological modeling in data sparse mountain regions 2

1.4 Understanding past hydrologic changes 3

1.5 Objectives and research questions 3

1.6 Study areas 5

1.7 Overview 6

2 Evaluation of areal precipitation estimates based on downscaled reanalysis and station data by hydrological modeling **7**

2.1 Introduction 8

2.2 Study area 9

2.3 Data and methods 10

2.3.1 Precipitation data and interpolation approaches 10

2.3.2 Point based evaluation of the precipitation data 12

2.3.3 Evaluation of areal precipitation estimates based on simulated discharge 13

2.4 Results and discussion 16

2.4.1 Characteristics of the precipitation data sets 16

2.4.2 Evaluation of the precipitation data based on simulated discharge 19

2.5 Conclusions 23

3 The value of satellite-derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia **27**

3.1 Introduction 28

3.2 Study area 29

3.3 Data and methods 30

3.3.1 Extraction of snow cover from AVHRR satellite imagery 30

3.3.2 Hydrological model 31

3.3.3 Model calibration 32

3.4 Results and discussion 35

3.4.1 Characterization of the trade-offs between good performance for discharge and snow cover area 35

3.4.2 Model performance in terms of discharge and snow cover area 37
3.4.3 Influence of the two objective functions on constraining model parameters 39
3.4.4 Value of increasing the number of snow cover observations during the calibration period 42
3.5 Conclusions 43

4 Attribution of streamflow trends in snow- and glacier melt dominated catchments of the Tarim River, Central Asia 47
4.1 Introduction 48
4.2 Study area and observed hydrometeorological changes 49
4.3 Data and methods 51
4.3.1 Streamflow and meteorological data 51
4.3.2 Trend analysis 52
4.3.3 Data-based analysis 52
4.3.4 Simulation-based analysis 53
4.4 Results 56
4.4.1 Data-based analysis 56
4.4.2 Simulation-based analysis 59
4.5 Discussion 67
4.5.1 Glacier melt contribution and trend attribution for two headwater catchments of the Aksu River 67
4.5.2 Data-based versus simulation-based trend attribution 67
4.5.3 Multiobjective model calibration 68
4.5.4 Limitations 68
4.6 Conclusions 69

5 Discussion and conclusions 71
5.1 Main results 71
5.2 Discussion and future research questions 72
5.2.1 Interpolation of station-based precipitation time series with precipitation fields from downscaled reanalysis data 72
5.2.2 Evaluating areal precipitation estimates by hydrological modeling 73
5.2.3 Satellite-derived snow cover for hydrological model calibration 73
5.2.4 Further data for multiobjective model calibration 75
5.2.5 Attribution of streamflow trends in two high mountain catchments using a data-based and a simulation-based approach 77
5.3 Concluding remarks 79

References 81

Acknowledgements 93

Author's declaration 95

List of figures

Fig. 1.1: Location of the study regions within the Aral Sea Basin and the Tarim River Basin. 1

Fig. 2.1: 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). 10

Fig. 2.2: 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). 17

Fig. 2.3: 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. 18

Fig. 2.4: 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. 18

Fig. 2.5: Monthly subcatchment mean precipitation (1960–1990) for six subcatchments of the Karadarya Basin and seven different precipitation estimates. 19

Fig. 2.6: 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. 19

Fig. 2.7: Mean annual precipitation (1960–1990) for six subcatchments of the Karadarya Basin for the precipitation data sets MLR-all, APHRODITE, and four global precipitation data sets. 20

Fig. 2.8: 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. 21

Fig. 2.9: Variation of the objective function values for different precipitation data sets (bars in each plot), different time periods (columns), and different subcatchments (rows). The colors indicate the range of objective function values for the best 20, 50, 100, and 150 parameter sets. 22

Fig. 2.10: Variation of the precipitation bias factor for different precipitation data sets (bars in each plot), different time periods (columns), and different subcatchments (rows). The colors indicate the range of the precipitation bias factor for the best 20, 50, 100, and 150 parameter sets. 22

Fig. 2.11: 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. 2.10, 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). 23

Fig. 2.12: 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. 23

Fig. 3.1: The six study catchments in the Karadarya Basin, including elevation, discharge gauges, and precipitation gauges in the area. The inlet shows the location of the upper Karadarya Basin in Kyrgyzstan. 29

Fig. 3.2: Distribution of the catchment area with altitude for the six study catchments (by 200 m elevation zones). 30

Fig. 3.3: Procedure for calculating the minimum, median, and maximum snow cover error during the validation period, when using *n* snow cover images for model calibration. 34

Fig. 3.4: Trade-off curves of model performance against SCA (*objf_sca*) and discharge (*objf_q*) for the six study catchments. The *x* axis is plotted in reverse order so that optimum solutions plot in the lower left corner. Dots: calibration period; crosses: validation period; blue: simulation period A (1978–1981 and 1986–1987); red: simulation period B (1982–1985 and 1988–1989). For example, blue dots indicate the model performance in simulation period A of the solutions from the calibration in simulation period A, and blue crosses show the model performance in

simulation period A of solutions from the calibration in simulation period B. Additionally, the stars show the performance of the uncalibrated model. Please note the different scales of the *x* axes. 35

Fig. 3.5: Scatterplots of the objective function with respect to SCA (*objf_sca*) against the objective function with respect to discharge (*objf_q*) showing both the solutions generated by Monte Carlo sampling (blue) and the solutions from the multiobjective optimization algorithm (red). The *x* axis is plotted in reverse order so that optimum solutions plot in the lower left corner. a) Calibration in simulation period A (1978–1981 and 1986–1987), and b) calibration in simulation period B (1982–1985 and 1988–1989). 36

Fig. 3.6: Time series of simulated versus observed SCA during the validation period, for selected catchments and elevation zones. The simulations show a compromise solution in terms of snow cover and discharge performance from the multiobjective optimization, and solutions from the Monte Carlo sample with comparable performance in terms of discharge but contrasting performance in terms of SCA. 37

Fig. 3.7: Simulated and observed SCA in the six study catchments, for the 28 April 1986, 11 May 1987, and 13 April 1988. Simulated snow cover is shown for the validation period and a selected compromise solution generated with the multiobjective optimization algorithm. The simulations also refer to discrete 200 m elevation zones but are drawn as continuous lines for better visibility. 38

Fig. 3.8: SCA bias by elevation zones averaged over the validation periods (top) 1986–1987 and (bottom) 1988–1989. The black dots show results for a compromise solution in terms of snow cover and discharge performance from the multiobjective optimization. Bars indicate ranges of values from selected Monte Carlo simulations with comparable performance in terms of discharge but contrasting performance in terms of SCA. 39

Fig. 3.9: Time series of simulated versus observed discharge for a compromise solution (good performance for discharge and SCA) from the multiobjective optimization algorithm and the time period 1986–1989. Gray area: observed discharge; red line: simulated discharge for model calibration in simulation period A (1978–1981 and 1986–1987); blue line: simulated discharge for model calibration in simulation period B (1982–1985 and 1988–1989). 40

Fig. 3.10: Observed discharge and simulated discharge, for the catchment Ak-Tash and the year 1987. Simulated discharge is shown for selected parameter sets from the Monte Carlo sample with overall comparable performance in terms of discharge, but good and poor performance with respect to SCA. 41

Fig. 3.11: Cumulative distribution functions of parameter sets selected from the Monte Carlo sample as the 5% best performing solutions in terms of *objf_q* (red lines) and after further constraining these solutions selecting the 5% best performing solutions in terms of *objf_sca* (blue lines) for calibration to simulation period A. The six lines indicate the different catchments (for more details please refer to the text). 41

Fig. 3.12: Value of the objective function with respect to SCA (*objf_sca*) in the validation period as a function of the number of snow cover images *n* used in the calibration period. The black line shows the median over 10,000 repetitions of selecting *n* images, the gray lines show the minimum and maximum, respectively. (top row) Results for calibration to simulation period A and validation in simulation period B, and (bottom row) results for the reversed case. 42

Fig. 3.13: Median effect of individual images on *objf_sca* in the validation period when drawn as first (filled dots), second (crosses), or third (gray diamonds) image, for the catchment Salamalik with simulation period A for model calibration. 43

Fig. 3.14: Simulated snow cover ranges of 1000 Monte Carlo simulations of each elevation zone over the time period 1986–1989 for the catchment Cholma (gray-shaded area). Red crosses show the snow cover observations. The blue line shows simulated snow cover for the validation period from a balanced solution of the multiobjective optimization. 44

Fig. 4.1: The study catchments, including elevation, discharge gauges, meteorological stations, and locations of Karabatkak glacier and geodetic mass balance estimates. The inset shows the location of the study area (dark gray outline) in relation to national borders and the Tarim Basin (white outline). 50

Fig. 4.2: Monthly changes in discharge (a–d), in catchment average precipitation (e–h), and in temperature (i–j) over the period 1957–2004 for the Kakshaal and Sari-Djaz catchment.

Changes in discharge and precipitation are reported in absolute (mm) and relative values (%). Bar colors indicate the significance level of the Mann-Kendall trend test. 57

Fig. 4.3: Observed (blue) and modeled (gray) seasonal discharge for the Kakshaal and Sari-Djaz catchment. The modeled discharge is based on a regression model with temperature and precipitation of the current season and of the last year as possible explanatory variables..... 57

Fig. 4.4: Trade-off curves of model performance with respect to daily discharge (*OF_Q1*) against model performance with respect to interannual discharge variations (*OF_Q2*), glacier mass balance time series at Karabatkak glacier (*OF_G1*), and the geodetic glacier mass balance (*OF_G2*) for the two study catchments. The *x* axes are plotted in reverse order so that optimum solutions always plot in the lower left corner. Blue dots show solutions used for further analyses, gray dots show solutions which were excluded due to too low performance with respect to *OF_Q2*, *OF_G2*, or representation of observed trends (see section 4.3.4.4). 59

Fig. 4.5: Simulated and observed daily runoff for the period 1964–1987 for the Kakshaal catchment. 60

Fig. 4.6: Simulated and observed daily runoff for the period 1964–1987 for the Sari-Djaz catchment. Gray triangles mark the dates of reported GLOFs as listed in Glazirin (2010). 61

Fig. 4.7: Interannual variation of the observed glacier mass balance at Karabatkak glacier, and simulated catchment average glacier mass balance for (a) the Kakshaal and (b) the Sari-Djaz catchment. 62

Fig. 4.8: Observed and simulated (a-b) annual, and (c–j) seasonal discharge for the Kakshaal and the Sari-Djaz catchment..... 64

Fig. 4.9: Annual variation of the input from rain, snowmelt, and glacier melt, as represented in the WASA model for the selected model parameterizations for (a) the Kakshaal and (b) the Sari-Djaz catchment. 66

List of tables

Table 2.1: Area, glacierization, elevation range, and mean annual runoff over the period 1961–1990 of the studied subcatchments of the Karadarya Basin. 9
Table 2.2: Overview of the precipitation data sets used in this study. 12
Table 2.3: Calibration parameters including values for the lower and upper bounds. 14
Table 2.4: Climate, plant, and soil inputs selected for the sensitivity analyses. 16
Table 2.5: Comparison of observed and WRF simulated precipitation at the station locations: bias, daily and monthly squared correlation coefficient calculated over the period 1960–1990. 17
Table 3.1: Area, glacier coverage, elevation range, mean annual runoff, and runoff coefficient of the studied subcatchments of the Karadarya Basin. 30
Table 3.2: Calibration parameters including values for the lower and upper bounds. 32
Table 4.1: Area, glacier coverage, elevation range, mean annual runoff, and mean annual precipitation of the two catchments. "Interpolated precip." is the areal precipitation estimate from interpolation of station-based observations, as described in section 4.3.1. The last column shows areal precipitation estimated from hydrological modeling (see section 4.3.4.1 and 4.4.2.1), as mean values and with the ranges over the selected model solutions in brackets. 49
Table 4.2: Spatial data sets used for setting up the WASA model. 54
Table 4.3: Results of the linear regression analysis of discharge: coefficient of determination (R^2), trend in observed discharge (*Qobs*) and corresponding *p*-value, trend in discharge as assessed with the regression model (*Qregr*) and corresponding *p*-value, as well as the included variables (T: temperature, P: precipitation) with standardized coefficients, trend, corresponding *p*-value, and the coefficient multiplied with the trend in the explanatory variable. Trends are reported as cumulative values over the period 1957–2004. 58
Table 4.4: Correlations between explanatory variables (T: temperature, P: precipitation) used in the regression models. * and ** indicate correlations with *p*-values <0.05 and <0.01, respectively. 58
Table 4.5: Range of Nash-Sutcliffe efficiency (*NSE*), Nash-Sutcliffe efficiency calculated for logarithmic discharge values (*logNSE*), and volume bias for daily and monthly discharge during the calibration and validation period of the WASA model. 62
Table 4.6: Coefficient of determination (R^2) between annual and seasonal simulated (*Qsim*) and observed (*Qobs*) discharge, as well as corresponding trends over the period 1957–2004. For model performance and trends of simulated discharges the table gives ranges over the selected model parameterizations. 64
Table 4.7: Trends over the period 1957–2004 and *p*-values of the Mann-Kendall test for time series of annual and seasonal discharge, as simulated by different WASA model versions with and without considering changes in glacier area and elevation (ranges over the selected model parameterizations). The simulations were performed with the original precipitation and temperature time series. 65
Table 4.8: Trends over the period 1957–2004 and *p*-values of the Mann-Kendall test for simulated time series of annual and seasonal discharge, as ranges over the selected model parameterizations. The simulations were performed with different combinations of original (orig.) and no-trend precipitation and temperature time series, using a model which assumes a constant glacier geometry. 65
Table 4.9: Trends over the period 1957–2004 in the simulated inputs to the hydrological system by rain, snowmelt, and glacier melt for the two study catchments, calculated with the Sen's slope estimator and the Mann-Kendall test. 66
Table 5.1: Factors for evaluating the credibility of a trend attribution study. 78

Abstract

In Central Asia, water resources from the mountain regions have a high relevance for the water supply of the water scarce lowlands. A good understanding of the water cycle in these snow and glacier melt dominated mountain regions provides a foundation for developing water management strategies. Hydrological modeling is a tool which helps to improve our knowledge on the regional water cycle. It can also be used to gain a better understanding of past changes, or estimate future hydrologic changes in view of projected changes in climate. However, due to the scarcity of hydrometeorological data, hydrological modeling for mountain regions in Central Asia involves large uncertainties. In this thesis, it was therefore investigated which approaches and additional data can be used to increase the credibility of hydrological models in such data sparse regions. Furthermore, both hydrological modeling and a data-based approach were applied to gain a better understanding of past changes in the hydrological cycle.

In data sparse mountain regions, the generation and evaluation of areal precipitation estimates are important problems for hydrological modeling. In mountain areas, it is usually of advantage to consider the elevation dependence of precipitation for the interpolation of precipitation. Given a small number or unequal distribution of stations, it is, however, often not possible to derive the spatial distribution of precipitation including the relationship to topographic variables from the station data alone. As a possible alternative, the spatial distribution of precipitation may be derived from precipitation patterns simulated by atmospheric models. In this thesis, it was therefore investigated whether monthly precipitation patterns from reanalysis data, downscaled by a regional climate model to a resolution of 12 km, can be useful for the interpolation of station-based precipitation time series. This approach was then compared to other precipitation estimates based on multilinear regression, inverse distance weighting, an existing gridded precipitation data set, and the direct use of the downscaled reanalysis data.

Different methods for precipitation interpolation are typically evaluated with cross-validation, but this method can be misleading if important parts of the catchment are underrepresented by the precipitation gauges. In this study, hydrological modeling and a comparison of simulated and observed discharge was used for the evaluation of different precipitation data sets. This approach for the comparison of precipitation data sets was extended in order to consider parameter uncertainties and to allow a more informed differentiation between deviations with respect to the overall bias and temporal dynamics of the precipitation data.

The precipitation estimate which used precipitation patterns from downscaled reanalysis data for the interpolation of station data performed clearly better than the direct use of the downscaled reanalysis data, and overall showed a high performance in most catchments. Poor performance of this method was only observed in one of six investigated catchments. A possible reason for the poor performance in this catchment might be the still relatively coarse resolution of the orography in the regional climate model. As a result of the differentiated evaluation, it could be shown that the investigated precipitation data sets mostly differed in their overall bias, while the performance with respect to the temporal dynamics was comparable.

In data sparse areas, larger uncertainties in the input data increase the risk that a good discharge simulation is the result of compensating errors. Considering additional data besides discharge data for model calibration can increase the internal consistency of a hydrological model. In this thesis, snow cover data and glacier mass balance data were applied as additional calibration criteria. Satellite-derived snow cover data are also available for remote areas and are well suited for regions where snowmelt is an important runoff generation process, like in many mountain areas. Using a multiobjective genetic algorithm, the trade-off between good model performance in terms of discharge and snow cover was explicitly characterized. Only small trade-offs between good simulations with respect to discharge and snow cover area were observed, meaning that with the same parameterization it was possible to achieve good discharge and good snow cover simulations. High model performance with respect to discharge did, however, not exclude low performance in terms of snow cover area, demonstrating that snow cover data were important for increasing the internal model consistency.

In this context, the number of snow cover scenes required for model calibration was investigated for the first time. The results showed that in this study at least 10–16 snow cover scenes were required for model calibration in order to avoid that the calibration data were dominated by individual snow cover scenes with unrepresentative deviations between model and observations, which result in the selection of unrepresentative parameter sets for the studied catchment.

The analysis of hydrologic changes can give valuable insights into hydrologic systems. In two headwater catchments of the Aksu River, streamflow increased by around 30% over the period 1957–2004. In order to investigate to what extent these

changes may be caused by increases in precipitation or by increases in temperature and higher glacier melt, a trend attribution study was performed. The analysis included a data-driven approach using multilinear regression and a simulation-driven approach based on hydrological modeling. The hydrological model considered changes in glacier area and glacier surface elevation. Within a multiobjective calibration framework, criteria based on daily and interannual variation of streamflow, as well as glacier mass balance data from direct observations and two geodetic estimates were taken into account. The individual contributions to the overall streamflow trends from changes in glacier geometry, temperature, and precipitation were assessed using simulation experiments with a constant glacier geometry, and with temperature and precipitation time series without trends. The results showed that the observed changes in streamflow were consistent with the changes in temperature and precipitation. In the catchment with a stronger glacierization, increasing temperatures were identified as the dominant driver, while in the other catchment precipitation and temperature both contributed to the streamflow increases. Comparing the two approaches, the evidence provided by the simulation-based approach was regarded as stronger than the one provided by the data-based approach. The latter was hampered by correlations between the explanatory variables of the multilinear regression, and the fact that it is based on statistical links instead of cause-effect relationships.

The study areas of this thesis are headwater catchments of the Karadarya Basin, located in Kyrgyzstan, and of the Aksu Basin, located in Kyrgyzstan and China. The six headwater catchments in the Karadarya Basin have a size of 170–3800 km^2. The glacierization of these catchments is low; two of these catchments have no glaciers and in the others glacier cover ranges between 0.5 and 2.3%. Average annual areal precipitation was estimated between 700 and 1200 mm a^{-1}, and average annual runoff ranges between less than 300 and nearly 800 mm a^{-1}. The studied headwater catchments in the Aksu Basin are larger than those in the Karadarya Basin (18,400 and 12,900 km^2), and are characterized by a stronger glacierization of 4.4% and 21%. With estimated average annual precipitation of 370 and 530 mm a^{-1}, the climate in the headwater catchments of the Aksu Basin is generally drier than in the Karadarya Basin. Average annual discharge of the two study catchments is 150 and 380 mm a^{-1}, respectively.

The approaches developed in this thesis improved hydrological modeling for data sparse catchments in mountainous regions in Central Asia by using additional data from remote sensing and atmospheric modeling. As a result a better understanding of the hydrological cycle in these catchments was achieved, including the water balance, hydrological processes, and observed hydrological changes. These results contribute to an improved scientific basis for decisions related to water management and planning. The developed approaches for hydrological modeling may also be transferred to other data sparse mountain regions.

Zusammenfassung

Wasserressourcen aus den Gebirgsregionen Zentralasiens haben eine große Bedeutung für die Wasserversorgung der niederschlagsarmen Tiefebenen. Ein gutes Verständnis des Wasserkreislaufs in diesen durch Schnee- und Gletscherschmelze dominierten Gebirgsregionen ist eine Grundlage für die Entwicklung von Wassermanagementstrategien. Hydrologische Modelle helfen, den regionalen Wasserkreislauf dieser Gebiete besser zu verstehen. Darüber hinaus können sie dazu beitragen, Änderungen des Wasserkreislaufs in der Vergangenheit nachzuvollziehen oder zukünftige hydrologische Änderungen infolge des Klimawandels zu projizieren. Allerdings ist die hydrologische Modellierung in Gebirgsregionen in Zentralasien aufgrund der nur spärlich vorhandenen hydrometeorologischen Daten mit großen Unsicherheiten verbunden. In dieser Arbeit wurde deshalb untersucht, welche Ansätze und zusätzlichen Daten verwendet werden können, um die Glaubwürdigkeit von hydrologischen Modellen in solchen datenarmen Regionen zu verbessern. Des Weiteren wurden hydrologische Modellierung sowie ein datenbasierter Ansatz angewandt, um beobachtete Veränderungen im Wasserkreislauf in der Vergangenheit zu verstehen.

In datenarmen Gebirgsregionen ist die Abschätzung des Gebietsniederschlags ein wichtiges Problem für die hydrologische Modellierung, und auch die Bewertung unterschiedlicher Interpolationsverfahren ist erschwert. Im Allgemeinen sind Interpolationsverfahren, die die Höhenabhängigkeit des Niederschlags berücksichtigen, in Gebirgsregionen von Vorteil. Bei einer geringen Anzahl von Niederschlagsstationen bzw. einer ungünstigen Verteilung ist es allerdings oft nicht möglich, die räumliche Verteilung des Niederschlags einschließlich seiner Beziehung zur Geländehöhe allein aus den Stationsdaten abzuleiten. Eine mögliche Alternative ist es, die räumliche Niederschlagsverteilung aus simulierten Niederschlagsmustern aus der atmosphärischen Modellierung abzuleiten. In dieser Studie wurde daher untersucht, inwieweit monatliche Niederschlagsmuster von Reanalysedaten, die mit einem regionalen Klimamodell auf eine Auflösung von 12 km herunterskaliert wurden, für die Interpolation stationsbasierter Zeitreihen von Nutzen sein können. Dieser Ansatz wurde mit anderen Verfahren für die Niederschlagsschätzung, basierend auf multilinearer Regression, dem Inverse-Distanzen-Verfahren, einem bereits vorhandenen Niederschlagsdatensatz und der direkten Verwendung der herunterskalierten Reanalysedaten, verglichen. Traditionellerweise werden unterschiedliche Interpolationsverfahren über Kreuzvalidierung bewertet. Diese Methode kann allerdings zu falschen Schlussfolgerungen führen, wenn wichtige Teile des Gebietes durch Niederschlagsstationen unterrepräsentiert sind. In dieser Arbeit wurden die verschiedenen Niederschlagsdatensätze daher über hydrologische Modellierung und einen Vergleich von simulierten und beobachteten Abflusszeitreihen bewertet. Dieser Ansatz für den Vergleich verschiedener Niederschlagsdatensätze wurde um die Berücksichtigung von Parameterunsicherheiten und um eine differenziertere Unterscheidung zwischen Abweichungen bezüglich des Niederschlagsvolumens und der zeitlichen Dynamik der Niederschlagsdaten erweitert.

Der Niederschlagsdatensatz, der auf Interpolation von Stationsdaten mithilfe von Niederschlagsmustern aus herunterskalierten Reanalysedaten basierte, war deutlich besser als die direkte Verwendung der herunterskalierten Reanalysedaten und zeigte insgesamt in den meisten Gebieten eine sehr hohe Güte. Eine geringere Güte wurde lediglich in einem von sechs betrachteten Gebieten beobachtet. Ein möglicher Grund für die schlechtere Niederschlagsschätzung in diesem Gebiet wurde in der immer noch recht groben Auflösung der Orographie des regionalen Klimamodells gesehen. Durch die differenziertere Bewertung der Niederschlagsdatensätze konnte gezeigt werden, dass sich die untersuchten Niederschlagsdatensätze insbesondere im Niederschlagsvolumen unterschieden, während ihre Güte bezüglich der zeitlichen Dynamik vergleichbar war.

Aufgrund der in datenarmen Gebieten insgesamt höheren Unsicherheiten besteht eine größere Gefahr, dass das Modell zwar den Abfluss gut simulieren kann, dies aber eine Folge von Fehlerkompensation ist. Die Verwendung zusätzlicher Daten neben Abflussdaten für die Modellkalibrierung kann ein guter Weg sein, um die interne Konsistenz eines hydrologischen Modells zu erhöhen. In dieser Arbeit wurden Schneebedeckungsdaten und Gletschermassenbilanzdaten als zusätzliche Kalibrierkriterien verwendet. Satellitenbasierte Schneebedeckungsdaten sind auch für abgelegene Gebiete verfügbar und eignen sich sehr gut, wenn – wie in vielen Gebirgsregionen – Schneeschmelze ein wichtiger Abflussbildungsprozess ist. Mittels multiobjektiver Kalibrierung wurde der Zielkonflikt zwischen einer guten Simulation des Abflusses und der Schneebedeckung explizit untersucht. Es zeigte sich, dass dieser Zielkonflikt in den meisten Fällen nur klein war, d.h., dass mit der gleichen Parametrisierung eine hohe Modellgüte bezüglich des Abflusses und der Schneebedeckung erreicht werden konnte. Eine hohe Modellgüte bezüglich des Abflusses schloss allerdings nicht unbedingt Simulationen mit einer geringen Modellgüte bezüglich der Schneebedeckung aus. Dadurch konnte demons-

triert werden, dass die Schneebedeckungsdaten entscheidend zu einer Verbesserung der internen Modellkonsistenz beigetragen haben.

In diesem Zusammenhang wurde erstmalig untersucht, wie viele Schneebedeckungsszenen für eine effiziente Modellkalibrierung benötigt werden. Das Ergebnis zeigte, dass in dieser Studie mindestens 10–16 Schneebedeckungsszenen für die Modellkalibrierung erforderlich waren. Bei einer geringeren Anzahl besteht die Gefahr, dass die Kalibrierdaten durch wenige Schneebedeckungsszenen dominiert werden, an denen die Abweichungen zwischen Modell und Beobachtung nicht repräsentativ sind und die daher zu einer Auswahl nicht repräsentativer Parametersätze führen.

Durch die Untersuchung von hydrologischen Veränderungen können wertvolle Erkenntnisse über hydrologische Systeme gewonnen werden. In zwei Teileinzugsgebieten des Aksu-Gebiets wurde über den Zeitraum 1957–2004 eine Zunahme des Abflusses um etwa 30% beobachtet. Um die möglichen Ursachen dieses Abflussanstiegs zu verstehen, wurde untersucht, inwieweit diese Änderungen einer Zunahme des Niederschlags oder einer Zunahme der Temperatur und der damit verbundenen Zunahme der Gletscherschmelze zugeordnet werden können. Die Analyse umfasste einen datenbasierten Ansatz unter Verwendung von multilinearer Regression und einen simulationsbasierten Ansatz unter Einsatz eines hydrologischen Modells. Für die hydrologische Modellierung wurden Veränderungen in der Gletscherfläche und -höhe berücksichtigt. Im Rahmen einer multiobjekiven Kalibrierung wurden Kriterien in Bezug auf tägliche und zwischenjährliche Abflussvariationen, sowie Gletschermassenbilanzdaten in Form einer Massenbilanzzeitreihe für einen Gletscher und zwei geodätische Beobachtungen verwendet. Die jeweiligen Beiträge zu den Abflussänderungen durch Änderungen der Gletschergeometrie, der Temperatur und des Niederschlags wurden mithilfe von Simulationsexperimenten mit konstanter Gletschergeometrie bzw. Niederschlags- und Temperaturzeitreihen ohne Trend abgeschätzt. Die Ergebnisse zeigten, dass die beobachteten Änderungen im Abfluss mit den Änderungen in der Temperatur und im Niederschlag konsistent waren. In dem Gebiet mit einer stärkeren Vergletscherung wurden zunehmende Temperaturen als Hauptursache für die Abflusszunahme identifiziert, während in dem weniger stark vergletscherten Gebiet sowohl Niederschlags- als auch Temperaturzunahmen zum Abflussanstieg beigetragen haben. Im Vergleich der beiden Ansätze wurde die Aussagekraft der auf hydrologischer Modellierung beruhenden Methode als höher angesehen als die der auf multilinearer Regression basierenden Methode. Wichtige Einschränkungen des datenbasierten Ansatzes waren Korrelationen zwischen den erklärenden Variablen der multilinearen Regression, sowie die Tatsache, dass dieser auf rein statistischen Beziehungen und nicht auf Ursache-Wirkungsbeziehungen beruht.

Die Untersuchungsgebiete dieser Arbeit liegen im Karadarya-Gebiet in Kirgistan und im Aksu-Gebiet, welches sich in Kirgistan und in China befindet. Die sechs Einzugsgebiete im Karadarya-Gebiet haben eine Größe von 170–3800 km^2 und weisen nur eine geringe Gletscherbedeckung auf. Zwei der Gebiete haben keine Gletscher und in den anderen Gebieten liegt die Gletscherbedeckung zwischen 0.5 und 2.3%. Der mittlere Gebietsniederschlag wurde auf 700 bis 1200 mm a^{-1} geschätzt, und der mittlere Jahresabfluss erreicht Werte von knapp 300 bis fast 800 mm a^{-1}. Die untersuchten Einzugsgebiete im Aksu-Gebiet sind mit 18,400 und 12,900 km^2 deutlich größer als die im Karadarya-Gebiet, und zeichnen sich durch eine höhere Gletscherbedeckung von 4.4% und 21% aus. Mit geschätzten mittleren Jahresniederschlägen von 370 und 530 mm a^{-1}, ist das Klima im Aksu-Gebiet allgemein trockener als das im Karadarya-Gebiet. Der mittlere Jahresabfluss liegt bei 150 und 380 mm a^{-1}.

Durch die in dieser Arbeit entwickelten Ansätze konnte die Robustheit der hydrologischen Modellierung für datenarme Gebirgsgebiete in Zentralasien erhöht werden. Dadurch wurde ein besseres Verständnis des Wasserkreislaufs in diesen Gebieten erzielt, einschließlich der Wasserbilanz, hydrologischer Prozesse und hydrologischer Veränderungen. Die Ergebnisse tragen zu einer Erweiterung der wissenschaftlichen Grundlage für Entscheidungen im Bereich von Wassermanagement und Planung bei. Die entwickelten Ansätze für die hydrologische Modellierung können auch in andere datenarme Gebirgsregionen übertragen werden.

1 Introduction

1.1 Motivation

Water is a vital resource. It serves as drinking water, sustains ecosystems, and is required for irrigation, generation of electricity, and industrial water supply. In Central Asia, water is a scarce resource and plays an important economic role as agriculture in the lowland regions largely depends on irrigation, and hydropower is the most important source for electricity in some Central Asian countries. Since the lowland regions are characterized by a semiarid to hyperarid climate with little runoff generation, the mountain regions in Central Asia are of great importance for water resources (Viviroli et al., 2007).

A good understanding of the water cycle provides a basis for decisions related to water management and planning of water infrastructure. Hydrological modeling is a tool which helps to gain a better understanding of the water cycle and can give insights into hydrologic processes in mountain areas. This can aid in better understanding past changes, and estimating possible future changes of the hydrological cycle.

However, the application of hydrological models in Central Asia is challenging and involves large uncertainties since hydrometeorological data are sparse; this applies in particular for remote mountain regions. These difficulties raised the question of which approaches and additional data can be used to increase the credibility of hydrological models in such data sparse regions.

1.2 Water in Central Asia—a scarce resource

Rivers originating from the mountain ranges of the Tien Shan and Pamir in Central Asia feed two major watersheds: the Aral Sea Basin, with an area of 1.76×10^6 km^2 and a population of 46 million (FAO, 2013), located in Kyrgyzstan, Tajikistan, Afghanistan, Uzbekistan, Kazakhstan, and Turkmenistan; and the Tarim River Basin, with an area of 1.02×10^6 km^2 and a population of 10.4 million (Zhou et al., 2012), located primarily in the Xinjiang Uygur Autonomous Region of the People's Republic of China (for an overview of the region see Fig. 1.1). Dissent about the use of the water resources frequently leads to tensions between neighboring countries. While the upstream countries are interested in water releases from reservoirs in winter for generating hydro power energy, the downstream countries require water for irrigation during the growing season in summer (Abbink et al., 2010).

Water scarcity and the wasteful use of water resources led to substantial environmental and socio-economic problems. In the Aral Sea Basin, irrigated area increased by 80% from 4.7 to 8.3 million ha over the time period from 1960 to 2000 (Beek et al., 2011). The drastic expansion of irrigated areas had severe consequences, including the well-known disaster of the shrinkage of the Aral Sea, its increase in salt concentrations, the disappearance of fish in the Aral Sea, the degradation of the deltaic ecosystems, the distribution of salt and dust from the dried out Aral Sea into the region, and a change of the regional climate toward drier and warmer conditions (Micklin, 2007).

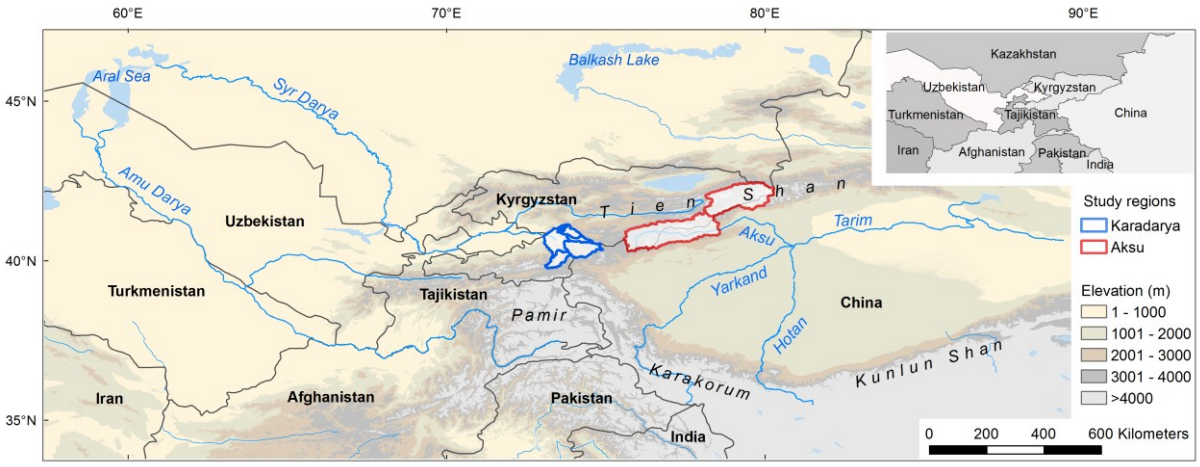

Fig. 1.1: Location of the study regions within the Aral Sea Basin and the Tarim River Basin.

A comparable development could also be observed in the Tarim Basin. In this region, irrigated areas nearly tripled over the time period 1950–2008 from 0.7 to 2 million ha (Zhou et al., 2012). As a consequence of the water withdrawals, the river bed of the Tarim River dried out over a stretch of 320 km, and the previous terminal lakes, Lop Nur and Taitema Lake, fell dry in 1970 and 1972 (Hao et al., 2009). The vegetation along the lower Tarim River degraded, and dust storms and soil salinization increased (Giese et al., 2005).

Even though these problems have been recognized and some improvements have been made, there is no simple solution, and water management in these regions faces great challenges. In addition to the existing problems, significant changes in the water cycle in Central Asia are expected as a consequence of climate change (Jarsjo et al., 2012; Siegfried et al., 2012; Hagg et al., 2013; Lutz et al., 2013; Sorg et al., 2014). Runoff in Central Asia is strongly influenced by snow and glacier melt, and is therefore particularly vulnerable to temperature changes (Barnett et al., 2005). An improved understanding of the hydrological cycle serves as a basis for decisions concerning present day water management, and its adaptation to climate change.

1.3 Additional data sources and approaches for hydrological modeling in data sparse mountain regions

Data required for hydrological modeling typically include time series of climate variables, information on the land surface—such as elevation, land cover, and soil types—and discharge time series for the calibration of model parameters. In catchments with a low density of hydrometeorological stations, it can be particularly useful to consider further data sources besides the usually applied data. Such additional data may, for example, be gained from atmospheric modeling or remote sensing.

A key challenge of applying a hydrological model in mountainous regions is the generation of reliable areal precipitation estimates, due to the high variability of precipitation and the low density of station networks (Weingartner et al., 2003). The spatial distribution of precipitation in mountainous regions is usually strongly related to the orography, and it is in most cases advantageous to explicitly consider the relation to the orography for the interpolation of station-based time series, e.g. within geostatistical or regression approaches (Hevesi et al., 1992; Phillips et al., 1992; Goovaerts, 2000). These methods however require that at least the relation between precipitation and the orography can be derived from the gauges, which may not be the case in data sparse catchments. As an alternative solution, the spatial distribution of precipitation may be derived from reanalysis data downscaled by a regional climate model (RCM) but this approach has hardly been investigated so far (Haberlandt and Kite, 1998; Tobin et al., 2011). RCMs simulate the formation of precipitation, considering the orography and the simulated wind field, and are therefore expected to be able to simulate a plausible precipitation distribution. However, as the performance of downscaled reanalysis data in representing the daily variability of precipitation may be low, their direct use in hydrological models may compromise the model performance. One possible solution therefore is to use the RCM to derive the spatial distribution of precipitation, which is then applied for an interpolation of station-based observations.

Traditionally, interpolation methods are assessed using cross-validation, where a part of the data is held back from the interpolation and used to evaluate the interpolated data. This method can however be misleading if the precipitation gauges are not in representative locations. As an example, one may assume a catchment with higher precipitation along a mountain range but with all precipitation gauges located in the lowlands. In this case, a superior precipitation estimate which reflects the higher precipitation along the mountain range will not be recognized as being superior to other methods not reflecting the higher precipitation along the mountain range. A well-suited alternative for the evaluation of different precipitation data sets in data sparse mountain catchments is to apply the precipitation data sets to a hydrological model and to compare the precipitation estimates based on the deviations between simulated and observed discharge (Heistermann and Kneis, 2011). This approach has the advantage that the precipitation data are evaluated not at the point scale but at the catchment scale.

In data-poor catchments, high uncertainties in the input data result in a high risk of internal inconsistencies of the model. A model may be able to well represent the observed discharge, but this may be due to error compensation resulting in the correct discharge response for the wrong reasons (Klemeš, 1986; Seibert and McDonnell, 2002). It has been recognized in the literature that in order to increase the internal consistency of the model, the model should not be evaluated based on discharge only, but also based on other variables (e.g., Güntner et al., 1999; Seibert, 2000). In remote catchments, satellite data are a well-suited source for additional data for model calibration or evaluation. Satellite data are often available for large regions or even globally and thus also in remote, inaccessible terrain. Satellite data have been applied for hydrologi-

cal modeling to derive estimates for state variables or fluxes, such as soil moisture (e.g., Brocca et al., 2010), snow cover (e.g., Li and Williams, 2008), total water storage (e.g., Werth and Güntner, 2010), or evaporation (e.g., Winsemius et al., 2008).

Satellite-derived snow cover data are particularly promising to be valuable for calibration in mountain catchments in Central Asia, as snowmelt is an important runoff generation process in these catchments. Satellite-derived snow cover can be regarded as a reliable product (Maurer et al., 2003; Parajka and Blöschl, 2006; Wang et al., 2009; Zhao and Fernandes, 2009; Gafurov et al., 2013), and products with suitable spatial and temporal resolutions are available. Snow cover data therefore have been applied for model calibration in a number of studies (Engeset et al., 2003; Udnaes et al., 2007; Koboltschnig et al., 2008; Parajka and Blöschl, 2008; Corbari et al., 2009; Sorman et al., 2009; Finger et al., 2011; Pellicciotti et al., 2012; Shrestha et al., 2013). As the information on snow cover is spatial, it is particularly suited for the evaluation of spatially distributed models. Snow cover data, which do not contain any information on snow water equivalents, can be regarded as complementary to discharge data, which provide information on the water balance integrated over the catchment area (Finger et al., 2011).

Remote sensing data on glacier thickness changes are a further valuable data source. In glacierized catchments, an underestimation of precipitation may be compensated by an overestimation of ice melt, and vice versa (Braun and Aellen, 1990; Schaefli et al., 2005; Stahl et al., 2008). It is therefore beneficial to include glacier mass balance estimates in the model calibration procedure in order to avoid such compensating errors (Schaefli et al., 2005; Stahl et al., 2008; Konz and Seibert, 2010; Schaefli and Huss, 2011; Mayr et al., 2013; Engelhardt et al., 2014). Measured mass balance estimates using the glaciological method are however only available for few individual glaciers. As an alternative, glacier mass balances derived from remote sensing data can be applied (Jost et al., 2012). Areal glacier thickness changes can be estimated from two digital elevation models of different points in time. Together with estimates for ice density, these elevation changes can be converted to glacier mass changes (Surazakov and Aizen, 2006; Pieczonka et al., 2013).

1.4 Understanding past hydrologic changes

Studying hydrologic changes observed in the past can give valuable insights in hydrologic systems. A better understanding of past changes provides an improved basis for estimating future changes. Hydrologic changes can become apparent as trends or step changes in hydrologic fluxes or state variables. This thesis concentrates on streamflow trends, as one indicator for hydrological change. While trend detection is the first step in trend analysis, the focus of my research was on attribution of the detected trends to their possible causes, since by studying this question we also learn more about the functioning of the hydrologic system in the investigated catchments. The attribution of streamflow trends may be performed using data-based or simulation-based approaches (Merz et al., 2012). Examples for data-based approaches are the analysis of concurrent trends in climate variables (e.g., Birsan et al., 2005; Pellicciotti et al., 2010; Kriegel et al., 2013), correlation analyses, multilinear regression (Aguado et al., 1992; Dettinger and Cayan, 1995; Moore and Demuth, 2001; Stahl and Moore, 2006), or generalized linear models (e.g., Bates et al., 2010). In contrast, in simulation-based approaches a hydrological model is applied to understand the observed changes (Hamlet et al., 2005; Hidalgo et al., 2009).

1.5 Objectives and research questions

The main objectives of this thesis are:

- The development of hydrological modeling approaches which explicitly consider the peculiarities of data sparse mountain catchments in order to obtain credible and internally consistent hydrological models, despite the data sparse situation.
- Achieving a better understanding of past streamflow changes in two snow and glacier melt dominated mountain catchments by attributing streamflow changes to their possible drivers.

In the following, the detailed research questions are introduced.

Can monthly precipitation fields from downscaled reanalysis data be used for interpolating gauge observations?

The use of reanalysis data for the interpolation of precipitation has so far been investigated in only few studies. Haberlandt and Kite (1998) used daily precipitation from the NCAR reanalysis (without downscaling) as external drift variable for the geostatistical interpolation of observed precipitation time series in the 1.8 million km^2 Mackenzie River Basin in north-western Canada, and more 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 for two catchments in the central Alps in Switzerland. To my knowledge, spatial fields from downscaled reanalysis data have so far not been used for precipitation interpolation in more data sparse mountain catchments, where they could be particularly useful. The performance of reanalysis data is usually higher at monthly as compared to daily time steps (Hurkmans et al., 2008). The present study therefore investigates whether monthly precipitation fields from downscaled reanalysis data are a useful data source for the spatial interpolation of station-based time series.

How can areal precipitation estimates be compared by hydrological modeling?

The advantages of evaluating precipitation data sets using observed discharge and hydrological modeling have been recognized by a number of studies (Vieux and Bedient, 2004; Gourley and Vieux, 2005; Yilmaz et al., 2005; Artan et al., 2007; Cole and Moore, 2008; Stisen and Sandholt, 2010; Behrangi et al., 2011; Bitew and Gebremichael, 2011a; Heistermann and Kneis, 2011; Shrestha et al., 2013). The idea is to use the different precipitation data sets as input to the hydrological model, and to evaluate these data sets based on the comparison between observed and simulated discharge. The model is either recalibrated for each precipitation data set, or Monte Carlo simulations over the whole parameter range are performed (Gourley and Vieux, 2005; Heistermann and Kneis, 2011). In contrast, using a model calibrated to only one specific precipitation data set may lead to biased results and is therefore not recommended. In this thesis, the calibration approach has been chosen since the Monte Carlo approach can in some cases lead to wrong conclusions, e.g., if the model parameters have a linear influence on the discharge volume, and the precipitation data sets differ primarily by a different systematic error (Heistermann and Kneis, 2011).

Previous applications of this method did not allow a separate evaluation of the temporal dynamics and the overall bias (over- or underestimation) of the precipitation estimate. While a systematic bias of a precipitation data set may also be derived from the comparison of observed and simulated discharge, this has two disadvantages: First, it is not possible to specify by how much a precipitation data set over- or underestimates because the observed percentage over- or underestimation of discharge does usually not correspond to the same percentage bias in precipitation. Second, it is not possible to separately evaluate the temporal performance of the precipitation data set since it will be influenced by the over- or underestimation, which may lead to a different behavior of the hydrological system.

This study therefore aimed at extending the approach in order to separately consider the temporal dynamics and the overall bias. This was achieved by adding a precipitation bias factor to the calibration parameters. Over- or underestimation of the precipitation data set is then evaluated using this precipitation bias factor, while the temporal performance of the precipitation data set is evaluated using a Nash-Sutcliffe based performance criterion. Uncertainties in both criteria are considered within the model calibration framework.

What are the trade-offs between good model performance with respect to discharge and snow cover?

While the benefit of applying satellite-derived snow cover data for model calibration has been demonstrated in a number of studies (Engeset et al., 2003; Udnaes et al., 2007; Koboltschnig et al., 2008; Parajka and Blöschl, 2008; Corbari et al., 2009; Sorman et al., 2009; Finger et al., 2011; Pellicciotti et al., 2012; Shrestha et al., 2013), studies which explicitly dealt with characterizing the trade-off between model performance in terms of snow cover and discharge are rare. Notable exceptions are Parajka et al. (2007), who applied snow cover data derived from interpolated measurements of snow water equivalent and not from remote sensing, and Parajka and Blöschl (2008), who used a weighted sum approach to combine the objective functions for satellite-derived snow cover area and discharge into one criterion but also analyzed the effects of varying the weights. This thesis therefore aimed at explicitly characterizing the trade-off curves between good model performance with respect to discharge and snow cover, and further at contrasting the results from the multiobjective optimization against the results from a Monte Carlo approach and single-objective optimizations.

Furthermore, this thesis aimed at an improved approach for the evaluation of a semidistributed model with respect to gridded snow cover data from remote sensing. Some studies compared the fractional area of model units with simulated snow to the fractional area with observed snow cover over the whole catchment (Engeset et al., 2003; Udnaes et al., 2007; Sorman et al., 2009). Parajka and Blöschl (2008) additionally made use of the information on the snow cover distribution with elevation by summarizing the satellite snow cover data per elevation zone and comparing it to the simulated snow water equivalent for the different elevation zones. Their approach however required thresholds

for the comparison between snow water equivalents and snow cover area, and the comparison may be influenced by the choice of these thresholds. By introducing a parameterization for the simulation of fractional snow cover areas (Liston, 2004), this thesis therefore aimed at enabling a comparison between simulated and observed snow cover data which considers the information on the snow cover distribution with elevation and does not rely on subjective thresholds.

How many snow cover scenes should be used for model calibration?

As the processing of snow cover scenes can be time consuming and costly, for example due to the need for manual georeferencing, the question how many snow cover scenes should be used for model calibration is a very relevant one. While a lot of work has focused on the amount of discharge data required for model calibration (e.g., Perrin et al., 2007, and references cited therein; Juston et al., 2009), there are, to my knowledge, no comparable studies which investigated the number of snow cover scenes required for model calibration. In the current study I therefore investigated how the snow cover error in the validation period changes with an increasing number of snow cover scenes in the calibration period.

To what extent can streamflow increases in two headwater catchments of the Aksu River be attributed to increases in temperature and precipitation?

A number of studies used data-based approaches for the attribution of observed streamflow trends in snow and glacier influenced catchments (Aguado et al., 1992; Dettinger and Cayan, 1995; Moore and Demuth, 2001; Birsan et al., 2005; Stahl and Moore, 2006; Pellicciotti et al., 2010; Kormann et al., 2014). In contrast, only few studies applied simulation-based approaches (Hamlet et al., 2005; Zhao et al., 2013; Engelhardt et al., 2014), despite their advantages particularly in mountainous catchments, where one needs to consider variations in temperature and precipitation, snow storage, and possibly glacier storage, at different elevations. For example, in order to check whether it is quantitatively plausible that a certain increase in winter temperature caused an increase in streamflow by a certain amount because of increased snowmelt, one would need to extrapolate the temperature to the different elevations and make a rough estimate on the possible impacts on snowmelt—which may be performed more easily using a simulation model.

In the headwater catchments of the Aksu River, the most important tributary to the Tarim River, streamflow increased by around 30% over the period 1957–2004. This area therefore forms a well-suited example for studying hydrologic change in snow- and glacier melt dominated mountain catchments. The increases are assumed to be caused by higher precipitation and higher temperatures, leading to higher glacier melt, but a thorough attribution study is still missing. While Zhao et al. (2013) attributed the runoff increase in the two catchments to 96% and 100% to an increase in precipitation, these results might not be reliable since the study neglected parameter uncertainties and observations on glacier mass balances. Due to the large uncertainties in precipitation there is a great risk of compensating errors in the areal precipitation estimate by a corresponding over- or underestimation of glacier melt.

This thesis applied both a data-based approach, using multiple linear regression analysis, and an approach based on hydrological simulations. For the simulation-based approach, the particularities of data sparse catchments are taken into account by considering uncertainties in the areal precipitation estimates and by using a multiobjective calibration approach with criteria based on discharge and glacier mass balance data. In highly glacierized areas affected by climatic changes, also changes in glacier area and elevation need to be considered. However, while changes in glacier geometry have been taken into account in hydrological models for climate impact assessments (e.g., Stahl et al., 2008; Huss et al., 2010; Farinotti et al., 2012), there are, to my knowledge, no studies considering glacier geometry changes in hydrological models applied for the attribution of streamflow trends. In the current study, impacts of changes in temperature, precipitation, and glacier geometry on the streamflow trends were analyzed using simulations with detrended temperature and precipitation time series and simulations where the glacier geometry was held constant. The investigated research questions were (1) whether the temperature and precipitation changes were consistent with the streamflow increases, (2) to what extent the streamflow increases in the two headwater catchments of the Aksu River can be explained by increases of temperature (increasing glacier melt), or by increases in precipitation, and (3) how glacier melt, snowmelt and rainfall changed over time.

1.6 Study areas

The study areas of this thesis were headwater catchments of the Karadarya Basin and of the Aksu Basin. The Karadarya Basin is located in southern Kyrgyzstan. The Karadarya flows into the Syrdar-

ya, which is the second largest inflow to the Aral Sea (Fig. 1.1). The six studied headwater catchments in the Karadarya Basin have a size of 170–3800 km^2. Elevations range between 1250 and 4750 m. The land cover is dominated by grasslands, and in the lower elevations there is also a high percentage of croplands. Glaciers are present in four of these catchments with glacier cover ranging between 0.5 and 2.3%. Average annual areal precipitation was estimated between 700 and 1200 mm a^{-1}, and average annual runoff ranges between less than 300 and nearly 800 mm a^{-1}.

The Aksu Basin is located in Kyrgyzstan and the Xinjiang Uyghur Autonomous Region in China (Fig. 1.1). The confluence of Aksu, Hotan, and Yarkand forms the Tarim River, and with a discharge contribution of about 80% to the Tarim, the Aksu is the most important tributary. This study focuses on two headwater catchments of the Aksu Basin, with a size of 18,400 and 12,900 km^2. Average elevations are around 3600 and 3700 m, while maximum elevations reach up to 5900 and 7100 m. The land cover consists of grassland, barren or sparsely vegetated land, and snow and ice. Glaciers cover 4.4% and 21% of the two catchments. With estimated average annual precipitation of 370 and 530 mm a^{-1}, the climate in the headwater catchments of the Aksu Basin is generally drier than in the Karadarya Basin. Average annual discharge in the two catchments is around 150 and 380 mm a^{-1}, respectively. More detailed descriptions of the study regions including maps and overview tables can be found in chapter 2 and 3 for the Karadarya Basin and in chapter 4 for the Aksu Basin.

1.7 Overview

This thesis is written in a cumulative form, and chapters 2 to 4 are articles published in or submitted to international peer-reviewed journals. These articles address the research questions outlined in Section 1.5. Chapter 5 finally summarizes the main results with respect to the research questions, provides an overall discussion, and outlines questions for further research. The first four research questions were investigated in the Karadarya Basin. Due to the better availability of discharge and meteorological data, as well as the lower influence of glacier melt on discharge, catchments in the Karadarya Basin were better suited for the methodological improvements than those in the Aksu Basin. The last research question focuses on the study catchments in the Aksu Basin. Since the observed streamflow in the Aksu headwater catchments showed clear increases over the last decades, these catchments represent ideal study sites for investigating the causes for the observed increases in streamflow, as an example of hydrological change in snow and glacier melt dominated catchments.

An overview of which research questions are addressed in which chapter is provided below:

The research questions

- Can monthly precipitation fields from downscaled reanalysis data be used for interpolating gauge observations?
- How can areal precipitation estimates be compared by hydrological modeling?

are investigated in chapter 2 *Evaluation of areal precipitation estimates based on downscaled reanalysis and station data by hydrological modeling.*

The questions

- What are the trade-offs between good model performance with respect to discharge and snow cover?
- How many snow cover scenes should be used for model calibration?

are addressed in chapter 3 *The value of satellite-derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia.*

Finally, the question

- To what extent can streamflow increases in two headwater catchments of the Aksu River be attributed to increases in temperature and precipitation?

is focus of chapter 4 *Attribution of streamflow trends in snow- and glacier melt dominated catchments of the Tarim River, Central Asia.*

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

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 multilinear 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 modeling 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 estimates 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 modeling and a comparison to observed discharge.

Published as:

Duethmann, D., J. Zimmer, A. Gafurov, A. Güntner, D. Kriegel, B. Merz, and S. Vorogushyn (2013): Evaluation of areal precipitation estimates based on downscaled reanalysis and station data by hydrological modeling, *Hydrol. Earth Syst. Sci.*, *17*(7): 2415–2434.

2.1 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 toward 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 (Hevesi et al., 1992; Phillips et al., 1992; Martinez-Cob, 1996; Goovaerts, 2000; 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 (Hevesi et al., 1992; Phillips et al., 1992; Garen et al., 1994; Martinez-Cob, 1996; Goovaerts, 2000; Lloyd, 2005; Tobin et al., 2011), or using multilinear or polynomial regression against various topographical variables (Basist et al., 1994; Daly et al., 1994; Goodale et al., 1998; Hay et al., 1998; Prudhomme and Reed, 1998; Ninyerola et al., 2000; Brown and Comrie, 2002; Cheval et al., 2003; Johansson and Chen, 2003; Marquinez et al., 2003; Johansson and Chen, 2005; Perry and Hollis, 2005; Ninyerola et al., 2007; 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 multilinear 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 (Stisen and Sandholt, 2010; Heistermann and Kneis, 2011). 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 modeling. The model may either be recalibrated for each precipitation data set (Yilmaz et al., 2005; Artan et al., 2007; Stisen and Sandholt, 2010; Behrangi et al., 2011; Bitew and Gebremichael, 2011b, a) or applied within a Monte Carlo framework (Gourley and Vieux, 2005; Heistermann and Kneis, 2011). 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 nonlinearities 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 modeling 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 modeling. 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.

2.2 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 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. 2.1). 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 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. 2.1 and important characteristics are listed in Table 2.1. For most of these subcatchments the mean annual runoff over the period 1961–1990 has values of 400 mm a^{-1} to 600 mm a^{-1}, outliers are Ak-Tash with nearly 800 mm a^{-1} and Gulcha in the south with less than 300 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 2.1: Area, glacierization, elevation range, and mean annual runoff over the period 1961–1990 of the studied subcatchments of the Karadarya Basin.

	Area	Glacier	Elevation (m)			Runoff
	(km^2)	(%)	min.	max.	mean	(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

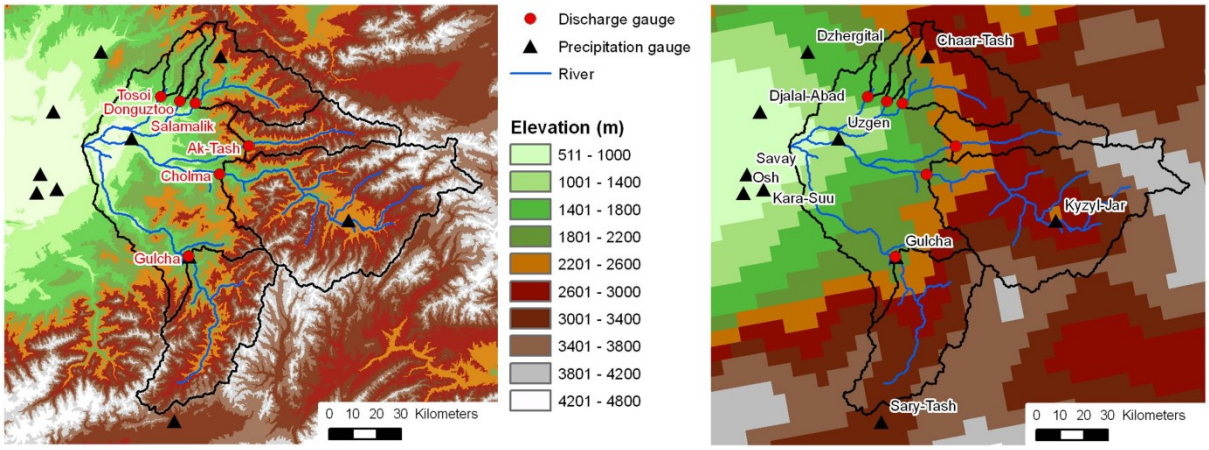

Fig. 2.1: The Karadarya Basin upstream of the Andijan reservoir. Left: SRTM elevation, right: elevation in the WRF model. Shown are the headwater subcatchments where the hydrological model is applied (black outlines), and their corresponding discharge gauges (red dots), as well as the precipitation gauges (black triangles).

2.3 Data and methods

2.3.1 Precipitation data and interpolation approaches

2.3.1.1 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 initialization and discarded. Fig. 2.1 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.

2.3.1.2 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 (equations 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 m s^{-1} is assumed. The undercatch correction results, on average, in an increase of the measured values by 10%.

2.3.1.3 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:

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

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

2.3.1.4 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 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:

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

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.

2.3.1.5 Interpolation of station data using monthly fields derived by multilinear 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 multilinear 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 equations 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 multilinear regression to generate the daily precipitation fields (see equation 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).

2.3.1.6 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 modeling. 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 2.2 lists all precipitation data sets used in this study.

2.3.2 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 x 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 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

Table 2.2: 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 multilinear regression, monthly maps from individual years
MLR-all	Station data interpolated using monthly precipitation maps from multilinear 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)

time series and evaluated using bias and mean absolute error of the daily time series.

2.3.3 Evaluation of areal precipitation estimates based on simulated discharge

2.3.3.1 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 modeling: 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 calibration 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 characterized 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 $\psi_{true} < 0.7$ ($\psi_{true} > 0.7$) 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 analyzing the model for possible ill-posed settings becomes a less sensitive issue. Another advantage is that by using an optimization 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 optimized 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.

2.3.3.2 Description of the hydrological model

The hydrological model WASA (Güntner, 2002; Güntner and Bronstert, 2004) is a semidistributed 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 discretization 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 2.3). 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).

2.3.3.3 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 digitized 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 digitizing. 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.

Table 2.3: 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} d^{-1}	1	7
	melt temperature	°C	−2	2
	glacier melt factor	mm °C^{-1} d^{-1}	0	7
Infiltration and percolation	*kf_corr_f*	-	0.01	100
	k_sat_f	-	0.01	100
	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

2.3.3.4 Model optimization 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 initialized 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 is applied:

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

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 penalized 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 optimization 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 optimization runs are started. These short optimization runs are meant to get into the region of the optimum, but the length of the optimization run is also short enough that they mostly do not reach the objective function maximum. For each optimization, 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 optimization 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.

2.3.3.5 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 2.4). 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 2.4 varying one factor at a time. This analysis is performed for all of the six subcatchments and all precipitation data

Table 2.4: 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 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 and 140 cm; mostly 50–100 cm	Multiply original value by 0.5 and 2
Minimum stomata resistance	Varies by land cover; e.g. grassland 126 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

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).

2.4 Results and discussion

2.4.1 Characteristics of the precipitation data sets

2.4.1.1 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 (section 2.4.1.2). 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 modeling 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 2.5). There is no relationship of the bias with elevation. 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. 2.2).

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 ERA-15 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 one year and found bias values up to −50% and +100% for individual seasons

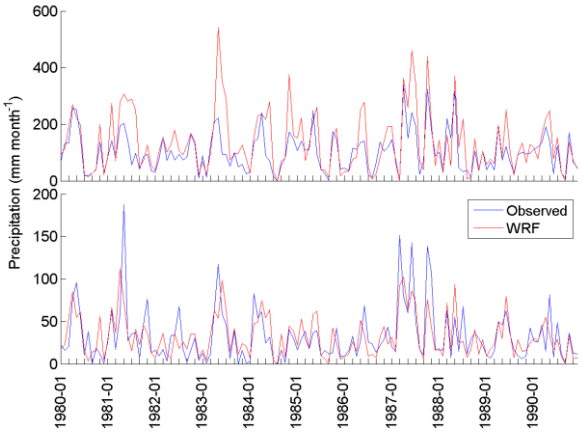

Fig. 2.2: Comparison of monthly time series over the period 1980–1990 of observed precipitation and WRF downscaled ERA-40 data at the gauge location for Chaar-Tash (top) and Kyzyl-Jar (bottom).

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 modeled 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), monthly fields from multilinear regression (MLR-all and MLR-ind), and inverse distance weighting are 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. 2.3). 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').

Table 2.5: 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

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 behavior when individual stations are omitted.

2.4.1.2 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 respect to their spatial distribution (Fig. 2.4). 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 multilinear regression show a strong increase of precipitation with increasing elevation; additionally precipitation also

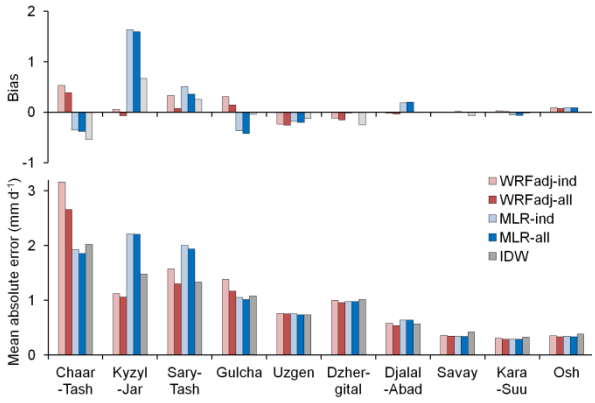

Fig. 2.3: Bias and mean absolute error calculated from cross-validation for the interpolation methods WRFadj-all, WRFadj-ind, MLR-all, MLR-ind, and IDW for precipitation stations in or close to the Karadarya catchment.

decreases to the south and to the east (Fig. 2.4 d, e). The WRF downscaled ERA-40 precipitation data indicate spots with very high precipitation values in the southern part of the catchment (Fig. 2.4 c). 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. 2.1). 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. 2.4 a, 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 markedly differs 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. 2.4 f, 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. 2.5). 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. 2.5 a-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.

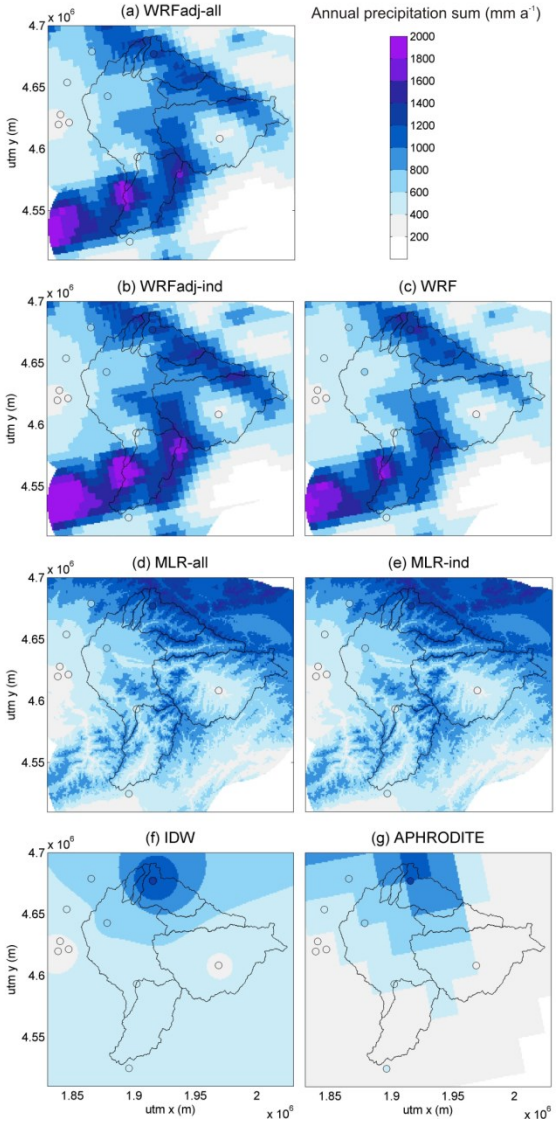

Fig. 2.4: Estimates of the mean annual precipitation (1960–1990) over the Karadarya catchment using different methods. Circles indicate measured precipitation at stations. Lines indicate subcatchment borders.

2.4.1.3 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. 2.6). 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

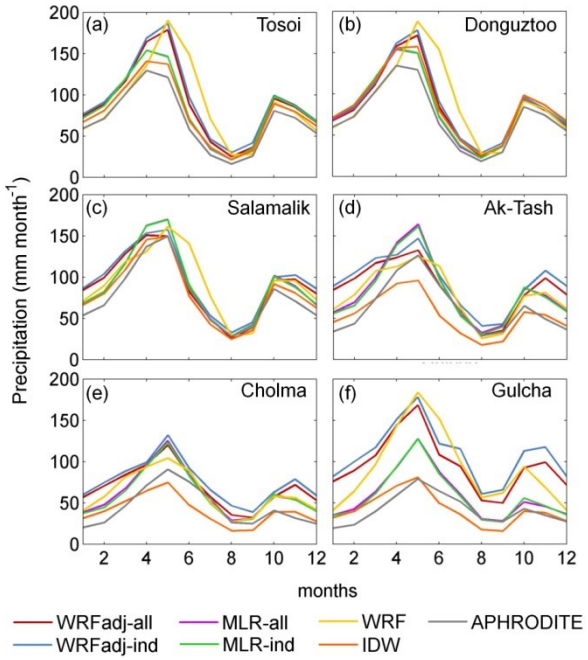

Fig. 2.5: Monthly subcatchment mean precipitation (1960–1990) for six subcatchments of the Karadarya Basin and seven different precipitation estimates.

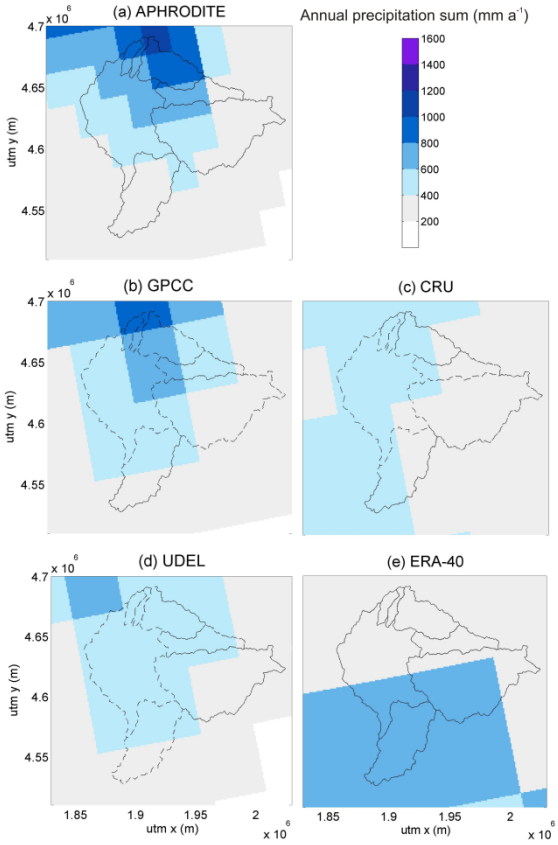

Fig. 2.6: Mean annual precipitation (1960–1990) for the Karadarya catchment from APHRODITE and four globally available gridded precipitation data sets GPCC, CRU, UDEL, and ERA-40.

are considerably lower than our estimates based on the methods WRFadj, WRF, or MLR (Fig. 2.4 a–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. 2.7). 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 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.

2.4.2 Evaluation of the precipitation data based on simulated discharge

2.4.2.1 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. 2.8). The general behavior 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

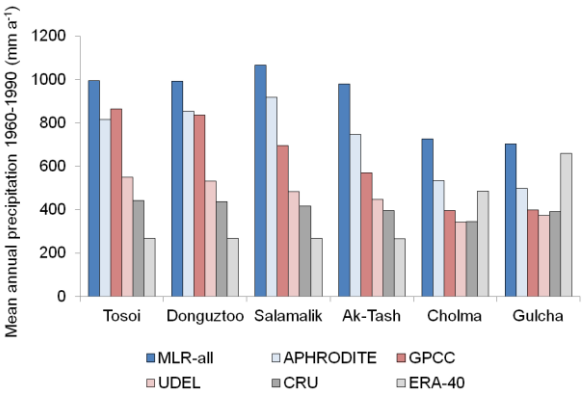

Fig. 2.7: Mean annual precipitation (1960–1990) for six subcatchments of the Karadarya Basin for the precipitation data sets MLR-all, APHRODITE, and four global precipitation data sets.

fact that very different parameter sets result in comparable model performances, unconstrained parameters can for example not be used for catchment characterization, and it is not possible to transfer individual parameters to catchments with similar characteristics. This does however not impede the objectives of this study.

Scatterplots 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 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.

2.4.2.2 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. 2.9), 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, section 2.4.1.1). 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 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. 2.9, 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.

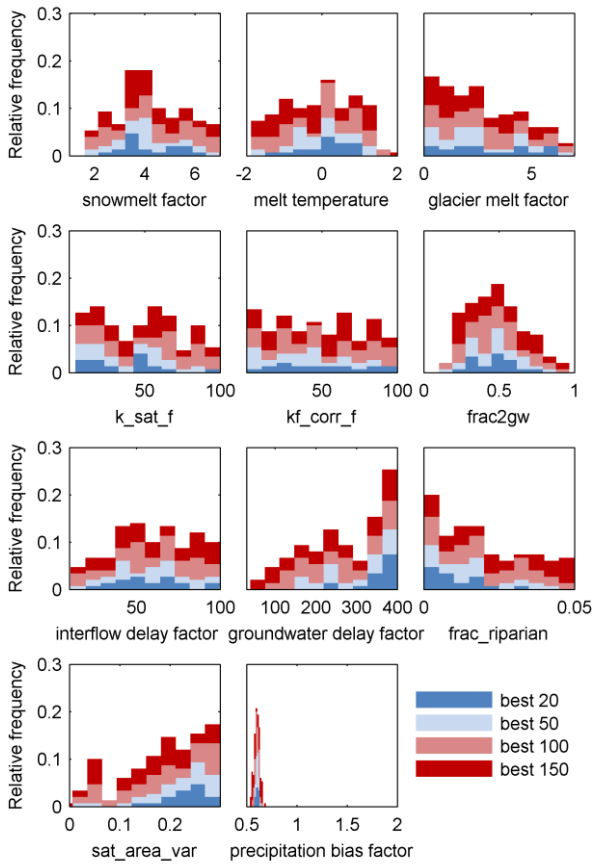

Fig. 2.8: Histograms of the parameter distributions for the best 20, 50, 100, and 150 parameter sets for the subcatchment Gulcha, precipitation estimate 'WRF', and time period 1979–1984.

2.4.2.3 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 multilinear regression seem to be the most suitable precipitation estimates (Fig. 2.10). 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 modeled 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. 2.10, and more clearly visible for one subcatchment in Fig. 2.11), 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 behavior 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 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. 2.1).

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 remains. 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

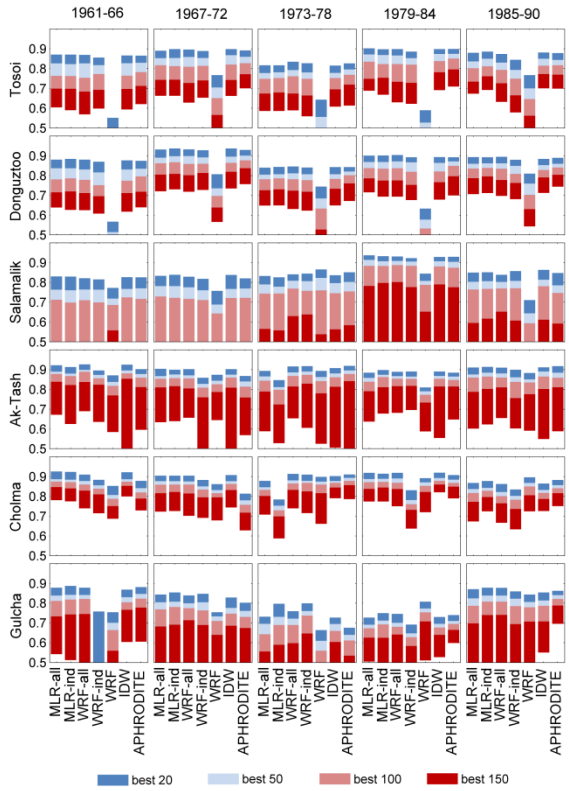

Fig. 2.9: Variation of the objective function values for different precipitation data sets (bars in each plot), different time periods (columns), and different subcatchments (rows). The colors indicate the range of objective function values for the best 20, 50, 100, and 150 parameter sets.

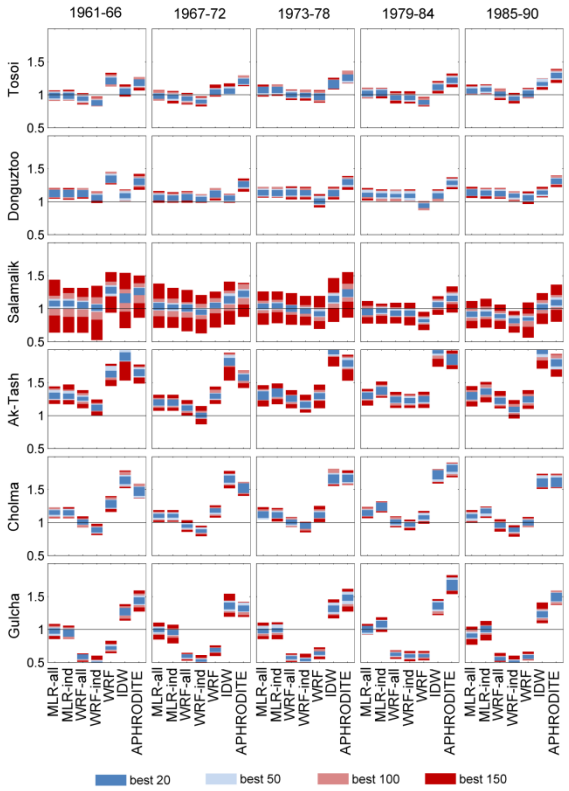

Fig. 2.10: Variation of the precipitation bias factor for different precipitation data sets (bars in each plot), different time periods (columns), and different subcatchments (rows). The colors indicate the range of the precipitation bias factor for the best 20, 50, 100, and 150 parameter sets.

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 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.

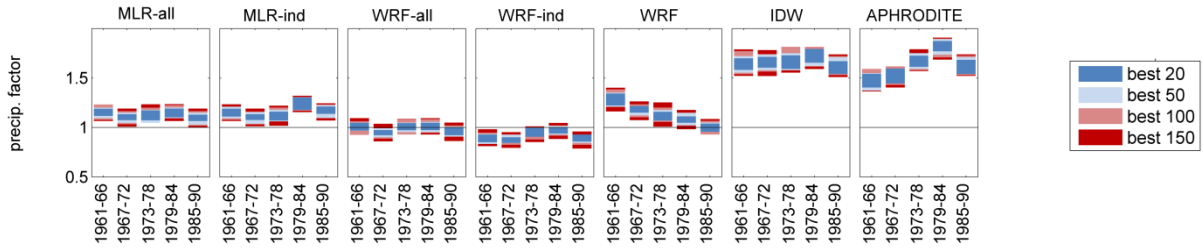

Fig. 2.11: Variation of the precipitation bias factor by time period for the subcatchment Cholma as an example. The data are the same as also shown in Fig. 2.10, but they are sorted in a different way in order to better demonstrate the variation by time period (particularly noticeable for the WRF and APHRODITE precipitation data sets).

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.

2.4.2.4 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 section 2.3.3.5), 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. 2.12. The boxplots summarize 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. 2.12 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

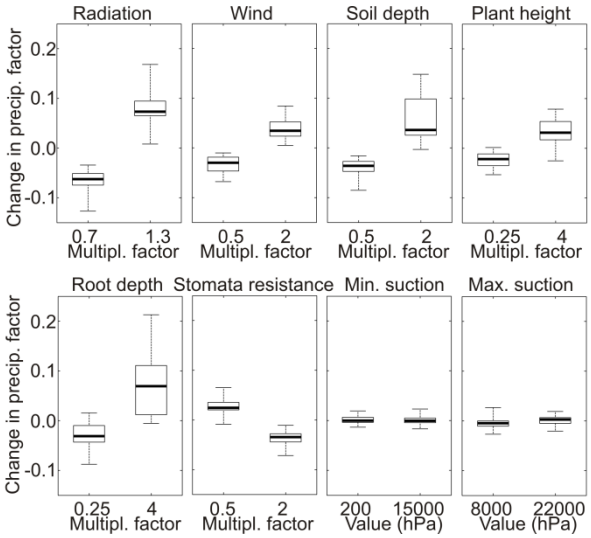

Fig. 2.12: Sensitivity analysis of the change in the calibrated precipitation bias factor as a result of changes in inputs for the time period 1979–1984. The boxplots show summaries of the results averaged over the six subcatchments and seven precipitation data sets with the thick black line indicating the median, the boxplot area the interquartile range and the whiskers the minimum and maximum change.

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 analyzed climate, soil and plant inputs are biased in a way that they would result in the same direction of change of the precipitation factor.

2.5 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 variability of precipitation cannot be derived from ground-based observa-

tions 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 modeling 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 modeling 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 multilinear 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 multilinear 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 modeling 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 multilinear 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 modeling and observed runoff might provide more reliable information.

When evaluating different precipitation estimates by hydrological modeling 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 modeling, 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 re-

gions where the differences in the precipitation estimates are only small, an evaluation by hydrological modeling 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 modeling can be assumed to be smaller than the differences between the precipitation data sets, making this approach very well suited for such regions.

Acknowledgements

This work was carried out within the projects CAWa (Water in Central Asia; http://www.cawa-project.net), funded by the German Federal Foreign Office (grant no. AA7090002), and SuMaRiO (Sustainable Management of River Oases along the Tarim River; http://www.sumario.de/), funded by the German Federal Ministry of Education and Research (grant no. 01LL0918A). We thank the Central Asian Institute for Applied Geosciences (CAIAG) for digitizing soil and discharge data and the Hydrometeorological Service of Kyrgyzstan for providing discharge and meteorological data.

3 The value of satellite-derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia

Abstract

Including satellite-derived snow cover data for hydrologic model calibration can be a good way to improve model internal consistency. This study applied a multiobjective genetic algorithm to characterize the trade-off curve between model performance in terms of discharge and snow cover area (SCA). Using a Monte Carlo-based approach, we further investigated the additional information content of an increasing number of SCA scenes used in the calibration period. The study was performed in six snowmelt-dominated headwater catchments of the Karadarya Basin in Kyrgyzstan, Central Asia, using the hydrological model WASA and snow cover data from four melt seasons retrieved from AVHRR (Advanced Very High Resolution Radiometer). We generally found only small trade-offs between good simulations with respect to discharge and SCA, but good model performance with respect to discharge did not exclude low performance in terms of SCA. On average, the snow cover error in the validation period could be reduced by very few images in the calibration period. Increasing the number of images resulted in only small further improvements. However, using only a small number of images involves the risk that these particular images cause the selection of parameter sets which are not representative for the catchment. It is therefore advisable to use a larger number of images. In this study, it was necessary to include at least 10–16 images.

Published as:

Duethmann, D., J. Peters, T. Blume, S. Vorogushyn, and A. Güntner (2014): The value of satellite-derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia. *Water Resour. Res., 50*(3), 2002–2021.

3.1 Introduction

Mountain water resources play an important role for the water supply of downstream areas (Viviroli et al., 2011). Particularly in dry regions with large populations, such as Central Asia, they are essential for irrigation, hydropower generation, and sustaining ecosystems which depend directly or indirectly on the river flow. Potential impacts of climate change on mountain water resources are thus of great concern for water management. Assessments of the hydrologic impacts are required for developing strategies to cope with changing conditions, which is often approached using hydrological models in combination with climate scenarios.

For such a task, the hydrological model needs to be able to extrapolate beyond the conditions during model calibration. This implicates that also internal processes, which have not undergone validation against observations, need to be simulated correctly (Seibert, 2000). It has been recognized for a while that good discharge simulations at the catchment outlet cannot guarantee good internal functioning of the model (Refsgaard, 1997) as this may be an effect of error compensation (Klemeš, 1986; Seibert and McDonnell, 2002). Model consistency may be increased by including internal variables in the model calibration and evaluation procedure (Güntner et al., 1999; Seibert, 2000). Several studies have shown the utility of multivariable model calibration and validation, using variables like groundwater (Lamb et al., 1998; Madsen, 2000; Seibert, 2000; Juston et al., 2009), soil moisture (Franks et al., 1998; Parajka et al., 2009), streamflow salinity (Mroczkowski et al., 1997), snow cover (Parajka et al., 2007), or glacier mass balance (Stahl et al., 2008; Konz and Seibert, 2010; Finger et al., 2011; Schaefli and Huss, 2011).

In mountain catchments dominated by snowmelt runoff, the correct representation of snow processes is crucial. Snow cover patterns from remote sensing can therefore be a useful data source for constraining model parameters. While these data do not contain information on the snow water equivalent (SWE), they are spatially distributed, which makes them particularly useful for the evaluation of distributed models. They can thus be seen as complementary to the discharge time series, which are spatially integrated but give quantitative information on the water balance (Finger et al., 2011). Remotely sensed snow cover data have been applied for hydrological modeling in different ways: as model forcing for snowmelt runoff models (e.g. Li and Williams, 2008), for model updating in data assimilation approaches (Rodell and Houser, 2004; Andreadis and Lettenmaier, 2006; Zaitchik and Rodell, 2009; De Lannoy et al., 2012; Yatheendradas et al., 2012; Liu et al., 2013), and for model calibration (Engeset et al., 2003; Udnaes et al., 2007; Koboltschnig et al., 2008; Parajka and Blöschl, 2008; Corbari et al., 2009; Sorman et al., 2009; Finger et al., 2011; Pellicciotti et al., 2012; Shrestha et al., 2013). Parajka and Blöschl (2008) performed an extensive study over 148 catchments in Austria on the value of MODIS (Moderate-resolution Imaging Spectroradiometer) snow cover data for calibration of the HBV (Hydrologiska Byråns Vattenbalansavdelning) model. Their results are very encouraging as they showed that, on average, the inclusion of MODIS snow cover during model calibration led to better snow cover and runoff simulations in the validation period. Finger et al. (2011) applied a Monte Carlo approach for the calibration of a grid-based model against discharge, satellite snow cover, and glacier mass balance. Their investigation of which data combinations are particularly useful for model conditioning showed a superior performance of the combination of discharge and snow cover data. However, studies explicitly showing the trade-off between snow cover and discharge performance are rare. Parajka et al. (2007) analyzed trade-off curves of model performance for discharge and snow cover area (SCA) derived from interpolated measurements of snow water equivalent (SWE). Parajka and Blöschl (2008) used a weighted sum approach to combine the objective functions for satellite-derived SCA and discharge into one criterion, but also analyzed the effects of varying the weights. In this study, we therefore explicitly analyze trade-offs between good model performance with respect to discharge and SCA. The results from the multiobjective optimization are also contrasted with results from Monte Carlo sampling and single-objective optimizations.

A further question concerns the amount of snow cover scenes required for model calibration. Whereas some remote sensing snow cover data, such as those from MODIS, can be obtained as ready product, others, like AVHRR (Advanced Very High Resolution Radiometer) or Landsat data, first need to be processed as only the raw data are available. This can be very time consuming, e.g., due to the need for manual georeferencing and sensor calibration. It is therefore very valuable to know how the added information content of additional images declines with an increasing number of observation scenes. While for discharge data this question has been addressed by a number of studies (Perrin et al., 2007, and references therein; see, e.g., Juston et al., 2009), we could not find any information on how much remote sensing snow cover data should be used for model calibration. This study therefore investigates the value of increasing the number of snow cover scenes for the calibration of a hydrological model.

In the present study, a hydrological model is applied to headwater catchments of the Karadarya Basin in Kyrgyzstan, Central Asia. In these catchments, the runoff regime is strongly influenced by snowmelt runoff in spring and early summer, and snow cover

data are therefore expected to be very well suited for constraining the model and improving model consistency. Due to lower data availability after 1990, a time period before 1990 was selected for model calibration. MODIS snow cover data, which are only available from 2000, could therefore not be used and we instead resorted to AVHRR data.

We first analyze the trade-offs between the model performance with respect to discharge and SCA using a multiobjective genetic algorithm to identify Pareto optimal solutions. Second, the study investigates within a Monte Carlo approach how the model performance in terms of snow cover prediction changes with an increasing number of SCA scenes in the calibration period. For the model evaluation against gridded satellite snow cover data, we suggest an approach which makes use of the information on the snow cover distribution with elevation and does not require threshold values for the comparison between simulated SWE and observed SCA.

3.2 Study area

This study focuses on six headwater catchments of the Karadarya Basin located in Kyrgyzstan (Fig. 3.1). Together with the Naryn River, the Karadarya forms the Syrdarya, which drains into the Aral Sea. The selected catchments are all upstream of the Andijan Reservoir and the influence of water management is generally assumed to be only marginal, except for the catchment Gulcha, where water is sometimes diverted into an irrigation channel. The catchment areas range from 170 to 3840 km^2 (Table 3.1). The catchments are characterized by high elevation gradients and maximum elevations up to 4750 m (Fig. 3.2). The predominant land cover is mostly grassland, and two catchments with lower elevations (Tosoi and Donguztoo) also have a larger fraction of agricultural land. Forest, which often impedes remote sensing of snow cover (Raleigh et al., 2013), only covers small fractions with a maximum value of 5% in Gulcha. A previous study analyzed the precipitation input to these catchments using hydrological modeling and observed discharge (Duethmann et al., 2013). While average annual precipitation at the gauges in vicinity to the study catchments ranges from 350 mm a^{-1} (measured at the stations in the lowland west of the study catchments) to 1050 mm a^{-1} (for the station in the northern mountain range), catchment average precipitation in the six study catchments is estimated to be between 700 and 1200 mm a^{-1}. Mean annual runoff over the period 1961–1990 ranges from <300 mm a^{-1} in the catchment Gulcha to nearly 800 mm a^{-1} in Ak-Tash (Table 3.1), resulting in annual runoff coefficients of approximately 0.4–0.6. The discharge regime is dominated by snowmelt with maximum discharges in spring and early summer.

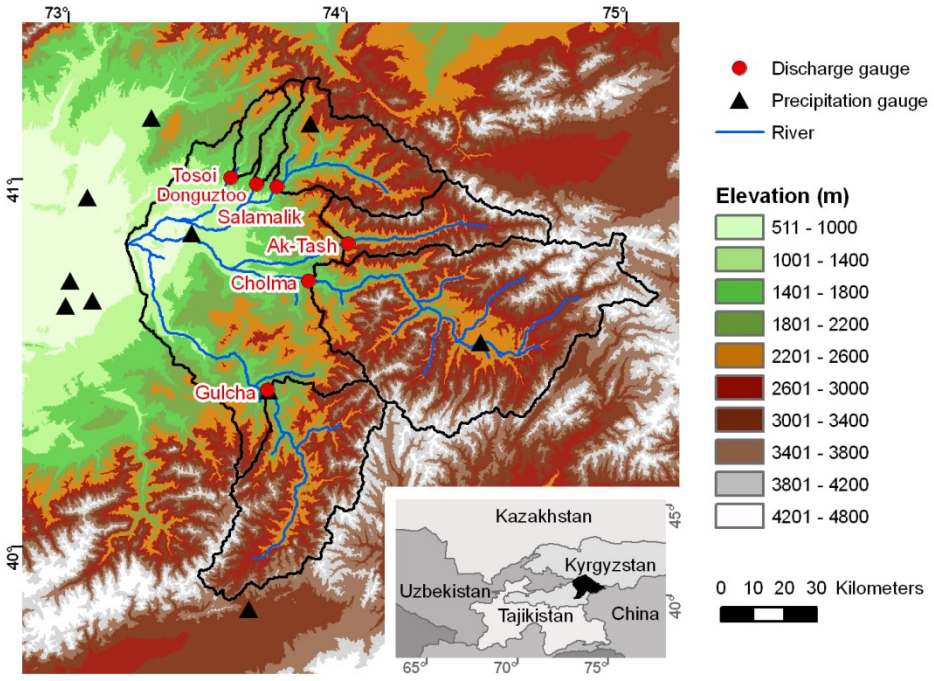

Fig. 3.1: The six study catchments in the Karadarya Basin, including elevation, discharge gauges, and precipitation gauges in the area. The inlet shows the location of the upper Karadarya Basin in Kyrgyzstan.

Table 3.1: Area, glacier coverage, elevation range, mean annual runoff, and runoff coefficient of the studied subcatchments of the Karadarya Basin.

	Area	Glacier coverage	Elevation (m)			Runoff*	Runoff coefficient**
	(km^2)	(%)	min.	max.	mean	(mm a^{-1})	(-)
Tosoi	216	0.0	1253	3165	2001	432	0.43
Donguztoo	166	0.0	1271	3502	1999	505	0.46
Salamalik	1180	0.5	1288	4381	2592	585	0.55
Ak-Tash	907	2.3	1728	4752	3121	778	0.62
Cholma	3840	1.9	1352	4753	3117	410	0.50
Gulcha	2010	0.7	1557	4623	3013	267	0.38

* Mean annual runoff over the period 1961–1990.
** Calculated for the period 1961–1990 using precipitation estimates by Duethmann et al. (2013).

3.3 Data and methods

3.3.1 Extraction of snow cover from AVHRR satellite imagery

Snow cover data were extracted from images of the AVHRR instrument onboard the NOAA-9 and NOAA-11 satellites. The AVHRR sensor provides images with a resolution of 1.1 km at nadir. Scenes were selected from the melt seasons (March–July) of the years 1986–1989, when both AVHRR data and suitable input data to drive the hydrological model were available. We focused on the melt season, as the parameters of the hydrological model were expected to be most sensitive to observed snow cover during this time. The time step between the images was approximately 9–14 days, but larger time steps were sometimes necessary in order to avoid high cloud coverage and poor data quality.

Before the images could be employed for snow cover delineation, careful calibration was required, including the calculation of albedo and radiances from the raw data as well as corrections for the sensor degradation and for nonlinearities in the measurements. In order to correct for the sensor degradation of channel 1 and 2, the calibration formulae of Rao and Chen (1995) were applied. Nonlinearities of the radiance measurements in channel 4 and 5 were adjusted using the corresponding error values from the NOAA Polar Orbiter Data User's Guide (National Climate Data Center, 1998). Finally, radiances of channel 3–5 were converted to temperatures using the inverted Planck's radiation equation, and the Normalized Differential Vegetation Index (NDVI) was calculated from channels 1 and 2.

The data were then classified into snow, no snow, and clouds using a dichotomous multichannel classification scheme based on a model developed by Voigt et al. (1999). Inputs to this model were differences in surface temperature, albedo, and NDVI, which is used for the extraction of the vegetation coverage. Initial values for the thresholds were taken from Höppner and Prechtel (2002). On the basis of visual interpretation of the scenes, these thresholds were subsequently adapted to the study region and to the respective season with different values for spring and summer. The seasonal variation proved to be particularly important for the correct discrimination between snow and clouds.

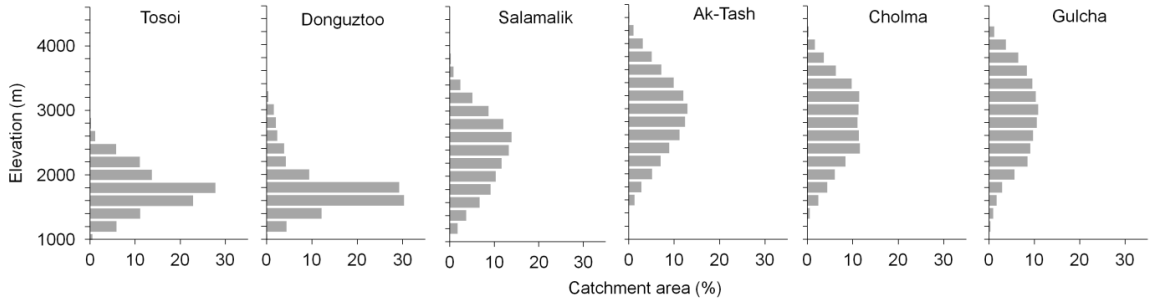

Fig. 3.2: Distribution of the catchment area with altitude for the six study catchments (by 200 m elevation zones).

3.3.2 Hydrological model

The hydrological model WASA (Model of Water Availability in Semi-Arid Environments) (Güntner, 2002; Güntner and Bronstert, 2004) was originally developed for large semiarid basins, and later extended for erosion and sediment transport (Mueller et al., 2010) and for mountainous regions influenced by snow and glacier melt (Duethmann et al., 2013).

The model uses a semidistributed approach. The spatial discretization may either be based on hillslopes (Güntner and Bronstert, 2004; Francke et al., 2008), which also allows for lateral surface and subsurface redistribution along the hillslope, or on hydrological response units (HRUs), which enables faster simulations. This study applied the latter approach. HRUs were delineated based on 200 m elevation bands, resulting in 11–18 HRUs with a median size of 51 km^2 and a range of 0.01–452 km^2. Each HRU was associated with its dominant soil type, dominant land cover type, and its glacier fraction.

The model includes routines for snow accumulation, snow and glacier melt, interception, infiltration, percolation through a multilayer soil, and evapotranspiration. Runoff processes considered by the model are infiltration and saturation excess surface runoff, interflow, and groundwater runoff. The model version used in this study separates subsurface flow into interflow and base flow based on a calibration parameter, and also considers that in a small fraction of the catchment (e.g., riparian areas, roads or rock areas connected to a stream) rainfall directly results into streamflow. The following section presents details of the snow module, detailed descriptions of other model components can be found in Güntner (2002) and Güntner and Bronstert (2004).

For the calculation of snow accumulation, precipitation is separated into rainfall or snowfall based on a calibrated threshold temperature. Snowmelt is simulated using a temperature index approach (Hock, 2003), where the melt factor varies in a sinusoidal form between a minimum value at the winter solstice and a maximum value at the summer solstice. The increase of the melt factor from winter to summer reflects an increase in incoming solar radiation and a decrease of the snow albedo associated with aged snow (Anderson, 2006).

Due to wind drift, avalanches, and the spatial variability of precipitation, snow is rarely uniformly distributed after snowfall, and variability in the melt processes (for example, due to variations in shading) may further increase the variability of SWE within an elevation zone, which is however often neglected in hydrological models. Neglecting this variability of SWE within an elevation zone or HRU implicates that an elevation zone can only be snow covered or snow free, resulting in abrupt changes from completely snow-covered conditions to completely snow free at the end of a melt season. Furthermore, the comparison to remotely sensed snow cover is easier if also fractional snow cover areas are simulated. In the WASA model, the variability of SWE within an elevation zone is parameterized using a snow depletion curve, as for example described by Liston (2004). The snow depletion curve in this case describes the fractional SCA as a function of the SWE divided by the maximum SWE at the end of the accumulation season (snow depletion curves may also be defined in other ways, for example, SCA as a function of time). During accumulation a spatially continuous snow cover is simulated; snowmelt then results in a gradually decreasing SCA. It has been observed that for a given catchment the spatial distribution of relative snow amounts is similar from year to year so that the shape of the snow depletion curve can be assumed constant in time (e.g., Luce and Tarboton, 2004). The distribution of the observed SWE is often approximated by a lognormal distribution (Donald et al., 1995). SCA can then be calculated analytically from the maximum simulated SWE at the end of the accumulation season and the cumulative melt depth (Liston, 2004). This parameterization for the description of the fractional SCA has only one additional parameter, the coefficient of variation of the SWE distribution. It assumes that within each elevation zone there is always some part with a very thin snow cover so that with the beginning of snowmelt a snow-free area is created. For cases where this is not considered appropriate, one could introduce an additional parameter defining the SWE that must melt before any snow-free area starts to be exposed (Donald et al., 1995). The situation that snowfall occurs during the melting phase needs to be considered separately. In this case, the model simulates a SCA of 100% and melt affects the whole area until the new snow is melted; the model then proceeds on the established snow depletion curve. If the sum of the remaining snow from the main accumulation period and the new snow exceed the previous maximum accumulation, the model considers this to be a new accumulation period and a new snow depletion curve is started. For HRUs containing a glacier, it is assumed that snow remains longest on the glacier so that glacier melt does not start until the snow-covered area in that HRU is smaller than the glacier area.

This version of the model has 13 calibration parameters (Table 3.2). The parameter ranges were established based on literature values and previous experiences with the WASA model. The glacier melt parameter is recalculated into a glacier melt factor with values ranging between the snowmelt factor and an upper boundary of 15 mm $°C^{-1}$ d^{-1}.

The following input data were used to set up the WASA model for the Karadarya Basin: the SRTM (Shuttle Radar Topography Mission) digital elevation model (Jarvis et al., 2008) with a resolution of 90 m, the MODIS land cover product at a resolution of

Table 3.2: Calibration parameters including values for the lower and upper bounds.

Routine	Parameter	Unit	Lower bound	Upper bound
Snow and glacier melt	min. snowmelt factor	mm °C^{-1} d^{-1}	1	15
	max. snowmelt factor	mm °C^{-1} d^{-1}	1	15
	glacier melt parameter	-	0	1
	threshold melt temperature	°C	−2	2
	coefficient of variation (*cv*)	-	0.001	1
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	200	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	-	0.75	1.5

500 m (MCD12Q1) (Friedl et al., 2002), mean monthly leaf area index (LAI) values from the 8 day MODIS LAI product for 2001–2008 at a resolution of 1 km (MOD15A2) (Myneni et al., 2002) aggregated by elevation zone, subcatchment and land cover class, a soil map (1 : 1 500,000) digitized from the Kyrgyz Atlas (Academy of Science of the Kyrgyz SSR, 1987), and glacier areas delineated from a Landsat Multispectral Scanner (MSS) scene (resolution 79 m) in summer 1977.

A previous analysis evaluated various precipitation estimates for this area (Duethmann et al., 2013) and concluded that the precipitation product interpolated from gauge data using monthly precipitation fields derived by multilinear regression against elevation, longitude, and latitude was best suited for this area. This precipitation product ("MLR-all" in Duethmann et al., 2013) was therefore also applied in the current study. The model further uses daily time series of solar radiation, temperature, temperature lapse rate, and humidity. Due to the lack of observational data these data were derived from the ERA-40 reanalysis data (Uppala et al., 2005) downscaled to a resolution of 12 km using the regional climate model Weather Research and Forecasting Model (WRF) (Skamarock et al., 2008). For details of the WRF simulations please refer to Duethmann et al. (2013). The meteorological time series were applied to the hydrological model as area-mean values for each catchment.

Discharge data used for model calibration were provided by the Hydrometeorological Service of Kyrgyzstan. Water stage is recorded twice daily manually and continuously at the larger rivers under high flow conditions using a float driven recording sensor. Discharge is estimated about four times per month during low flow and eight times per month during high flow conditions using the velocity-area method with velocities derived from current meter measurements. Data at the gauge Gulcha are influenced by diversions into an irrigation channel upstream of this gauge. As timing and volume of these abstractions were not known, they could not be considered in the model.

3.3.3 Model calibration

3.3.3.1 Objective functions

For the calibration to discharge data, the following objective function was used:

$$objf\_q = 0.5 \times (NSE + LogNSE),$$

$$\text{with } NSE = 1 - \frac{\sum_{t=1}^{T}\left(Q_{obs}(t) - Q_{sim}(t)\right)^2}{\sum_{t=1}^{T}\left(Q_{obs}(t) - \overline{Q_{obs}}\right)^2} \text{ and} \quad (1)$$

$$LogNSE = 1 - \frac{\sum_{t=1}^{T}\left(\log(Q_{obs}(t)) - \log(Q_{sim}(t))\right)^2}{\sum_{t=1}^{T}\left(\log(Q_{obs}(t)) - \overline{\log(Q_{obs})}\right)^2}.$$

NSE is the Nash-Sutcliffe efficiency value between daily observed ($Q_{obs}(t)$) and simulated ($Q_{sim}(t)$) discharge and *LogNSE* is the Nash-Sutcliffe efficiency calculated on logarithmic flows. As the Nash-Sutcliffe efficiency is particularly responsive to errors in high discharge values, and the Nash-Sutcliffe efficiency for logarithmic flows is more sensitive to errors in low flows, an average of these two measures results in a more balanced evaluation of high and low flows. The maximum possible value of the objective function

is 1, which would indicate perfect agreement between simulated and observed discharge.

Previous studies used different methods for evaluating model performance of semidistributed models with respect to raster-based remote sensing snow cover. Several studies simply compared catchment average simulated and observed SCA (Engeset et al., 2003; Udnaes et al., 2007; Sorman et al., 2009). This approach however neglects the information of the SCA distribution with elevation, which is also contained in the snow cover data. The method suggested by Parajka and Blöschl (2008) takes account of the observed SCA in each elevation zone. As the hydrological model they applied simulated a uniform distribution of SWE in each elevation zone, they used an indirect approach for the comparison between simulated SWE and observed SCA. They defined two kinds of errors—a snow underestimation error if the model simulated no snow in an elevation zone but the snow-covered area was larger than a threshold, and a snow overestimation error if the simulated SWE was above an SWE threshold but the remote sensing data indicated no snow for this elevation zone. A disadvantage of this approach is the subjective choice of the thresholds, which can influence the results of the comparison. For the snow underestimation error it is for example not clear to what value the threshold for observed SCA should be set. A SCA threshold of, e.g., 0.05 means that the model has to simulate at least some snow (SWE > 0) if the observed SCA is larger than 5%, otherwise this produces an underestimation error. By assuming a uniform SWE distribution, this however, also has the consequence that the model generates snowmelt from the entire elevation zone, while in reality snowmelt is only produced from the much smaller snow-covered area. In the present study, this problem was avoided by introducing the parameterization for the fractional SCA in an elevation zone as described in section 3.3.2, which then allowed a direct comparison between simulated and observed snow cover fraction for each elevation zone. While introducing such a parameterization increased the number of parameters by one, it also allowed a more realistic representation of the fractional SCA, which could potentially also improve the runoff simulations.

For the comparison to simulated snow cover, the AVHRR snow cover data were summarized by catchment and elevation zone. An AVHRR cell (1.1 km × 1.1 km) was assigned to an elevation zone according to its median elevation, calculated from the 90 m SRTM digital elevation model. In a next step, the fractional snow and cloud cover by catchment and elevation zone were calculated. This approach thus particularly considers the large-scale heterogeneity at a scale larger than the AVHRR cell size. Observed snow cover values were only considered for model comparison if >20% of that elevation zone was cloud free and if there were more than three cloud-free cells. The objective function with respect to snow cover was then defined as root-mean-squared error (RMSE) between simulated and observed SCA:

$$objf\_sca = \sqrt{\left(\sum_{i=1}^{m} n(i)\right)^{-1} \cdot \sum_{i=1}^{m}\sum_{j=1}^{n(i)} \left(SCA_{obs}(i,j) - SCA_{sim}(i,j)\right)^2}, \quad (2)$$

where $SCA_{obs}(i,j)$ and $SCA_{sim}(i,j)$ are the observed, respectively, simulated, SCA on observation day i in elevation zone j; m is the number of days with snow cover observations; and $n(i)$ the number of elevation zones with valid observations on day i. The best possible value of this objective function is 0, which would indicate perfect agreement between simulated and observed SCA.

Due to the formulation of the two objectives, multiobjective model calibration aims at identifying parameter sets which maximize *objf_q* and minimize *objf_sca*.

3.3.3.2 Multiobjective calibration to determine Pareto fronts

As the optimum solutions for different objectives in general do not converge, multiobjective optimization aims at identifying a set of Pareto optimal solutions instead of one single best solution. A solution is classified as Pareto optimal (or nondominated) if there is no other solution which improves in one or more objectives without degrading at least one objective. Evolutionary algorithms are particularly suited to solve such multiobjective problems, as due to their population-based approach they can return a set of solutions within a single run (Konak et al., 2006). For this study, we applied the Epsilon-Dominance Nondominated Sorted Genetic Algorithm II (ε-NSGAII) (Kollat and Reed, 2006). This algorithm was selected, as in comparison studies which also included hydrological model calibration in their test problems, this algorithm was competitive or superior to other state-of-the-art multiobjective algorithms (Kollat and Reed, 2006; Tang et al., 2006). ε-NSGAII is based on the NSGAII-algorithm (Deb et al., 2002), which uses a fast nondominated sorting algorithm and elitism. Elitism means that the nondominated solutions found so far are preserved and survive to the next generation. As an extension compared to the original NSGAII-algorithm, ε-NSGAII introduced the concept of ε-dominance, adaptive population sizing and automatic termination, reducing the number of algorithm parameters to be tuned. ε-dominance allows the user to specify the required precision in each objective. The user should set it to the difference in objective function values he or she considers to be relevant. The objective space is divided into multidimensional cells, with the dimension according to the number of objectives and the cell size in each dimension according to the ε-value in this objective. If there is more than one

solution within a cell, only one solution—in case of minimization of all objectives, the solution closest to the origin—is retained. This way the density of the final nondominated solutions is controlled. In a series of connected runs, the population size is adapted according to the number of archived nondominated solutions. A new run is started if there is no significant increase in the number or quality of the nondominated solutions over a selected number of generations. The overall search is terminated if there is no significant improvement between the archived solutions of two successive runs. For this study, we used an initial population size of 16 and a maximum run time of 40,000 generations. ε-values were set to 0.005 for *objf_q* and 0.001 for *objf_sca*. Other algorithm parameters were set to the values as suggested by Kollat and Reed (2006). In order to reduce run time, a parallel version of the code (Tang et al., 2007b) was applied.

Snow cover data only were available for the period 1986–1989. As a two year period was regarded as being too short for model calibration with respect to discharge, and the model could not be run in the 1990s due to missing data, the model was first calibrated in simulation period A (1978–1981 and 1986–1987) and validated in simulation period B (1982–1985 and 1988–1989), and then calibration and validation period were swapped (calibration in simulation period B and validation in simulation period A). Thus, calibration and validation periods usually included two years of snow cover and six years of discharge data. Simulations were always performed for a continuous period. For example, for calibration in period A, simulations were performed for the period 1976–1987. From these simulations, *objf_q* was evaluated over 1978–1981 and 1986–1987, and *objf_sca* was evaluated over 1986–1987. Due to gaps in the discharge data (21 May to 31 December 1980 in Ak-Tash and 1989 in Salamalik) the calibration or validation periods were accordingly shorter in these two catchments. An additional two year period prior to the actual simulation period was used for model initialization.

3.3.3.3 Monte Carlo parameter calibration for investigating the value of an increasing number of snow cover scenes

The additional information content of an increasing number of snow cover scenes was explored by calibrating the model using different subsets of snow cover images during the calibration period and evaluating the model performance against the snow cover images in the validation period. This resulted in a very large number of different calibration settings, as different possibilities of, e.g., selecting 5 snow cover images from 20 available images in the calibration period have to be considered. Optimizing the parameters for each subset of observations was therefore not a suitable approach. While for a small number of calibration settings multiobjective calibration usually results in better model performances with a lower number of simulations, Monte Carlo simulations can be more efficient if the number of different subsets of observations to which the model should be calibrated is very large. After once performing a large number of simulations, it is then possible to simply select well-performing simulations for a large number of different calibration settings without further computational cost. Assuming uniform distributions between the minimum and maximum bounds, 50,000 random parameter sets were generated using Latin hypercube sampling.

For each catchment, the RMSE of observed versus simulated SCA of each image was evaluated. Snow cover images for which the range of RMSE values over the 50,000 simulations was zero were not considered in the next step since they did not exert any constraining power. These were generally images where only elevation zones far below or above the snowline were cloud free and which were thus not challenging for the model. In the next step, the value of increasing the number of snow cover scenes was evaluated. Calibration and validation period were defined as described in section 3.3.3.2 and all simulations where *objf_q* in the calibration period was above 0.5 were selected. Now $n = 0, 1, \ldots m$ images were randomly chosen from the calibration period, where m denotes the number of available images in the calibration period. In order to calculate the median, minimum, and maximum value of *objf_sca* during the validation period, the following procedure was applied (summarized in Fig. 3.3). For each simulation, *objf_sca* with respect to the selected images in the calibration period was calculated and the best $s = 20$ parameter sets were selected (for $n = 0$, s parameter sets were chosen randomly). Selecting an ensemble of best performing parameter sets instead of selecting just one best parameter set is expected to result in a more robust behavior during the validation period. For the selected

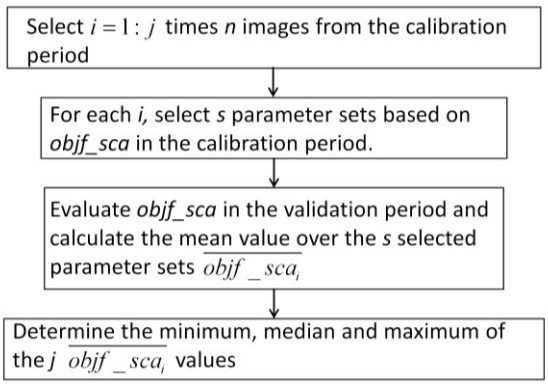

Fig. 3.3: Procedure for calculating the minimum, median, and maximum snow cover error during the validation period, when using n snow cover images for model calibration.

simulations, the model performance over the validation period was evaluated and averaged over the selected simulations. This was repeated j = 10,000 times in order to sample different combinations of selecting n from m images. Finally, the median, minimum, and maximum *objf_sca* of these 10,000 repetitions were calculated. The procedure depicted in Fig. 3.3 was repeated for $n = 0, 1, \ldots m$ images. It was also performed for the two different calibration periods (calibration in period A and validation in period B, and vice versa).

3.4 Results and discussion

3.4.1 Characterization of the trade-offs between good performance for discharge and snow cover area

Using a multiobjective evolutionary algorithm, it was investigated how well discharge and SCA can be modeled simultaneously. Overall the model achieved good performances with objective function values for discharge between 0.62 and 0.93 (optimum = 1) and objective function values for SCA between 0.1 and 0.29 (optimum = 0). Plots of the Pareto fronts show that the trade-offs between good simulations with respect to discharge and SCA are generally small in the studied catchments (Fig. 3.4). In this figure, the red and blue dots indicate the solutions for calibration period A and B, respectively. Optimum solutions plot in the lower left corner. Only for the catchment Gulcha, a larger spread of solutions can be seen, for all other catchments, only up to six solutions were found. Solutions which differ by less than the specified ε-values of 0.005 for *objf_q* and 0.001 for *objf_sca* were not retained by the algorithm, avoiding overly precise trade-offs, which are not meaningful (Kollat et al., 2012). Thus in most of the studied catchments a good model performance in terms of SCA did not preclude a good simulation of discharge. The model performances in the validation periods are close to the performances if the respective period was used for calibration (Fig. 3.4), except for the catchment Gulcha, where the model seems to be less transferrable between different periods. The peculiarities at Gulcha with overall lower performances, a larger trade-off between discharge and snow cover performance, and lower transferability between the calibration and validation period are likely to be caused by the unconsidered flow diversions into an irrigation channel upstream of this gauge.

We also compared the performance of the calibrated model to an uncalibrated model which simply uses mean parameter values between the upper and lower

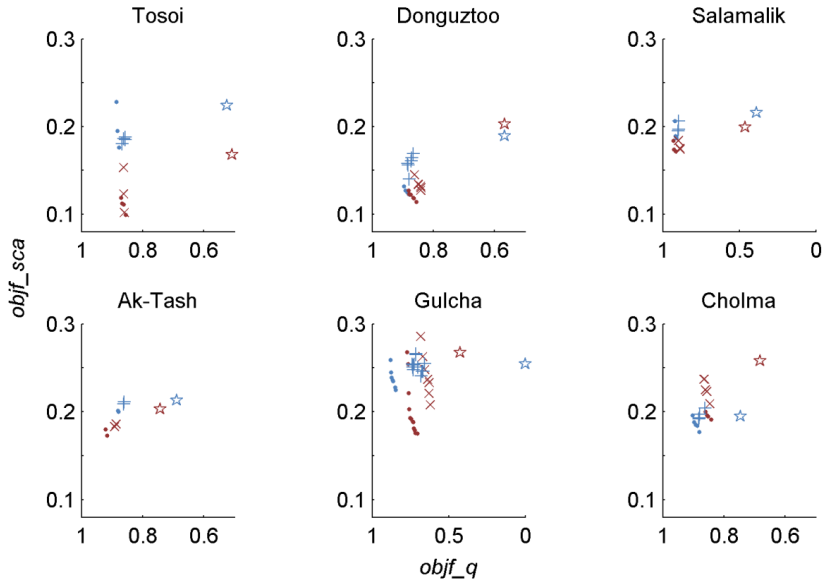

Fig. 3.4: Trade-off curves of model performance against SCA (*objf_sca*) and discharge (*objf_q*) for the six study catchments. The *x* axis is plotted in reverse order so that optimum solutions plot in the lower left corner. Dots: calibration period; crosses: validation period; blue: simulation period A (1978–1981 and 1986–1987); red: simulation period B (1982–1985 and 1988–1989). For example, blue dots indicate the model performance in simulation period A of the solutions from the calibration in simulation period A, and blue crosses show the model performance in simulation period A of solutions from the calibration in simulation period B. Additionally, the stars show the performance of the uncalibrated model. Please note the different scales of the *x* axes.

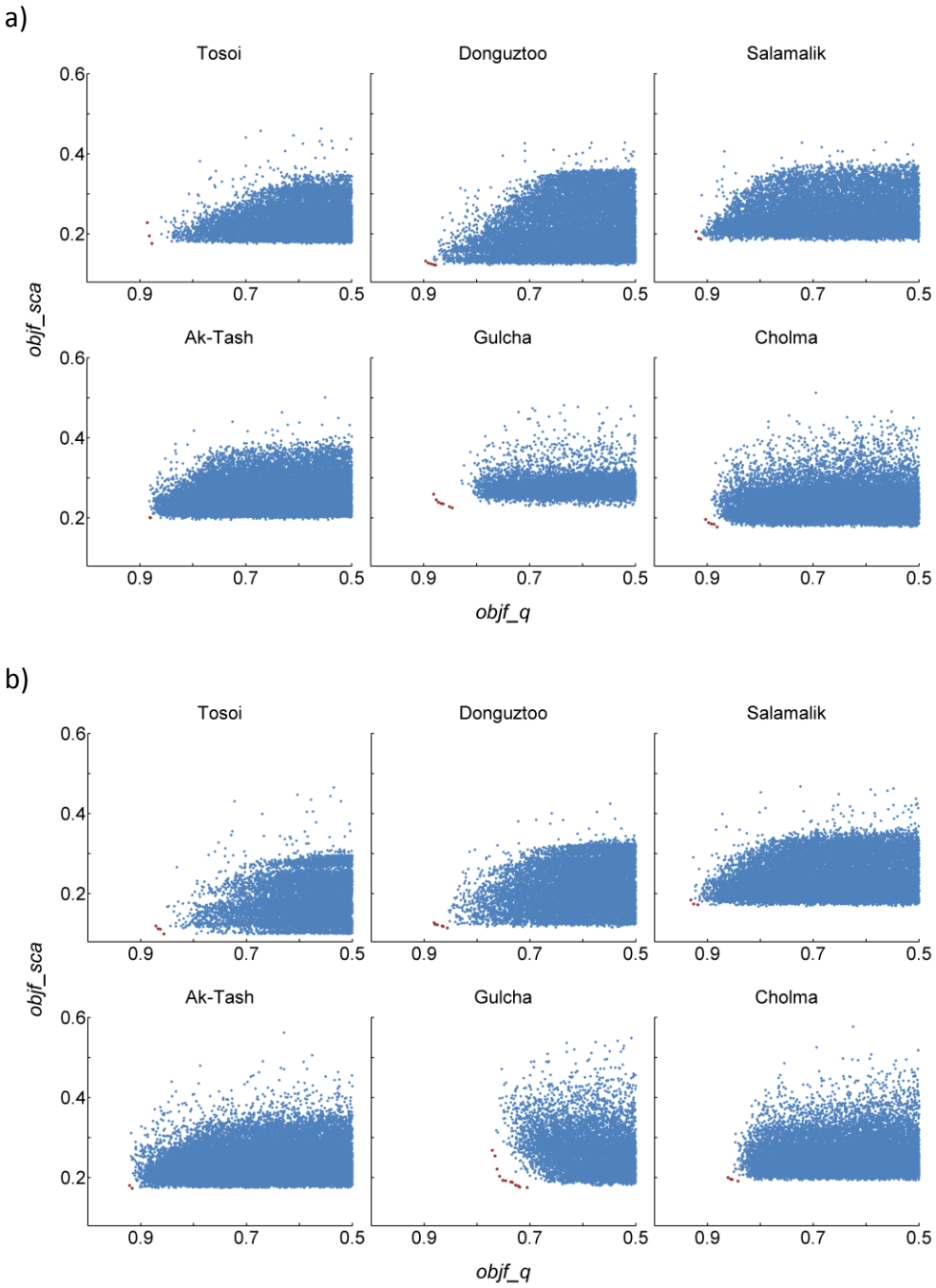

Fig. 3.5: Scatterplots of the objective function with respect to SCA (*objf_sca*) against the objective function with respect to discharge (*objf_q*) showing both the solutions generated by Monte Carlo sampling (blue) and the solutions from the multi-objective optimization algorithm (red). The *x* axis is plotted in reverse order so that optimum solutions plot in the lower left corner. a) Calibration in simulation period A (1978–1981 and 1986–1987), and b) calibration in simulation period B (1982–1985 and 1988–1989).

parameter bound of Table 3.2 (Fig. 3.4). This shows that in all cases model performance was improved by calibration, and on average, *objf_q* improved by 0.33 and *objf_sca* improved by 0.03 (Fig. 3.4).

The multiobjective optimization algorithm always outperformed the best solutions of the Monte Carlo simulations with randomly sampled parameters (Fig. 3.5). The number of model evaluations used by the multiobjective optimization was between 4168 and 12,472 with an average of 6988. Thus, even though the average number of model evaluations in the multi-objective optimization was about seven times smaller than the number of Monte Carlo simulations applied in this study, better solutions were achieved through multiobjective optimization. The scatterplots in Fig. 3.5 also show that it is possible to achieve good simu-

lations with respect to discharge with a large snow cover error and vice versa. This was also confirmed by single-objective optimizations where the model was calibrated only against discharge or only against SCA. For the criterion which was included in the optimization (either discharge or snow cover area), they resulted in virtually the same model performance as the best performing solution with respect to this criterion from the multiobjective optimization, but only much lower performances were achieved for the neglected criterion. Namely, for the optimizations against SCA only, *objf_q* was on average 0.62 worse than the worst solution from the multiobjective optimization; and for the optimizations against discharge only, *objf_sca* degraded on average by 0.05. It is therefore necessary to consider both objectives for model conditioning.

3.4.2 Model performance in terms of discharge and snow cover area

In addition to the objective function values, which assess the model performance in an aggregated way, this section provides a more detailed analysis of the model performance with a particular focus on SCA prediction. Time series of simulated and observed fractional SCA can be compared for each elevation zone in each subcatchment (see examples in Fig. 3.6). As the trade-offs between the solutions from the multiobjective optimization were only small, the figure depicts only one compromise solution from the multiobjective optimization indicated by the black line. In many cases, the observed decrease in snow cover is well represented by the model. Not only start and end point in time of snow cover depletion but also the fractional snow cover often compares well to the observed one. The model has some difficulties with snow events during the melt season. Possible reasons for this are uncertainties in the precipitation input, the precipitation phase (solid or liquid), and the model assumptions for snow events during the melt season. The interpolated precipitation is only based on relatively few stations so that the timing of precipitation events cannot always be representative for the entire catchment. Uncertainties in the precipitation phase are due to uncertainties in the temperature and lapse rate, as well as the fact that the temperature below which precipitation falls as snow is not a fixed value but may vary from event to event. If there are snow events during the melt season, the simulated SCA of the elevation zones with snowfall is increased to 100% (assuming that snowfall covers the entire elevation zone, see section 3.3.2) and only decreases again after this new snow has completely melted, though in reality this new snow might also disappear more gradually.

In order to visualize the difference between simulations with comparable performance in terms of *objf_q* but different performance in terms of *objf_sca*, Fig. 3.6 also shows results from the Monte Carlo sample, selected as the five best and the five worst simulations in terms of *objf_sca* from all solutions with $objf\_q \geq objf\_q_{max} - 0.05$, where $objf\_q_{max}$ is the best Monte Carlo solution in terms of *objf_q* achieved for this catchment. The solutions with a poor performance

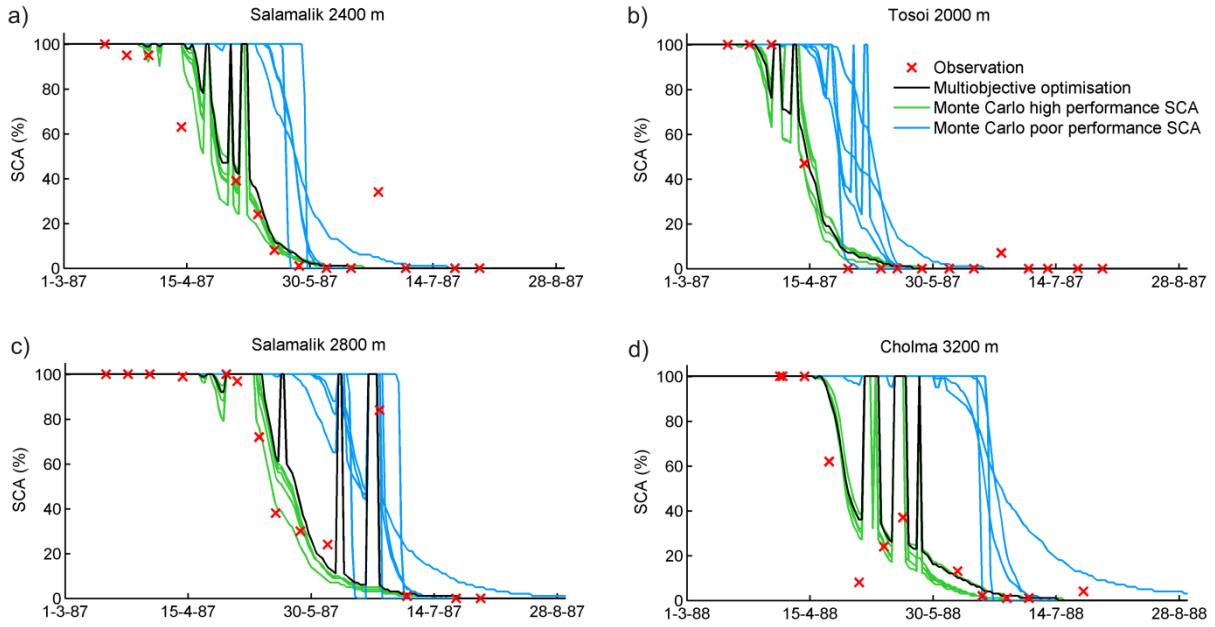

Fig. 3.6: Time series of simulated versus observed SCA during the validation period, for selected catchments and elevation zones. The simulations show a compromise solution in terms of snow cover and discharge performance from the multiobjective optimization, and solutions from the Monte Carlo sample with comparable performance in terms of discharge but contrasting performance in terms of SCA.

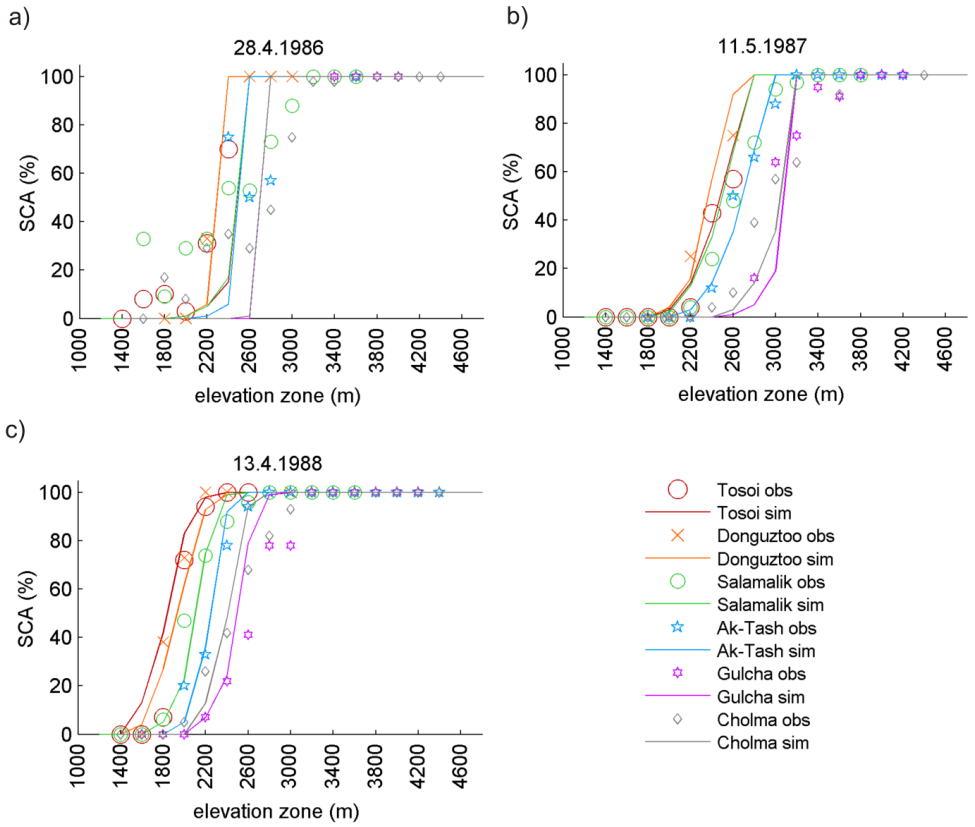

Fig. 3.7: Simulated and observed SCA in the six study catchments, for the 28 April 1986, 11 May 1987, and 13 April 1988. Simulated snow cover is shown for the validation period and a selected compromise solution generated with the multi-objective optimization algorithm. The simulations also refer to discrete 200 m elevation zones but are drawn as continuous lines for better visibility.

considerably overestimate the snow cover. Despite a comparable performance with respect to discharge, large errors in the simulation of the snow cover can occur if snow cover observations are not considered during model calibration.

From a different perspective, Fig. 3.7 shows simulated and observed SCA for all elevation zones of all study catchments as examples for three selected days. These plots generally show a consistent increase of the observed SCA with elevation with only little noise. A typical pattern, apparent on many dates, is that the elevation zone where for example around 50% of the area is snow covered is at a higher elevation in Cholma and Gulcha than in Tosoi and Donguztoo (Fig. 3.7 b and 3.7 c). The faster decrease of the snow cover in Cholma and Gulcha may partly be explained by lower precipitation in these two catchments compared to Tosoi and Donguztoo. Differences in temperature, aspects or shading may furthermore explain the differences in snow cover decrease between the catchments. In accordance with the observations, the simulations also represent the gradual increase of SCA with elevation over several elevation bands. On some dates, the model shows a very abrupt increase of snow cover from 0% to 100% over only one or two elevation zones not reflected by the observations (e.g., 28 April 1986). This behavior can again be related to snow events during the melt season.

In order to summarize the behavior over all images, the mean bias by elevation zone averaged over all dates within the validation period was evaluated. The black dots in Fig. 3.8 show the result for a compromise solution from the multiobjective optimization. This generally reveals only relatively small systematic biases with elevation. In the catchments with a larger elevation range, there is a tendency for overestimation at high elevations (e.g., Salamalik and Ak-Tash for both periods (Fig. 3.8 a and 3.8 b) and Cholma and Gulcha for the model calibrated in period B (Fig. 3.8 b)). Moreover, in the catchment Gulcha, SCA is overestimated at low elevations if the model is calibrated to simulation period A and validated in simulation period B (Fig. 3.8 a). In this case, the number of observations was very low; there are only four observations for the 1600 m elevation zone in the catchment Gulcha in the period 1986–1987 so that an overestimation of SCA at two of these dates caused a large average error. Fig. 3.8 additionally also illustrates the bias by elevation zone for the selected sets of solutions from the Monte Carlo sample with comparable

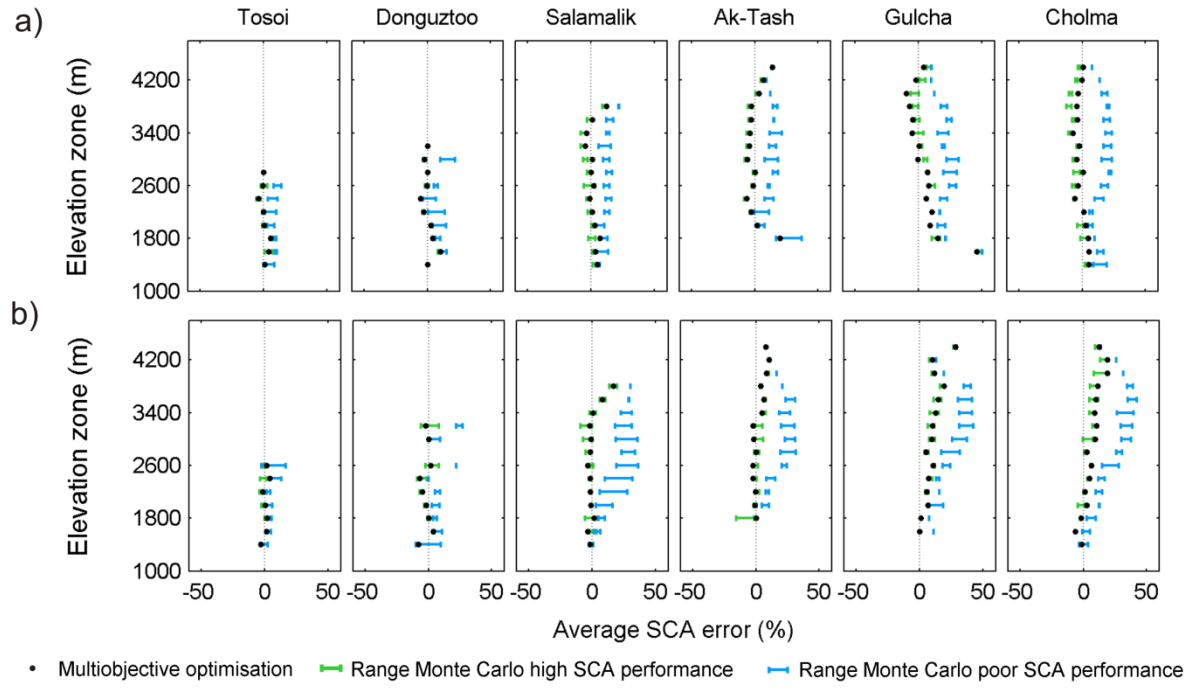

Fig. 3.8: SCA bias by elevation zones averaged over the validation periods (top) 1986–1987 and (bottom) 1988–1989. The black dots show results for a compromise solution in terms of snow cover and discharge performance from the multiobjective optimization. Bars indicate ranges of values from selected Monte Carlo simulations with comparable performance in terms of discharge but contrasting performance in terms of SCA.

performance for discharge, but contrasting performance in terms of SCA. The simulations with a poor performance in terms of SCA generally show an overestimation of snow cover. This overestimation is however smaller in the catchments Tosoi and Donguztoo. In contrast to this, the selected solutions from the Monte Carlo sample with a good performance in terms of snow cover only have low bias values, comparable to the solutions from the multiobjective optimization.

As a result of the small trade-offs, the simulated hydrographs of the two extreme solutions from the multiobjective optimization which result in the best model performance for discharge and SCA, respectively, are most of the time very similar (results not shown). Therefore, Fig. 3.9 only shows compromise solutions that are good both in terms of discharge and snow cover. The time series indicates that the degradation between calibration and validation period in Gulcha is to a large part the result of an underestimation of discharge in the years 1988–1989 of the model calibrated in simulation period A (1978–1981 and 1986–1987) and an overestimation in the years 1986–1987 of the model calibrated in simulation period B (1982–1985 and 1988–1989). These differences are likely a result of different abstraction volumes for irrigation during the two periods. Fig. 3.9 also demonstrates some other deficiencies of the model, like too low variability of the simulated discharge in Salamalik, or a relatively strong overestimation of discharge of the base flow during the winter period 1987–1988 in Cholma. However, overall the model performs well, particularly considering the sparse data availability in this region.

As an example, Fig. 3.10 also shows simulated discharge for solutions from the Monte Carlo sample, selected as the five best and five worst solutions with respect to *objf_sca* of all solutions with $objf\_q \geq objf\_q_{max} - 0.05$, for one year in the catchment Ak-Tash. The shown example is typical also for other catchments and other years. The most striking difference between the two sets of simulations is that, compared to the solutions with poor SCA performance, the solutions with good SCA performance generally result in higher discharge at the beginning and lower discharge toward the end of the melting period. Thus, the simulations with a good performance with respect to snow cover have a higher tendency to overestimate the observed discharge at the beginning of the melting season and to underestimate toward the end, while this is the other way round for the simulations with a high snow cover error. Overall, this results in a comparable performance with respect to discharge for the two sets of solutions.

3.4.3 Influence of the two objective functions on constraining model parameters

In order to illustrate the effect of the two different objectives on constraining the different model parame-

ters, sets of solutions which perform well with respect to the two objective functions were selected from the Monte Carlo sample and evaluated with respect to their parameter distributions. For each catchment and for the two simulation periods, first the 5% best performing solutions in terms of *objf_q* were selected from the Monte Carlo sample. From these solutions, the 5% best performing solutions in terms of *objf_sca* were retained. Fig. 3.11 shows cumulative density functions of the resulting parameter distributions. In these plots, uniform prior distribution would appear as straight lines. Note that the minimum and maximum melt factors were swapped when the minimum melt factor was larger than the maximum melt factor so that the prior distributions are not uniform. The two

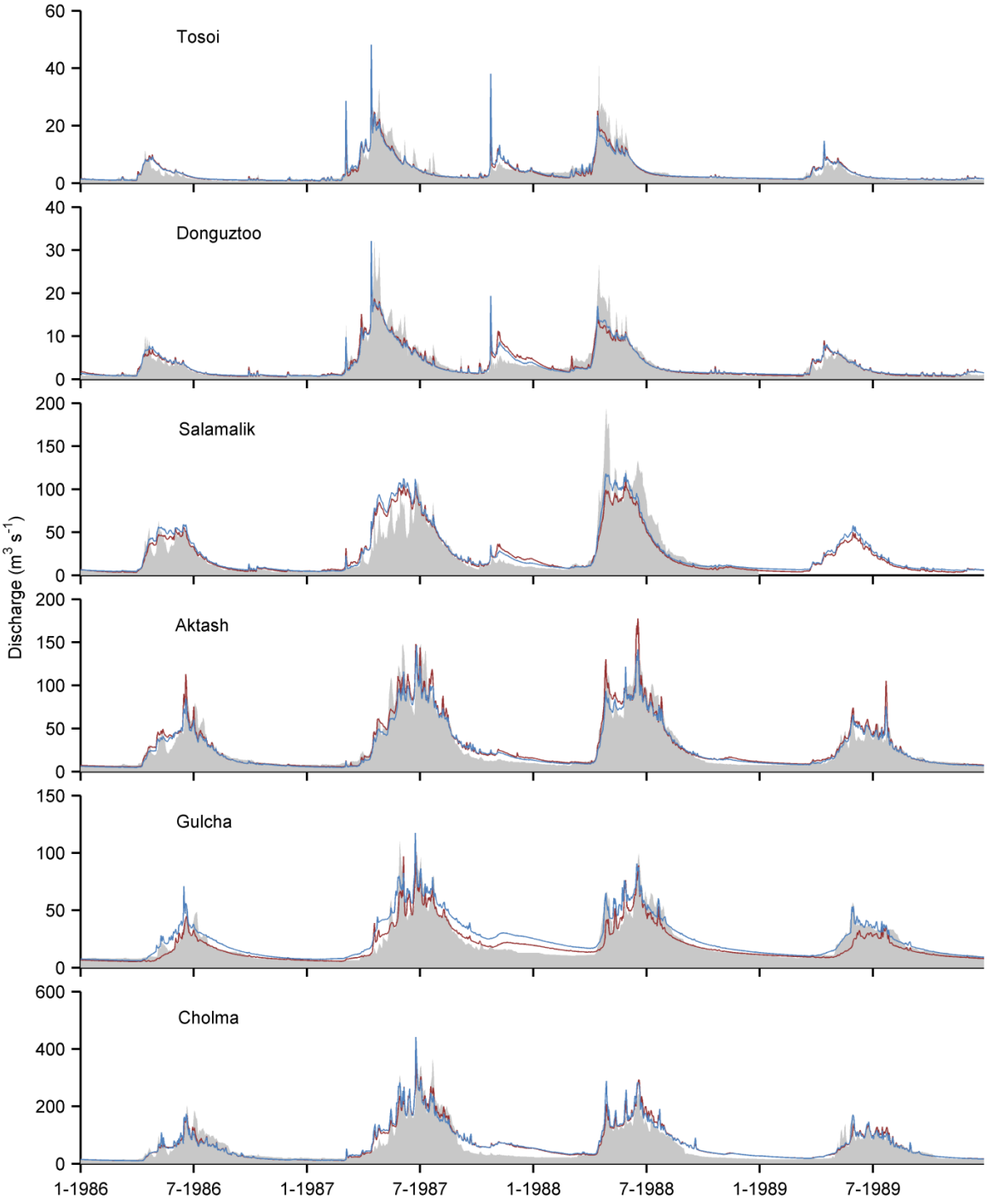

Fig. 3.9: Time series of simulated versus observed discharge for a compromise solution (good performance for discharge and SCA) from the multiobjective optimization algorithm and the time period 1986–1989. Gray area: observed discharge; red line: simulated discharge for model calibration in simulation period A (1978–1981 and 1986–1987); blue line: simulated discharge for model calibration in simulation period B (1982–1985 and 1988–1989).

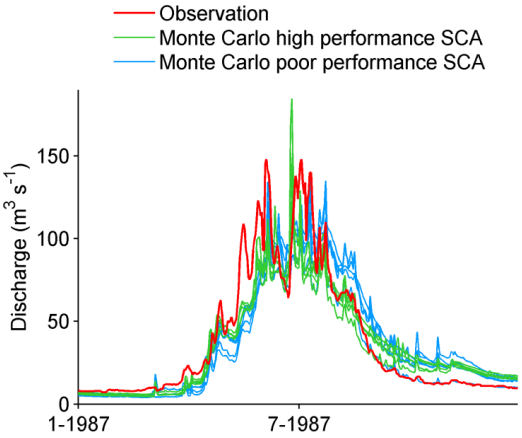

Fig. 3.10: Observed discharge and simulated discharge, for the catchment Ak-Tash and the year 1987. Simulated discharge is shown for selected parameter sets from the Monte Carlo sample with overall comparable performance in terms of discharge, but good and poor performance with respect to SCA.

snowmelt factors and the threshold melt temperature are conditioned by both objectives. The coefficient of variation (*cv*), which determines the heterogeneity of the snow distribution, is largely controlled by *objf_sca*. Most other parameters, like *frac2gw*, the interflow delay factor, *sat_area_var*, and the precipitation bias factor are controlled by *objf_q*. Furthermore, some of the model parameters, in particular the glacier melt parameter and *kf_corr_f* hardly get constrained. If one selects all solutions with a comparable discharge performance ($objf\_q \geq objf\_q_{max} - 0.05$) and then contrasts the five best and five worst simulations in terms of SCA (as also done for Fig. 3.6, Fig. 3.8, and Fig. 3.10), the solutions with a good SCA performance are generally characterized by larger values for the parameter *cv*, higher melt factors, and in some cases (Cholma and Gulcha) lower threshold melt temperatures than the simulations with a poor SCA performance. Additionally, the solutions with low SCA performance have in some cases also lower values for the interflow delay factor, indicating that the too late snowmelt is compensated by faster subsurface transport.

Considering how the various parameters impact the model response, the model parameters are affected by the different objectives in a plausible way. A large variability in the parameter values is typical in many hydrological modeling applications, where often very different parameter sets may lead to similar performance of the objective function values (Beven and Binley, 1992). The parameter distributions indicate that this problem can to some extent be alleviated when in addition to observed discharge also the performance in terms of SCA is considered. This is, for example, clearly demonstrated for the parameter *cv*. In other cases, adding the snow cover criterion may lead to a shift in the distribution, with possibly only small effects on further constraining the distribution. The main advantage of taking into account snow cover data should however be seen in improving model consistency.

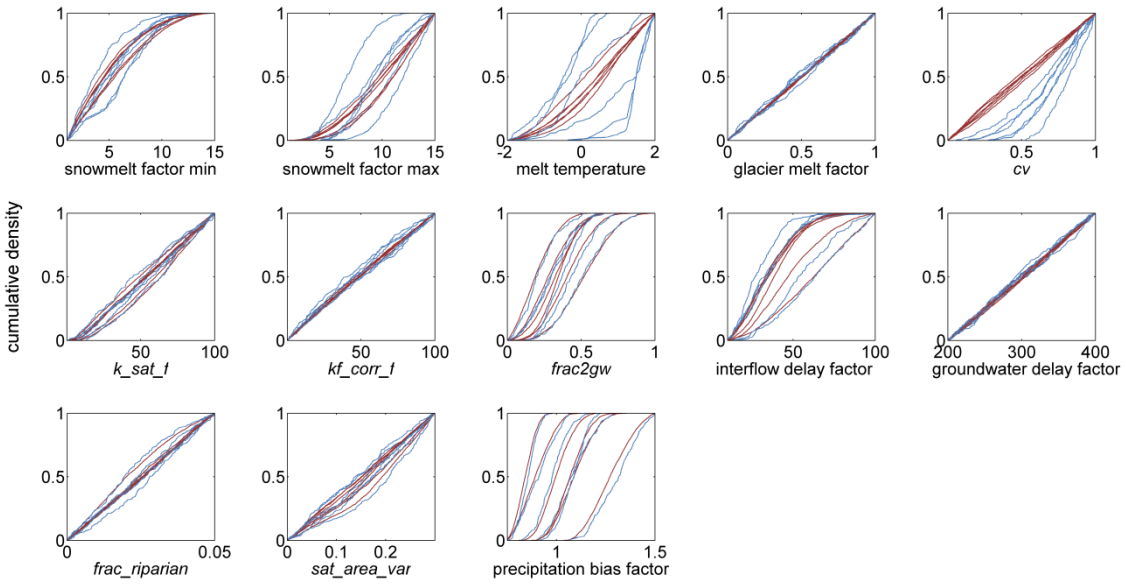

Fig. 3.11: Cumulative distribution functions of parameter sets selected from the Monte Carlo sample as the 5% best performing solutions in terms of *objf_q* (red lines) and after further constraining these solutions selecting the 5% best performing solutions in terms of *objf_sca* (blue lines) for calibration to simulation period A. The six lines indicate the different catchments (for more details please refer to the text).

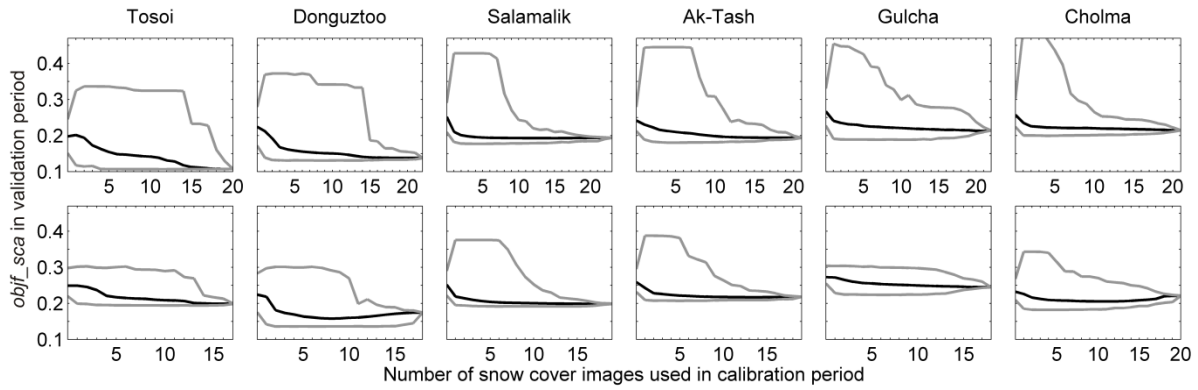

Fig. 3.12: Value of the objective function with respect to SCA (*objf_sca*) in the validation period as a function of the number of snow cover images *n* used in the calibration period. The black line shows the median over 10,000 repetitions of selecting *n* images, the gray lines show the minimum and maximum, respectively. (top row) Results for calibration to simulation period A and validation in simulation period B, and (bottom row) results for the reversed case.

3.4.4 Value of increasing the number of snow cover observations during the calibration period

In order to determine how many snow cover images are necessary for successful model calibration, we investigated how the snow cover objective function in the validation period changed with the number of snow cover images in the calibration period. Generally, the median SCA error continually decreased with an increasing number of snow cover images in the calibration period (Fig. 3.12). This decrease was strongest over the first 1–4 mages, while there was little further decrease when increasing the number of images from 10 to the maximum possible number of images in the calibration period (depending on catchment and period, 17–22 images).

However, a larger number of images was necessary to reduce the maximum SCA error. If the number of images is too low, the identified parameter sets may not be representative. A model calibrated to all images is likely to over- or underestimates individual observations, due to deficiencies of the model (including inputs and parameters) or errors in these observations. A small sample can in the worst case therefore be dominated by images where such a well-calibrated model would for example always overestimate the SCA. In such a case the SCA images would shift the model in the wrong direction. The maximum SCA error was higher when using few images in the calibration period than when using no image and randomly selecting 20 parameter sets. The reason for this is that there is a very large number of possibilities to draw 20 simulations from the available simulations (all simulations where the objective function value with respect to discharge during the calibration was above 0.5) so that it is very unlikely to draw the 20 worst simulations by random chance. An image which indicates an overestimation of the model, although the model actually overall underestimates SCA, is thus much more effective in selecting simulations with poor performance with respect to SCA. In our study, in most catchments around 10 images were necessary to reduce the maximum error below the value for the case where no image was used.

In the catchments Tosoi and Donguztoo up to 16 images were necessary to reduce the maximum error below the level where no image was used for model calibration. A possible reason why a larger number of images was needed in these two catchments is the lower elevation range (Fig. 3.2), resulting in a lower number of data points per image. The number of data points per images is determined by the number of elevation zones with maximum 80% cloud cover and more than three cloud-free cells (see section 3.3.1). The average number of data points per image in Tosoi and Donguztoo was 6.2 and 5.2, respectively; in contrast it was between 10.7 and 13.4 in the other catchments. Furthermore, many images in Tosoi and Donguztoo did not allow locating the transitional elevation zones between snow and snow-free areas, as they showed only snow-free or only snow-covered elevation zones. It is likely that images with a lower number of observations and images which do not allow locating the transitional elevation zones between snow and snow-free areas have a lower constraining power. Therefore, more images may be necessary to compensate the effect of images which tend to draw the model in the wrong direction.

Fig. 3.13 illustrates the effect of individual images on *objf_sca* in the validation period when drawn as first, second, or third image, shown as the median change in *objf_sca* in the validation period. The example depicts the catchment Salamalik with simulation period A for model calibration. When using only one image for model calibration, most images result in an improvement of *objf_sca* in the validation period. In contrast, there are some images (22 March 1986, 1 April 1986,

16 March 1987, and 24 June 1987) which cause a large deterioration of *objf_sca* in the validation period. The effect of such images can be compensated when further images are added. Generally, the influence of individual images declines with an increasing number of images. While the median changes in *objf_sca* in the validation period, as shown in Fig. 3.13, are mostly only small, the maximum positive or negative change can be much larger, since the influence of an image added as the second or third image of course also depends on which other image(s) are used for constraining the model. For example an image which has little constraining power when added as the first image, can in combination with particular other images still result in an improvement of the model performance in the validation period.

Snow cover scenes can only be useful for model calibration if there is variability in the simulated snow cover at that time. Therefore, the unconstrained model may be used for identifying suitable time periods in which snow cover data may be beneficial for model calibration. Fig. 3.14 shows the range of simulated SCA of 1000 Monte Carlo simulations (generated using Latin hypercube sampling and the parameter ranges from Table 3.2) which have not been constrained by any observational data. Early in the melt season, data points at high elevations have no value in constraining the model, as there is no variability in the simulations. However, these snow cover scenes deliver useful observation points at low to medium elevations. In contrast, snow cover scenes in the late melt season deliver useful data points at medium to high elevations. Such relatively fast initial simulations, which can be performed a priori before snow cover data for model calibration is available, can be used for identifying variability hotspots in the simulations. Based on this, one can then select suitable time periods for which snow cover data should be processed.

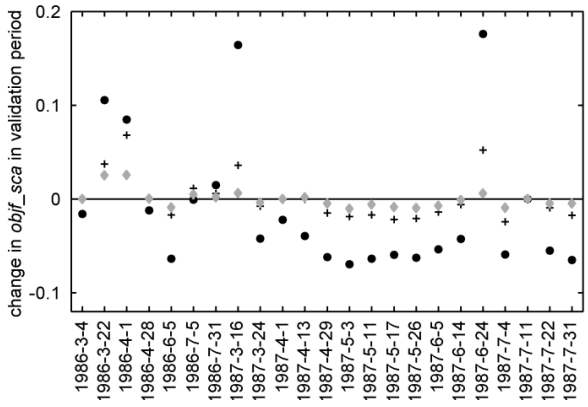

Fig. 3.13: Median effect of individual images on *objf_sca* in the validation period when drawn as first (filled dots), second (crosses), or third (gray diamonds) image, for the catchment Salamalik with simulation period A for model calibration.

3.5 Conclusions

This study evaluates the benefit of satellite-derived snow cover images for the calibration of a hydrological model in snow-dominated catchments in Central Asia. In most of the catchments, we found only small trade-offs between good simulations with respect to discharge and SCA. However, if the parameters were selected based on the discharge objective function only, this could lead to simulations with large snow cover errors. The fact that good discharge simulations were also achieved with large snow cover errors and thus for the wrong reasons demonstrates very clearly that SCA observations should be taken into account for model calibration in order to achieve higher internal consistency of the model. Using a parameterization for fractional SCA, allowed a direct comparison between simulated and observed SCA for each elevation zone without requiring any thresholds for the comparison of simulated SWE to observed SCA. This approach could also be advantageous for other semidistributed models, where the comparison to observed satellite-derived SCA is generally less straight-forward than in raster-based models.

While the average SCA error was already reduced with only few images, it is recommended to use a larger number of images (10–16 SCA scenes in this study) for model calibration in order to also reduce the maximum error in case of an unfavorable selection of satellite scenes. Our results should be seen as a starting point toward a more general understanding of the value of increasing the number of snow cover images for model calibration. As the six catchments investigated in this study are located in the same region, the results should at this point only be transferred to catchments with a similar physiographic setting and snowfall regime. Further studies should investigate how the results change with changes in the physiographic setting (e.g., catchments with a much lower elevation range) or snowfall regime (e.g., many snow events also during the melt season, or not one distinct snow season but rather a number of shorter snow events over the winter period). As a further step, it should also be explored whether there are characteristics of snow cover images (e.g., particular patterns of snow cover) which make them particularly useful for model calibration. It might for example be important that the set of images contains scenes with the snowline at low and at high elevations.

We recommend a wider application of SCA data for model calibration if snowmelt is an important runoff generation process in the catchment of interest. This study demonstrated that good discharge simulations could also be attained with large snow cover errors if snow cover data were not taken into account for calibration. As satellite-derived SCA data are available globally, considering these data for calibration is a very good opportunity to improve hydrological modeling also in remote, data sparse areas. This is particu-

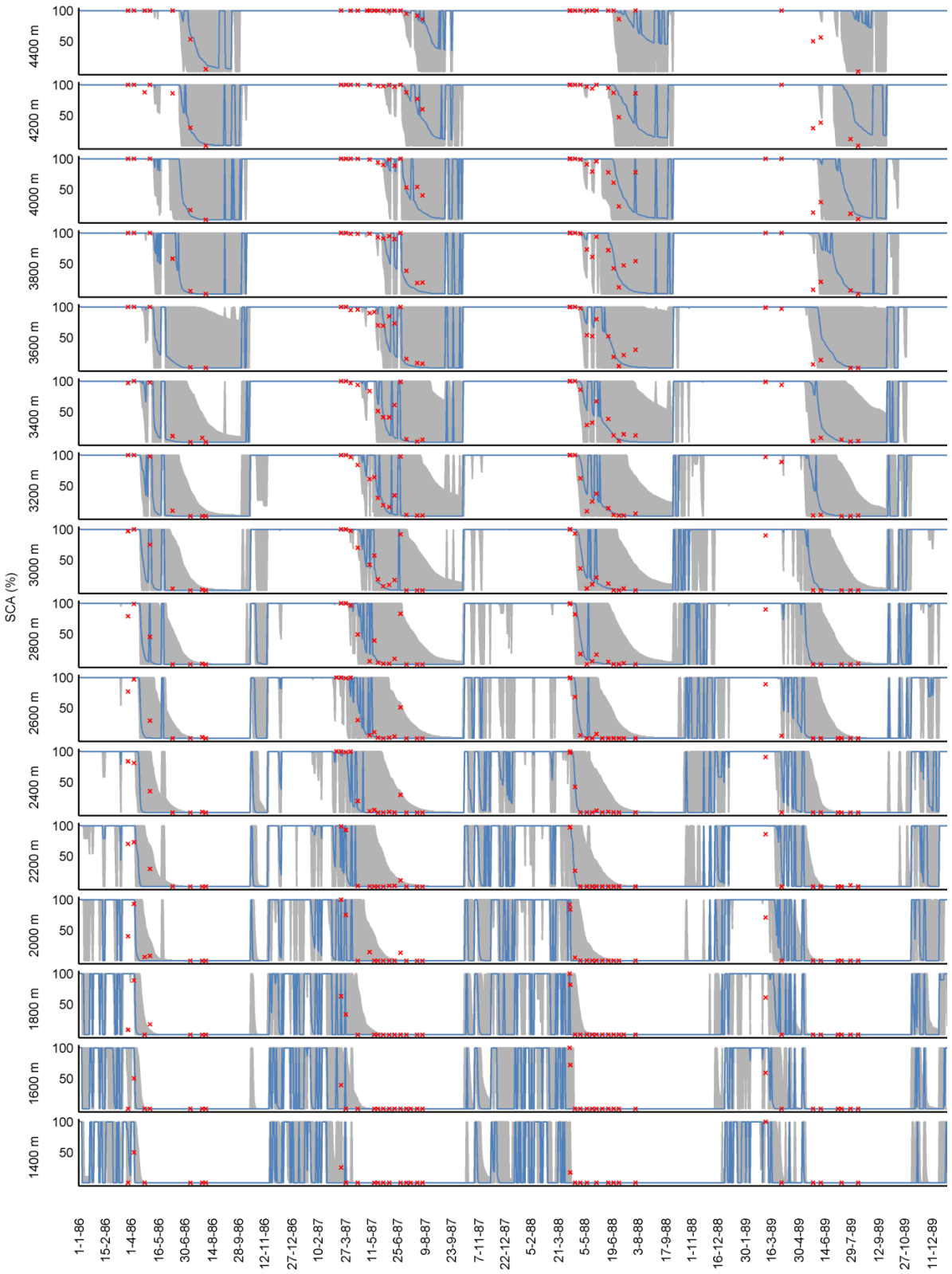

Fig. 3.14: Simulated snow cover ranges of 1000 Monte Carlo simulations of each elevation zone over the time period 1986–1989 for the catchment Cholma (gray-shaded area). Red crosses show the snow cover observations. The blue line shows simulated snow cover for the validation period from a balanced solution of the multiobjective optimization.

larly important if the model is applied for an analysis of hydrological processes or for climate change scenarios.

Acknowledgments

This study was conducted within the projects CAWa (Water in Central Asia; http://www.cawa-project.net), funded by the German Federal Foreign Office (grant AA7090002), and SuMaRiO (Sustainable Management of River Oases along the Tarim River; http://www.sumario.de/), funded by the German Federal Ministry of Education and Research (grant 01LL0918A). We thank Patrick Reed and Joshua Kollat from the Pennsylvania State University for providing the ε-NSGAII algorithm and Marina Köhler for her assistance with the computing cluster. The Central Asian Institute for Applied Geosciences (CAIAG) digitized soil and discharge data and the Hydrometeorological Service of Kyrgyzstan provided discharge and meteorological data. We would also like to thank the associate editor and two anonymous reviewers for their comments and suggestions which helped improving this manuscript.

4 Attribution of streamflow trends in snow- and glacier melt dominated catchments of the Tarim River, Central Asia

Abstract

Streamflow of headwater catchments of the Tarim River (Central Asia) increased by about 30% over the period 1957–2004. This study aims at attributing these trends in streamflow to changes in air temperature and precipitation. The analysis includes a data-based approach using multilinear regression, and a simulation-based approach using a hydrological model. The hydrological model considers changes in both glacier area and surface elevation. It was calibrated using a multiobjective optimization algorithm with calibration criteria based on glacier mass balance and daily and interannual variations of discharge. The individual contributions to the overall streamflow trends from changes in glacier geometry, temperature, and precipitation were assessed using simulation experiments with a constant glacier geometry, and with temperature and precipitation time series without trends. The results showed that the observed changes in streamflow were consistent with the changes in temperature and precipitation. In the Sari-Djaz catchment, increasing temperatures and related increase of glacier melt were identified as the dominant driver, while in the Kakshaal catchment, both increasing temperatures and increasing precipitation played a major role. Comparing the two approaches, the evidence provided by the simulation-based approach was regarded as stronger than the one provided by the data-based approach since the latter was hampered by correlations between the explanatory variables and the fact that it is based on statistical links instead of process-based cause-effect relationships. A complementary application of both approaches is recommended in order to evaluate and enhance the reliability of results from trend attribution studies.

Published as:

Duethmann, D., T. Bolch, D. Farinotti, D. Kriegel, S. Vorogushyn, B. Merz, T. Pieczonka, T. Jiang, B. Su, and A. Güntner (2015): Attribution of streamflow trends in snow- and glacier melt dominated catchments of the Tarim River, Central Asia, *Water Resour. Res., 51*(6), 24.

Note: The manuscript underwent modifications during the review process. Please refer to the journal for the final version.

4.1 Introduction

In mountain regions, snow and glaciers play a crucial role in runoff formation. This makes them one of the most vulnerable environments to climate change, as temperature changes affect the ratio of rain versus snowfall and snow- and glacier melt (Birsan et al., 2005). This study focuses on streamflow variability in the Aksu catchment, located in the Central Tien Shan in Kyrgyzstan and the Xinjiang Uyghur Autonomous Region in China (Fig. 4.1). The Aksu River is the most important tributary to the Tarim River, contributing about 80% of its streamflow (as calculated from average streamflow at the last gauges of the Aksu, Hotan, and Yarkand before their confluence forms the Tarim River, for the period 1964–1988). These water resources are of eminent importance for the population, industry, agriculture, and natural vegetation in the Tarim Basin, where the only significant water supply is via streams from the surrounding mountain regions. Several studies report increasing discharge of the Aksu headwater streams (e.g., Zhang et al., 2010; Krysanova et al., 2014). While it is assumed that these increases are related to concurrent trends in temperature, causing higher glacier melt, and precipitation (Chen et al., 2009; Zhang et al., 2010; Tao et al., 2011; Xu et al., 2011), a thorough attribution analysis is still missing.

The analysis of streamflow trends in mountain catchments influenced by snow and glacier melt has recently gained more attention from the scientific community, and a number of studies also attempt to attribute the detected changes to their possible drivers. The link between trends in streamflow and possible climatic drivers may be investigated by two groups of methods—data-driven or simulation-driven (Merz et al., 2012). Data-based approaches directly relate changes in runoff to changes in climate without the additional step of applying a hydrological model. This has the advantage that uncertainties from parameterization and structure of the hydrological model are avoided. Data-based approaches include the investigation of concurrent trends in climate variables (Birsan et al., 2005; Pellicciotti et al., 2010; Kriegel et al., 2013), correlation analyses (Dahlke et al., 2012), approaches based on multiple linear regression (Aguado et al., 1992; Dettinger and Cayan, 1995; Moore and Demuth, 2001; Stahl and Moore, 2006), or on generalized linear models (Bates et al., 2010). Further, climate variables have been used as covariates in time-varying distributions of streamflow (Delgado et al., 2014; Villarini and Strong, 2014).

Trend attribution in mountain catchments is complex due to the fact that one has to consider the influences of variations in temperature and precipitation, in seasonal snow storage, and in glacier mass balance and area at different elevations (Molnar et al., 2011). For example, concurrent trends in winter streamflow and winter temperature may be found, indicating that there are more snowmelt events or an increasing fraction of rain as compared to snow during winter. However, winter temperatures may still be well below 0°C in the largest part of the catchment and during most of the winter so that the increase in temperature would not be sufficient to explain the streamflow increase. In order to estimate whether it is physically plausible that the temperature increase caused the streamflow increase, one would have to extrapolate the temperature to different elevations in the catchment and estimate the area where rainfall or snowmelt could have occurred. This may be easier incorporated within a hydrological modeling approach. Furthermore, a hydrological model can also incorporate other available information (apart from streamflow data), such as estimates of glacier mass balances or snow cover time series. Hydrological modeling thus provides a framework for testing whether the hypotheses on causes for changes are plausible and whether different types of information are consistent with each other. A further advantage of the simulation-based approach is the possibility of analyzing trends in components of the water cycle which are not or cannot be measured individually (Hamlet et al., 2007), such as trends in soil moisture or glacier melt.

Despite these advantages, there are only few studies using simulation-based approaches for attributing streamflow trends in snow/glacier melt dominated catchments to their possible causes. Examples are Engelhardt et al. (2014), who used a glaciohydrological model to explain discharge changes between different time periods, or Hamlet et al. (2005), who applied a hydrological model for attributing the decline of snow water equivalents over the western United States. Using Global Climate Models runs with and without anthropogenic forcing, Hidalgo et al. (2009) were able to attribute shifts toward earlier streamflow timing in the western United States to anthropogenic climate change. Zhao et al. (2013) investigated the runoff increase in the Aksu catchment by applying the Variable Infiltration Capacity model, and they attributed the runoff increase mostly to an increase in precipitation. However, their study did not consider parameter uncertainties, and observations on mass balances were not taken into account for model calibration. In such a region with large uncertainties in the precipitation data, this can lead to wrong conclusions, as an underestimation of precipitation can be compensated in the model by an overestimation of glacier melt (Stahl et al., 2008; Schaefli and Huss, 2011).

Negative glacier mass balances result in reductions of the glacier area and lowering of the glacier sur-

face elevation. These changes have a feedback on the glacier mass balance: The reduction in glacier area in the ablation zone results in less glacier melt and thus a relatively more positive glacier mass balance, while the lowering of the glacier elevation and the associated increase in surface temperature causes both increased glacier melt and reduced snow accumulation. An analysis of these effects for glaciers in Switzerland showed that the negative effect caused by glacier surface lowering compensated on average about 50% of the positive effect due to glacier area change (Huss et al., 2012). While changes in glacier geometry have been taken into account in hydrological modeling for climate impact analyses (e.g., Stahl et al., 2008; Huss et al., 2010; Farinotti et al., 2012), to our knowledge they have so far not been considered for the attribution of observed discharge trends.

The presented study aims at demonstrating a comprehensive trend attribution study in two snow and glacier melt dominated catchments. We revisit streamflow changes in the Aksu River and apply both a data-based approach using multiple linear regression analysis, and a simulation-based approach based on hydrological modeling. For the simulation-based approach, glacier area and elevation changes are taken into account, and multiple objectives are considered for model calibration, including measures on the temporal variation of the simulated glacier mass balance (based on a glacier mass balance time series of a nearby glacier), the cumulative glacier mass change (based on geodetic glacier mass balance estimates), the interannual variation of seasonal discharge, and the representation of discharge trends. The effects on streamflow trends caused by glacier geometry, temperature, and precipitation changes were investigated through simulation experiments with a constant glacier geometry, and temperature and precipitation time series without trends. This leads us to contrasting conclusions as compared to Zhao et al. (2013). Questions that motivate this study are (1) whether the observed changes in streamflow over the last decades are consistent with the changes in temperature and precipitation, (2) to what extent they can be explained by either changes in temperature (and glacier mass loss) or in precipitation, and (3) how these changes in runoff relate to changes in the input from rain, snowmelt and glacier melt.

4.2 Study area and observed hydrometeorological changes

The study area comprises two headwater catchments of the Aksu River, located in the Central Tien Shan (Fig. 4.1; Table 4.1). The lower part of the Aksu River is not considered as it hardly contributes to runoff generation and is strongly influenced by water management (Tang et al., 2007a). The Kakshaal catchment (also called Toxkan or Toshkan in China) upstream of the gauge Shaliguilanke (see Fig. 4.1 for location) has an area of 18,410 km^2. About 4.4% were covered by glaciers as revealed by Hexagon satellite imagery of the mid-1970s. With an area of 12,950 km^2, the Sari-Djaz catchment (also called Kumarik in China) upstream of the gauge Xiehela is about one third smaller, but has a higher glacierization (about 21% for the mid-1970s). Both catchments feature high elevations with mean elevations around 3600–3700 m, but the elevation range is higher in the Sari-Djaz (1450–7100 m) than in the Kakshaal catchment (1900–5900 m). The most dominant mountain range in the Kakshaal catchment is the Kakshaal-Too Range, which stretches through the whole catchment up to the south-eastern part of the Sari-Djaz catchment. Main mountain ranges in the Sari-Djaz catchment are the ones around Tomur Feng (Kyrgyz: Jengish Chokusu, Russian: Pik Pobedy) and Khan-Tengri in the east, the Teskey Ala-Too in the north, and the Ak-Shirak massif in the northwest (Fig. 4.1). Due to the high mountain terrain, land cover is dominated by grassland,

Table 4.1: Area, glacier coverage, elevation range, mean annual runoff, and mean annual precipitation of the two catchments. "Interpolated precip." is the areal precipitation estimate from interpolation of station-based observations, as described in section 4.3.1. The last column shows areal precipitation estimated from hydrological modeling (see section 4.3.4.1 and 4.4.2.1), as mean values and with the ranges over the selected model solutions in brackets.

	Area	Glacier coverage (~year 1975)	Elevation			Runoff (1957–2004)	Interpolated precip. (1957–2004)	Precip. estimated by hydr. modeling (1957–2004)
	(km^2)	(%)	min.	max.	mean	(mm a^{-1})	(mm a^{-1})	(mm a^{-1})
Kakshaal	18,410	4.4	1900	5900	3550	151	245	386 (372–399)
Sari-Djaz	12,948	20.9	1450	7100	3700	382	333	474 (450–526)

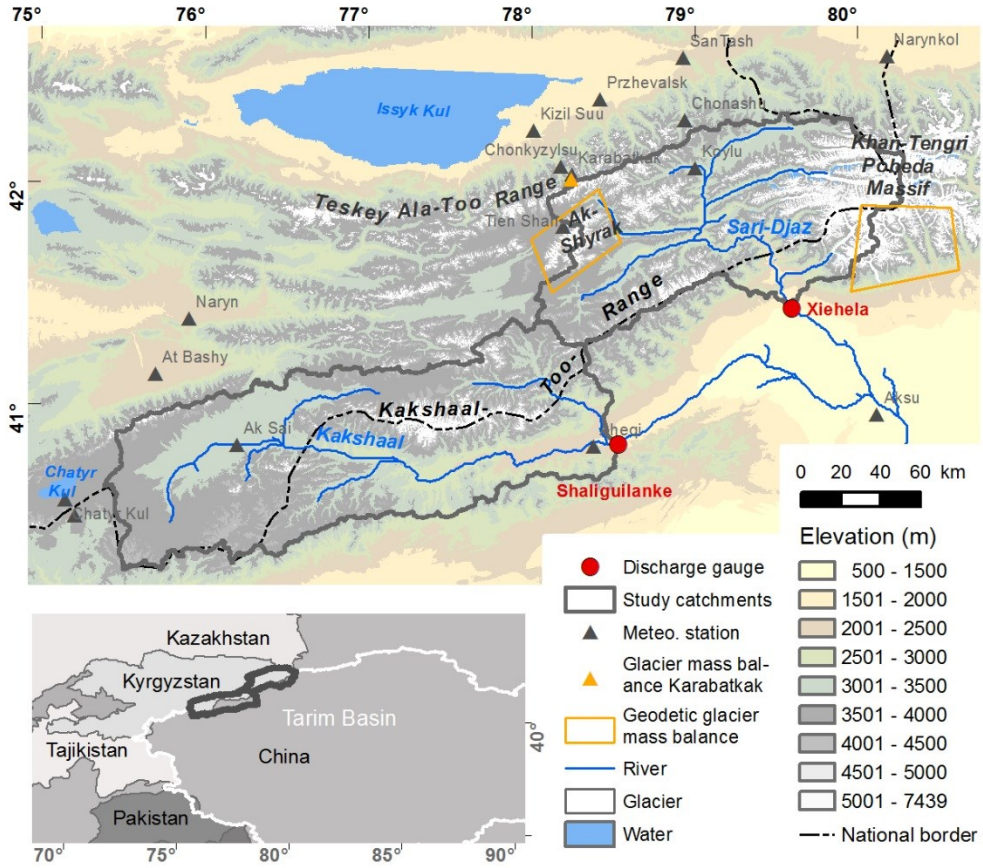

Fig. 4.1: The study catchments, including elevation, discharge gauges, meteorological stations, and locations of Karabatkak glacier and geodetic mass balance estimates. The inset shows the location of the study area (dark gray outline) in relation to national borders and the Tarim Basin (white outline).

barren or sparsely vegetated land, and snow and ice.

Average annual precipitation over the period 1960–1990 measured at precipitation gauges within the two catchments ranges from 190 to 320 mm a^{-1} (values without undercatch correction), but higher values are expected at exposed mountain locations. The precipitation regime shows a strong maximum in summer, typical for the Central and Southern Tien Shan (Aizen et al., 1995; Bothe et al., 2012). Average annual runoff over the period 1960–1990 was 140 and 360 mm a^{-1} for the Kakshaal and Sari-Djaz Basin, respectively (Wang, 2006). Maximum monthly discharge values occur in July (Kakshaal) and August (Sari-Djaz) and are caused by snowmelt, glacier melt, and the summer precipitation maximum. High discharge peaks at the Xiehela gauge can further result from quasi-periodic outbursts of the Merzbacher Lake, a proglacial lake, which refills every year (Glazirin, 2010). The outbursts occur about once a year, although years without or with two outbursts have been observed as well. The average volume of these floods was estimated at 170 × 10^6 m^3 (Jinshi, 1992; Wortmann et al., 2013), which corresponds to around 5% of the summer (JJA) discharge at Xiehela. There is no evidence of significant influence of water management in both catchments.

Observed changes in the hydrometeorological variables and glaciers have been described in a number of studies and are only briefly summarized here (for an overview, see Krysanova et al., 2014). Over the time period 1957–2004 discharge has increased by around 30% at the gauges Shaliguilanke and Xiehela (Xu et al., 2010; Kundzewicz et al., 2015). Over the same period, several studies showed a temperature and precipitation increase for stations in the Chinese part of the basin (Xu et al., 2010; Zhang et al., 2010; Fan et al., 2011; Tao et al., 2011), while stations close to the basin in Kyrgyzstan showed an increase in temperature but not in precipitation (Aizen et al., 1997; Aizen et al., 2006). Changes were also observed in glacier characteristics. Glacier area in the Tien Shan decreased over the last decades (see reviews in Sorg et al., 2012; Unger-Shayesteh et al., 2013). In the Central Tien Shan, from which the Aksu headwater streams drain, glacier retreat rates were generally lower

than in the outer ranges (Unger-Shayesteh et al., 2013). For the Chinese part of the Aksu Basin, Liu et al. (2006) reported a glacier area decrease of 3.3% for the period 1963–1999. For the period 1990–2010 and the Sari-Djaz Basin, Osmonov et al. (2013) found a glacier area loss of 3.7 ± 2.7%. For the Ak-Shirak massif, glacier area losses of 8.7% for the period 1977–2003 (Aizen et al., 2006), and 13.6% for the period from the mid-1970s to the mid-2000s and the part draining to the Naryn Basin (Kriegel et al., 2013) were reported.

Geodetic mass balances were available for two areas adjacent to our study area (see Fig. 4.1 for location). For the Ak-Shirak massif and the period 1977–1999, Surazakov and Aizen (2006) estimated glacier thickness changes of −15.1 ± 8.2 m, based on topographic maps and the SRTM3 DEM. Assuming an estimated ice density of 900 kg m^{-3}, this thickness change corresponds to a mass budget of −0.62 ± 0.33 m w.e. a^{-1}. For a region south of Tomur Peak and the period 1976–1999, Pieczonka et al. (2013) derived an average glacier mass change of −0.42 ± 0.23 m w.e. a^{-1} from differencing SRTM3 and Hexagon DEMs. Mass loss from Karabatkak glacier, close to the Aksu Basin in the Teskey-Ala-Too mountain range (Fig. 4.1), is in line with these geodetic estimates, and over the period 1977–1998 an average mass loss of −0.64 m w.e. a^{-1} was measured (Dyurgerov and Meier, 2005; WGMS, 2012).

Rigorous trend attribution studies should also comprise a search for possible alternative drivers, and evidence that these are inconsistent with the observed changes (Merz et al., 2012). Such alternative drivers could be changes in other meteorological variables than temperature or precipitation, in the water management, or in land cover. Measured time series for radiation, wind speed, or humidity were not available, and could therefore not be included in this analysis. Data for these variables were only available from reanalysis data, which, due to changes of the observing system over time, are less suited for trend-analyses (Bengtsson et al., 2004). Changes in water management were excluded since there was no evidence for significant influence of water management. A distinct change in the land cover was the reduction in glacier area and glacier surface elevation. This effect is estimated with the simulation-based approach. Other possible drivers were not known.

4.3 Data and methods

4.3.1 Streamflow and meteorological data

Daily discharge data for the gauges Xiehela and Shaliguilanke were available for the period 1964–1987 from the Hydrological Yearbook published by the Chinese Ministry of Water Resources. Monthly discharge data were available for 1957–2004 (Wang, 2006). The daily data were checked for outliers and jumps to remove possible digitization errors. In cases where monthly sums of daily data did not agree with the monthly data, more trust was given to the daily data.

Daily meteorological data with variable lengths over the study period 1957–2004 from stations within and close to the study area were obtained from the Kyrgyz Hydrometeorological Service. Daily data were available from 4 stations for temperature and 13 stations for precipitation. Further monthly station data (11 stations for temperature and 3 stations for precipitation) were downloaded from a collection of the AsiaCryoWeb group at the University of Idaho (http://www.asiacryoweb.org). The locations of the meteorological stations are shown in Fig. 4.1. Particular attention needs to be paid to the station Tien Shan, located at 3614 m in the Central Tien Shan. This station was relocated in 1998/99, and it is under debate whether this relocation resulted in inhomogeneous measurements (Mamatkanov et al., 2006; Kutuzov and Shahgedanova, 2009; Unger-Shayesteh et al., 2013). Monthly temperature anomalies of the station Tien Shan show a high correlation to neighboring stations, and visually no inhomogeneity could be identified from the comparison to temperature anomalies of neighboring stations. Identifying inhomogeneities in the precipitation data of the Tien Shan station is difficult, as the only other station in this region with a long precipitation time series which continues after 1999 is the Naryn station, located at a distance of 190 km and a considerably lower altitude. Comparing monthly precipitation anomalies of Tien Shan to those of Naryn indicated two possible inhomogeneities in 1983 and 1998, with a period of relatively low precipitation of the Tien Shan station between 1983 and 1998. As the correlation between the two precipitation time series is generally low and no reason for a possible inhomogeneity of the Tien Shan station in 1983 is known, it is not clear whether this is due to climatic variability or inhomogeneous measurements. Due to this ambiguity, and because the station Tien Shan is the only high elevation station for which daily precipitation data are available also after 1999, the station was finally kept in the analysis.

The temperature and precipitation station data were spatially interpolated to a 2 km × 2 km grid. When interpolated data are used for the analysis of trends, one needs to consider that a varying number of meteorological stations over time may lead to inhomogeneities of the resulting data set. However, in our study area, selecting only time series covering the whole period would leave only two stations and likely reduce the ability of a hydrological model driven with these data in representing the observed discharge. In order to use most of the available data but minimize the problem of possible inhomogeneities in the resulting data set, the temperature and precipitation data sets were interpolated based on anomalies to the reference period 1969–1988, selected due to good data availability (4 stations with monthly temperature and 13 stations with monthly precipitation time series). Individual missing precipitation (temperature) values in the monthly time series were filled assuming a constant ratio (difference) to the station with the highest correlation for that month. The monthly time series were then aggregated over all years of the reference period to average monthly values. Interpolation of the time series was performed as follows:

For temperature, monthly lapse rates and intercepts were estimated from the average monthly temperature values over the reference period using linear regression with elevation. An average monthly gridded climatology for the reference period was then estimated from a DEM and these lapse rates and intercepts. For the interpolation of the daily data for a particular day, (1) differences between the daily value and the average monthly temperature during the reference period were calculated for stations with available data for this day and monthly data during the reference period. (2) These anomalies were then interpolated using the inverse distance weighting (IDW) method. (3) The interpolated anomalies were added to the monthly climatology to generate daily temperature maps.

For precipitation, the average monthly gridded climatology during the reference period was based on APHRODITE data (Yatagai et al., 2012). This 0.25° gridded data set was chosen as topography is considered in the APHRODITE interpolation method so that this data set at least improves on a simple interpolation of the available station data by IDW. As daily APHRODITE data are likely to contain inhomogeneities from a varying number of stations, we used APHRODITE only to derive the monthly gridded climatology during the reference period. Daily precipitation was then generated in a similar way as the temperature data, but anomalies were calculated as ratios instead of differences to the average monthly values during the reference period.

In addition to temperature and precipitation data, time series of humidity and radiation were required to drive the hydrological model. As hardly any measured time series were available, the 0.25° Watch Forcing Data based on ERA-40 (WFD-E40) (Uppala et al., 2005; Weedon et al., 2011) were used for the time period 1957–2001. These were complemented with ERA-Interim based Watch Forcing Data (WFD-EI) (Dee et al., 2011; Weedon et al., 2014) for the period 2002–2004. Correspondence between the two data sets was ensured by multiplying the WFD-EI data with a bias factor, which was calculated for each day of the year over the common period 1979–2001 using 31 day moving averages.

4.3.2 Trend analysis

Throughout this paper, trends were estimated using Sen's slope estimator (Sen, 1968). This estimator is less affected by outliers and non-normally distributed data than linear regression (e.g. Yue et al., 2003). Trend significance was assessed using the nonparametric Mann-Kendall test (Mann, 1945; Kendall, 1975). This test is robust against outliers and also suited for non-normally distributed data, or data with nonlinear trends. Since serial correlation may lead to an overestimation of the trend, we applied so-called trend free pre-whitening (TFPW), thus removing the influence of a lag-one autoregressive process (Yue et al., 2002). In contrast to the direct application of pre-whitening to the original time series, TFPW preserves the trend magnitude. Test results are reported as *p*-values, i.e. the probability of being wrong if rejecting the null hypothesis of no trend.

4.3.3 Data-based analysis

Trends were analyzed for streamflow, precipitation, and temperature time series. Multiple linear regression analysis was conducted to estimate the influence of temperature or precipitation on interannual streamflow variations. This analysis was performed at a seasonal time scale in order to consider possibly varying influences in the different seasons. Seasons were defined as winter (December, January, February), spring (March, April, May), summer (June, July, August), and autumn (September, October, November). The dependent variable was represented by annual series of spring (summer, autumn, winter) runoff. Explanatory variables were temperature and precipitation of the current season, and the average temperature or accumulated precipitation over the previous 12 months. A stepwise backwards approach (e.g. Backhaus et al., 2003) was applied for variable selection, setting the *p*-value of an *F*-statistic for exclusion and inclusion of the explanatory variables to 0.1 and 0.05, respectively.

4.3.4 Simulation-based analysis

4.3.4.1 WASA hydrological model

The hydrological model WASA (Model of Water Availability in Semi-Arid Environments) (Güntner, 2002; Güntner and Bronstert, 2004) is a semidistributed model with conceptual and process-orientated approaches. The model has been adapted for applications in mountain regions with previous applications in Central Asia (Duethmann et al., 2013; Duethmann et al., 2014). The original spatial discretization based on hillslopes (Güntner and Bronstert, 2004; Francke et al., 2008) was modified into an approach based on hydrological response units (HRUs) as it allows faster computations. The model has routines for snow versus rain differentiation, snow and glacier melt, glacier mass balance, interception, evapotranspiration, infiltration and saturation excess surface runoff, interflow, and groundwater runoff. The snow and glacier modules are described below; descriptions of the other routines can be found in Güntner (2002).

If the air temperature falls below a calibrated threshold, precipitation is considered as snow. Preferential snow deposition, snow redistribution by wind and avalanches, and spatial variability in snowmelt result in spatial variability of the snow water equivalent. In the WASA model, the spatial variability of snow within an HRU is represented using a snow depletion curve (Liston, 2004). After snowfall, HRUs where snow fall occurred are simulated as completely snow covered. Assuming a lognormal distribution of snow depth (e.g. Donald et al., 1995), snowmelt results in a gradual melt-out of this elevation zone. This is in better agreement with observations than an instantaneous change from completely snow-covered to snow-free conditions and also allows for a more straightforward comparison to snow cover observations from remote sensing (Duethmann et al., 2014). The model additionally considers snow redistribution between elevation zones. At elevated mountain peaks, snowmelt can be very low due to low temperatures. However, snow does still not accumulate over years, which would result in the formation of glaciers, but is moved to lower elevations by avalanches and/or wind transport. In the model, snow which is not on a glacier but accumulates above a given snow water equivalent threshold is moved to lower elevation zones. Snowmelt is calculated with a temperature-index approach (Hock, 2003). The melt factor varies in a sinusoidal form between a minimum value at the winter solstice and a maximum value at the summer solstice, reflecting the increase in incoming solar radiation (Anderson, 2006). Glacier melt is simulated using the same approach as for snowmelt but with a different melt factor. The glacier mass balance is calculated from snowfall onto the glacier, minus glacier- and snowmelt from the glacier area. Glacier melt from clean ice and debris-covered glaciers is not differentiated since it is still under debate whether lower melt rates, that have to be expected under thick debris cover (e.g. Östrem, 1959), also apply at the regional scale (Gardelle et al., 2012; Kääb et al., 2012). In addition, Pieczonka et al. (2013) found significant glacier mass loss in the Tomur region despite thick debris cover. The effect of debris cover is implicitly accounted for in our results, as model parameters are calibrated against geodetic glacier mass balances, which include both glaciers with and without debris cover.

The model has 13 calibration parameters. Due to the large uncertainties in areal precipitation, the calibrated parameters include a precipitation bias factor. In mountain areas, estimating areal precipitation from sparsely distributed precipitation stations results in high uncertainties. In such cases, hydrological modeling in combination with observed discharge and additionally glacier volume changes can be a good alternative for evaluating and correcting areal precipitation estimates (Duethmann et al., 2013). The precipitation bias factor is applied as a constant factor to the precipitation input, e.g. a precipitation bias factor of 1.5 means that precipitation used in the model is 50% higher than the original precipitation input. The bounds for the calibration parameters were set according to Duethmann et al. (2014), but with larger ranges for the groundwater delay factor (200–1000 days) and the precipitation bias factor (0.5–2.5).

4.3.4.2 Model setup

The WASA model was set up for the two study catchments using the data sources listed in Table 4.2. The two study catchments were subdivided into 16 and 13 subcatchments with an average size of 1100 km^2. Within each subcatchment, HRUs were defined as 200 m elevation bands, since differences in elevation rather than differences in land cover or soil (which are furthermore in mountain regions often related to elevation) were assumed to be most important for differences in hydrological processes. This resulted in 202 (Kakshaal) and 219 (Sari-Djaz) HRUs. Each HRU was assigned its glacier fraction, dominant land cover, and dominant soil. All meteorological data were spatially averaged to the level of subcatchments. Temperature data were scaled to the average elevation of a HRU using the estimated lapse rates.

Table 4.2: Spatial data sets used for setting up the WASA model.

Data type	Data source	Resolution	Reference
DEM	SRTM DEM	90 m	Jarvis et al. (2008)
Glacier outlines 1970s	Delineated from Landsat MSS images and KH-9 Hexagon images (1970s)	~8m	Pieczonka and Bolch (in revision)
Glacier outlines ~2008	Delineated from Landsat TM and Landsat ETM+ images (2007–2010)	30 m	Osmonov et al. (2013), Pieczonka and Bolch (in revision)
Soil types	Harmonized World Soils Database	1:1,000,000	FAO/IIASA/ISRIC/ISS-CAS/JRC (2012)
Land cover	MODIS land cover product MOD12Q1 collection 5.1	500 m	Friedl et al. (2002)

4.3.4.3 Changes in glacier area and elevation

Over the 48 year study period, effects of changes in glacier area and glacier surface elevation cannot be neglected and were therefore considered in the hydrological model in the following way. Glacier area changes were derived from two glacier inventories. This way, it was assured that the considered glacier area changes were close to the observed ones, while deriving glacier area changes from estimated glacier volumes and simulated glacier thickness changes could involve larger uncertainties. Glacier cover was available for ~1975 largely from KH-9 Hexagon images (resolution 7.6 m), which were completed by Landsat Multispectral Scanner (MSS; resolution 79 m) data, and for ~2008 from Landsat Thematic Mapper (TM; resolution 30 m) and Landsat Enhanced Thematic Mapper Plus images (ETM+; resolution 30 m) (Osmonov et al., 2013; Pieczonka and Bolch, in revision). Based on the two glacier inventories, linear decrease rates in area were calculated for each individual glacier. Glacier areas in the model were then updated annually according to these decrease rates during the period 1975–2004. No further glacier inventory was available for the beginning of our study period. For the period before 1975, unchanged glacier outlines were assumed. This is a more plausible assumption than extending the linear trend backwards in time, as observed glacier mass balances in Central Asia were not far from balance in the 1960s and started to become predominantly negative in the mid-1970s (Unger-Shayesteh et al., 2013).

In order to account for glacier elevation changes, the Δh-approach by Huss et al. (2010) was applied. Due to ice flow, glacier ice thickness changes in the ablation zone are less negative than the local glacier mass balance. The Δh-approach uses predefined functions to distribute the glacier-wide mass balance across glacier elevation bands. These functions were parameterized based on glacier thickness changes reported for the Ak-Shirak massif and the period 1977–1999 by Surazakov and Aizen (2006). The calculation of glacier thickness changes is performed on an annual basis individually for each glacier. Glaciers were subdivided into 50 m elevation bands for this purpose. Glacier elevation changes were restricted by glacier thickness, i.e. no further reduction in glacier elevation was allowed if the accumulated elevation change exceeded the estimated glacier thickness. A glacier ice thickness distribution of each individual glacier was estimated based on the Glabtop2 model introduced by Linsbauer et al. (2012) and modified by Frey et al. (2014) using the filled SRTM DEM and the 1970s glacier inventory as input.

4.3.4.4 Multiobjective model calibration

For model calibration, criteria based on glacier mass balance and river discharge were taken into account. As the main precipitation season is in summer, snowmelt, rainfall, and glacier melt occur at a similar time. Considering both discharge and glacier mass balance is therefore important for this region, as the uncertainties in precipitation are high and require a calibrated precipitation bias factor. An underestimation of precipitation can be compensated by an overestimation of glacier melt and vice versa. Two criteria were used for the calibration to discharge. *OF_Q1* (Eq. 3) was applied to ensure a good model performance with respect to the observed hydrograph. In order to achieve a balanced solution with a good performance both in high and low flow periods, we used a balanced Nash-Sutcliffe efficiency for daily discharge values, which was defined as average of NSE_d, the Nash-Sutcliffe criterion of daily discharge values, and $LogNSE_d$, the Nash-Sutcliffe criterion calculated from daily logarithmic discharge values:

$$OF\_Q1 = 0.5 \times \left(NSE_d + LogNSE_d \right), \tag{3}$$

with

$$NSE_d = 1 - \frac{\sum_{t=1}^{T}\left(Q_{obs}(t) - Q_{sim}(t)\right)^2}{\sum_{t=1}^{T}\left(Q_{obs}(t) - \overline{Q_{obs}}\right)^2} \text{ and}$$

$$LogNSE_d = 1 - \frac{\sum_{t=1}^{T}\left(\log\left(Q_{obs}(t)\right) - \log\left(Q_{sim}(t)\right)\right)^2}{\sum_{t=1}^{T}\left(\log\left(Q_{obs}(t)\right) - \overline{\log\left(Q_{obs}\right)}\right)^2}. \quad (4)$$

In Eq. (4), Q_{obs}(t) and Q_{sim}(t) are the observed and simulated discharge at time t, respectively, and T is the number of time steps.

This criterion is, however, not very sensitive to interannual discharge variations, and the performance with respect to interannual variations of discharge may be low despite good performance for daily discharge variations. Therefore, a criterion based on annual time series of seasonal discharge was formulated additionally:

$$OF\_Q2 = balNSE_{DJF} + balNSE_{MAM} + balNSE_{JJA} + balNSE_{SON} \quad (5)$$

where $balNSE_{DJF}$ ($balNSE_{MAM}$, $balNSE_{JJA}$, $balNSE_{SON}$) is the balanced Nash-Sutcliffe criterion (cf. Eq. 3) calculated from annual series of observed and simulated discharge for the winter (spring, summer, autumn) season.

Concerning glacier mass balance, a time series of observed glacier mass balances for Karabatkak glacier for the period 1957–1998, and geodetic mass balances for two areas adjacent to our study area were available (see section 4.2). Based on these data, two calibration criteria with respect to glacier mass balances were formulated. The correlation between the simulated catchment-average mass balance and the glacier mass balance time series of Karabatkak (*OF_G1*) was used as a criterion reflecting the year-to-year variation of the glacier mass balance, which is not given by the geodetic estimate. The geodetic estimates, which are representative for a larger area but only give one estimate over the measuring period, were additionally applied as criterion for the cumulative mass change over the period 1976–1999. Using the mean of the two geodetic estimates and the larger of the two uncertainty ranges, the criterion *OF_G2* was formulated as a triangular fuzzy membership function with a value of 1 if the simulated mass balance over 1976–1999 matches the mean value of −0.52 m w.e. a^{-1}, a value of 0 if the simulated mass balance was outside the estimated uncertainty range of −0.85 to −0.19 m w.e. a^{-1}, and linearly increasing/decreasing values in between.

Snow cover data from remote sensing are a further data source suitable for model calibration in snow-dominated, data sparse regions (Duethmann et al., 2014). MODIS snow cover data (MOD10A1/MYD10A1; Hall et al., 2002) were summarized within elevation zones and subcatchments. In many subcatchments, the data showed negligible snow cover fractions during winter in elevations zones up to 3600 m, interrupted by few, short-lasting snowfall events. This is probably due to the low snowfall during the winter period which, according to the data set described in section 4.3.1, was only 10 and 12 mm $month^{-1}$ (average value for the period overlapping with the MODIS snow cover data 2001–2004). Such thin snow layers may be redistributed by wind so that snow accumulates in small fractions of the basin, while large parts remain snow free. Furthermore, snow may also be removed by sublimation. Factors favoring sublimation are low humidity and high radiation input, both consistent with the low winter precipitation. The model in its current form cannot represent the observed snow cover data, as sublimation is not considered and the parameterization for snow redistribution does not become effective before the onset of melt. Satellite snow cover data were therefore not used for model calibration since calibration against such data would result in compensating effects (e.g. model parameters might be drawn toward allowing melt at very low temperatures), thus increasing errors in the internal functioning of the model. Due to the very low amounts of winter precipitation, we assume that neglecting the processes of sublimation and snow redistribution before melt onset has only small effects on the overall performance of the hydrological model.

In the context of multiobjective optimization, we seek a set of Pareto optimal (or nondominated) solutions instead of a single optimum solution. For each of the solutions belonging to this Pareto optimal set, none of the objective functions can be improved without degradation in at least one of the other objective functions. In this study, the location of the Pareto optimal set was estimated using the Epsilon-Dominance Nondominated Sorted Genetic Algorithm II (ε-NSGAII; Kollat and Reed, 2006), which proved to be efficient and effective in studies comparing different multiobjective optimization algorithms (Kollat and Reed, 2006; Tang et al., 2006). The density of the final nondominated solutions is controlled using the concept of ε-dominance. For each objective function, the user defines an ε-value, which is the smallest difference in the function that is considered as relevant. For this study, the initial population size was set to 16, the maximum run time to 30,000 simulations, and ε-values were set to 0.01 for all four objectives as the different objectives vary in the same order of magnitude. Other algorithm parameters were adopted from Kollat and Reed (2006).

The model was calibrated for the period 1976–1999, according to the available geodetic glacier

mass balance estimates. A two year period prior to the calibration period was used for model initialization. Due to missing daily runoff data after 1987, *OF_Q1* only refers to the period 1976–1987. The remaining part of the available time series from 1957–1975 and 2000–2004 was used as validation period.

After model calibration, solutions were selected from the Pareto optimal set for further evaluation. Solutions with a glacier mass balance outside the uncertainty ranges of the geodetic estimate ($OF\_G2 = 0$), or low performance with respect to daily or interannual variations of discharge were excluded. After inspecting the achieved solutions, the threshold for exclusion with respect to interannual variations of discharge was set to $OF\_Q2 < -2$. As the remaining solutions all showed a good performance with respect to daily discharge ($OF\_Q1 > 0.75$), no additional threshold was set for *OF_Q1*. Since the model was applied in order to analyze causes for changes in discharge, the ability of the model to represent observed discharge trends was set as a further criterion. For this purpose, trends in seasonal discharge time series were calculated over the whole study period, as the calibration or validation periods were regarded as too short for the calculation of trends. Solutions which did not simulate a significant positive (negative) trend ($p > 0.1$), even though the observed time series showed a positive (negative) trend, were excluded from further evaluation.

4.4 Results

4.4.1 Data-based analysis

In both catchments the trend analysis revealed increasing discharge in all months over the period 1957–2004 (Fig. 4.2 a–d), in agreement with previous studies. For the Kakshaal catchment, trends were significant at the 5% level in nearly all months. The highest absolute increases in monthly discharge occurred during May to July. In most months, discharge increased by around 30% (relative changes with respect to the mean of the study period), with larger changes in May and lower changes in August. In the Sari-Djaz catchment, the largest absolute increases in discharge were observed for July and August, where monthly discharge increased by 33 mm, or 32%.

Changes in precipitation show a less clear picture, with both increases and decreases (mostly not significant) (Fig. 4.2 e–h). The summer months, which are the months with the highest precipitation, only show insignificant changes in both catchments. Temperature, in contrast, increased during all months in both catchments (Fig. 4.2 i–j). In the Kakshaal catchment, temperatures increased most strongly during the winter months, with changes of 3–4 K over the 48 year study period. In the Sari-Djaz catchment, significant trends with *p*-values ≤0.05 occurred mostly during the second half of the year from July to December. Relatively large positive trends but with slightly lower significance levels occurred in January and February.

In order to analyze the relation between discharge and climate parameters on a seasonal scale, a regression model for estimating the seasonal discharge was applied. Explanatory variables were temperature and precipitation of the current season, and average temperature and accumulated precipitation of the last year. The regression model resulted in coefficients of determination (R^2) between 0.1 and 0.6 (Table 4.3, Fig. 4.3). Only relatively low R^2-values were achieved for spring in both catchments, and also for winter in the Sari-Djaz catchment. Trends of discharge from the regression model were generally lower than trends of the observed discharge, except for autumn (Table 4.3). The identification of the regression model was impaired by significant correlations between the explanatory variables, for example summer temperatures and summer precipitation were significantly negatively correlated ($p < 0.01$) with correlation coefficients of −0.6 and −0.4 for the Kakshaal and Sari-Djaz catchment, respectively (Table 4.4).

The influence of the variables included in the regression can be evaluated by standardizing the coefficients (Table 4.3). For the high flow season in summer, and also for the following autumn season, the variables with the highest influence were precipitation in the Kakshaal catchment, and temperature in the Sari-Djaz catchment. The stronger influence of precipitation on discharge for the Kakshaal catchment is related to a higher contribution of snowmelt and rain for the Kakshaal as compared to the Sari-Djaz River, where glacier melt plays a larger role (Krysanova et al., 2014). In winter, the most important explanatory variables were precipitation of the last year for the Kakshaal and temperature of the last year for the Sari-Djaz catchment. A large part of the variability in winter can be explained by discharge of the previous year (analysis not shown), indicating that the greatest discharge fraction in winter is release from storage. This results in a stronger influence of last year's precipitation in the Kakshaal and of last year's temperature in the Sari-Djaz catchment. The performance of the regression models is lowest in spring, and the variables with the strongest influence are mean temperature of the previous year for the Kakshaal, and spring temperature for the Sari-Djaz catchment.

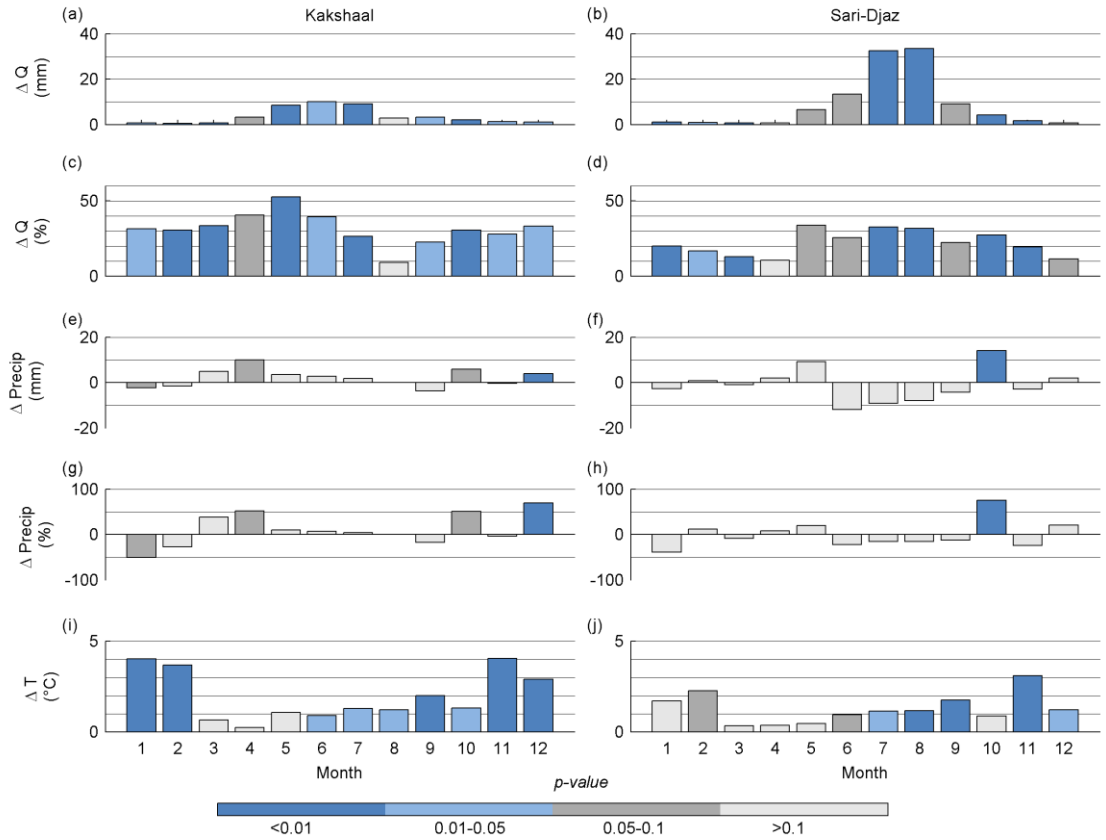

Fig. 4.2: Monthly changes in discharge (a–d), in catchment average precipitation (e–h), and in temperature (i–j) over the period 1957–2004 for the Kakshaal and Sari-Djaz catchment. Changes in discharge and precipitation are reported in absolute (mm) and relative values (%). Bar colors indicate the significance level of the Mann-Kendall trend test.

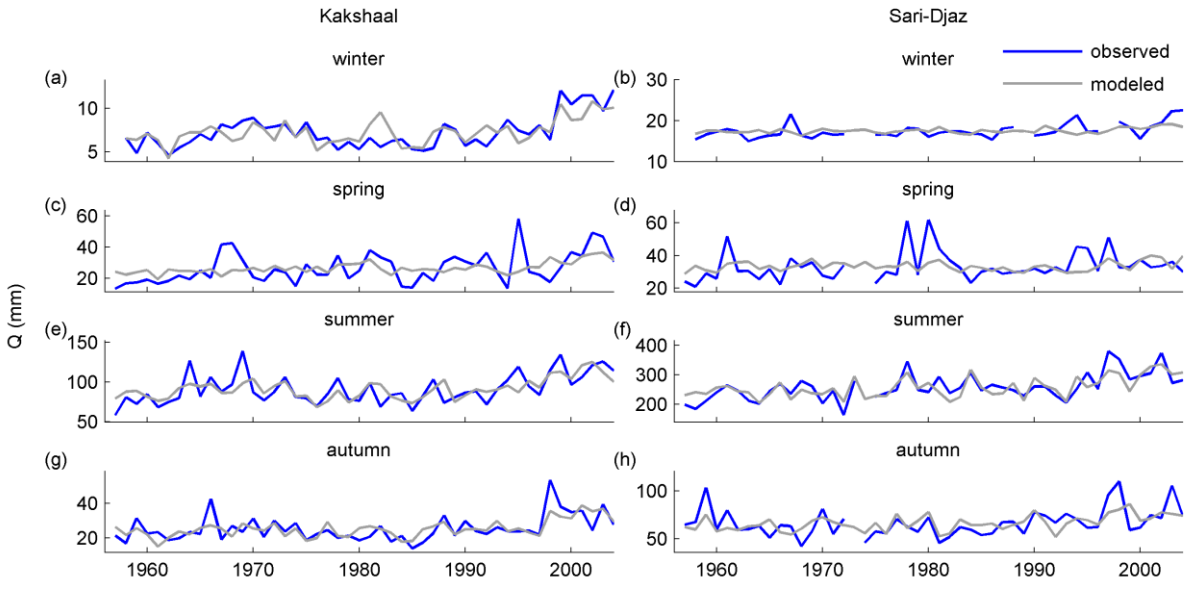

Fig. 4.3: Observed (blue) and modeled (gray) seasonal discharge for the Kakshaal and Sari-Djaz catchment. The modeled discharge is based on a regression model with temperature and precipitation of the current season and of the last year as possible explanatory variables.

Table 4.3: Results of the linear regression analysis of discharge: coefficient of determination (R^2), trend in observed discharge (*Qobs*) and corresponding *p*-value, trend in discharge as assessed with the regression model (*Qregr*) and corresponding *p*-value, as well as the included variables (T: temperature, P: precipitation) with standardized coefficients, trend, corresponding *p*-value, and the coefficient multiplied with the trend in the explanatory variable. Trends are reported as cumulative values over the period 1957–2004.

						Included explanatory variables				
	R^2	Trend *Qobs* (mm)	*p*-value trend *Qobs*	Trend *Qregr* (mm)	*p*-value trend *Qregr*	Variable	Standardized coefficient	Trend of explanatory variable	*p*-value of trend	Trend × coefficient (mm)
					Kakshaal					
Winter	0.55	3	0.00	2	0.01	T last year	0.39	1.6°C	0.00	1.3
						P last year	0.58	15%	0.23	0.8
Spring	0.13	14	0.00	7	0.00	T last year	0.36	1.7°C	0.00	6.6
Summer	0.44	27	0.00	19	0.00	T	0.48	1.2°C	0.01	12.6
						P	0.62	4%	0.81	1.4
						P last year	0.35	17%	0.17	5.9
Autumn	0.46	6	0.02	7	0.00	T	0.31	2.4°C	0.00	4.8
						P	0.40	28%	0.16	2.1
						P last year	0.29	19%	0.16	2.2
					Sari-Djaz					
Winter	0.14	3	0.01	1	0.04	T last year	0.35	1.0°C	0.00	0.8
Spring	0.11	7	0.01	1	0.82	T	0.33	0.4°C	0.42	1.1
Summer	0.59	74	0.00	63	0.00	T	0.67	1.1°C	0.00	47.0
						T last year	0.22	1.3°C	0.00	16.1
Autumn	0.30	12	0.10	13	0.00	T	0.55	1.8°C	0.00	13.1

Table 4.4: Correlations between explanatory variables (T: temperature, P: precipitation) used in the regression models. * and ** indicate correlations with *p*-values <0.05 and <0.01, respectively.

	T P	T last year P last year	T T last year	T P last year	P P last year	P T last year
			Kakshaal			
Winter	0.0	0.1	0.5**	−0.1	0.2	0.1
Spring	−0.1	0.1	0.4**	0.0	0.0	0.5**
Summer	−0.6**	0.1	0.3*	0.2	0.0	0.2
Autumn	0.0	0.2	0.5**	0.3*	0.2	0.3*
			Sari-Djaz			
Winter	−0.2	0.3*	0.5**	−0.1	0.3	0.2
Spring	0.4**	0.2	0.4**	0.3	0.1	0.5**
Summer	−0.4**	0.3*	0.4*	0.3*	0.1	0.3
Autumn	0.1	0.3*	0.3*	0.6**	0.2	0.5**

Trend analyses of the climate parameters that were included in the regression models can give indications of the possible causes for the changes in discharge. Based on this, increasing streamflow in the Kakshaal catchment in summer, winter, and autumn can be explained by significantly increasing temperatures ($p < 0.05$). Increases of precipitation likely further contributed to this trend, even though these increases were not significant ($p > 0.1$) (Table 4.3). The calculated trend multiplied with the respective (not standardized) coefficient of the regression equation can give an indication for the relative influence of the temperature or precipitation trend on the discharge trend. This points to a stronger influence of the temperature as compared to the precipitation trend. In the Sari-Djaz catchment, seasonal streamflow was only related to the temperature of the respective season or of the last year, and not related to precipitation. Discharge increases in the Sari-Djaz catchment may therefore primarily be attributed to the temperature increases (Table 4.3), pointing at a stronger importance of snow and glacier melt. However, even though plausible explanations for the statistically identified relations could be found, the results should be taken with care, particularly considering the correlations between some of the explanatory variables.

4.4.2 Simulation-based analysis

4.4.2.1 General model performance

While there were no clear trade-offs between the two objective functions for discharge (*OF_Q1* and *OF_Q2*), trade-offs were apparent between the discharge and glacier objective functions, and to a very small extent between the two glacier objective functions (Fig. 4.4). A distinct trade-off is observed between *OF_Q1* and *OF_G1*, and this trade-off was larger for the Sari-Djaz than for the Kakshaal catchment. From the resulting Pareto set, solutions were selected for further evaluations if they showed an acceptable performance in terms of interannual discharge variation, if the simulated glacier mass balance was inside the range given by the geodetic estimate, and if a significant trend for seasonal streamflow was simulated when the observed time series displayed a trend (see section 4.3.4.4), resulting in the selection of 28 solutions for the Kakshaal and 6 solutions for the Sari-Djaz catchment.

Daily and monthly discharge variations are well represented by the model (Table 4.5; Fig. 4.5–4.6). The slightly lower model performance in the Kakshaal catchment is likely related to the stronger

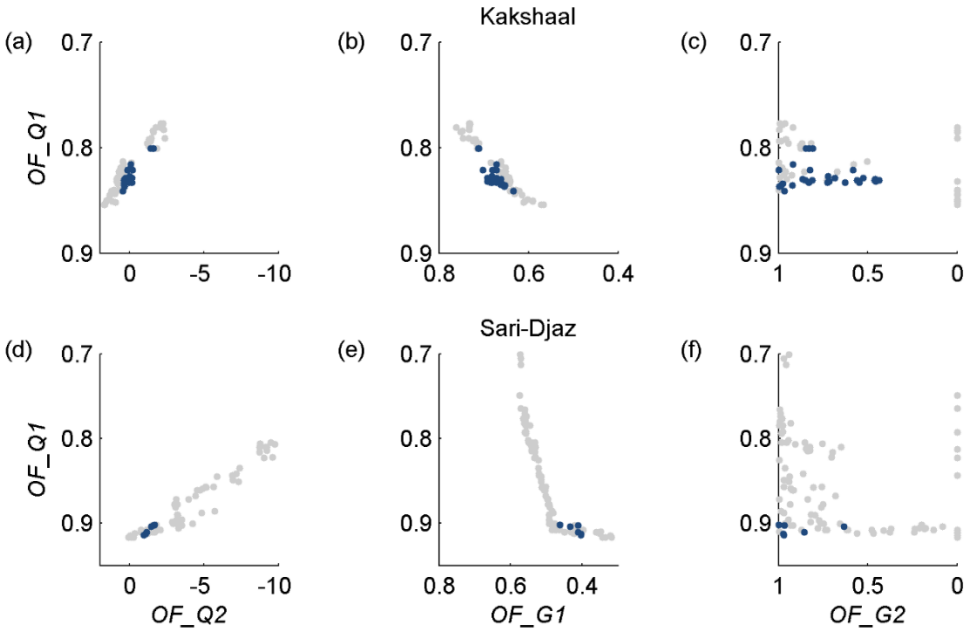

Fig. 4.4: Trade-off curves of model performance with respect to daily discharge (*OF_Q1*) against model performance with respect to interannual discharge variations (*OF_Q2*), glacier mass balance time series at Karabatkak glacier (*OF_G1*), and the geodetic glacier mass balance (*OF_G2*) for the two study catchments. The *x* axes are plotted in reverse order so that optimum solutions always plot in the lower left corner. Blue dots show solutions used for further analyses, gray dots show solutions which were excluded due to too low performance with respect to *OF_Q2*, *OF_G2*, or representation of observed trends (see section 4.3.4.4).

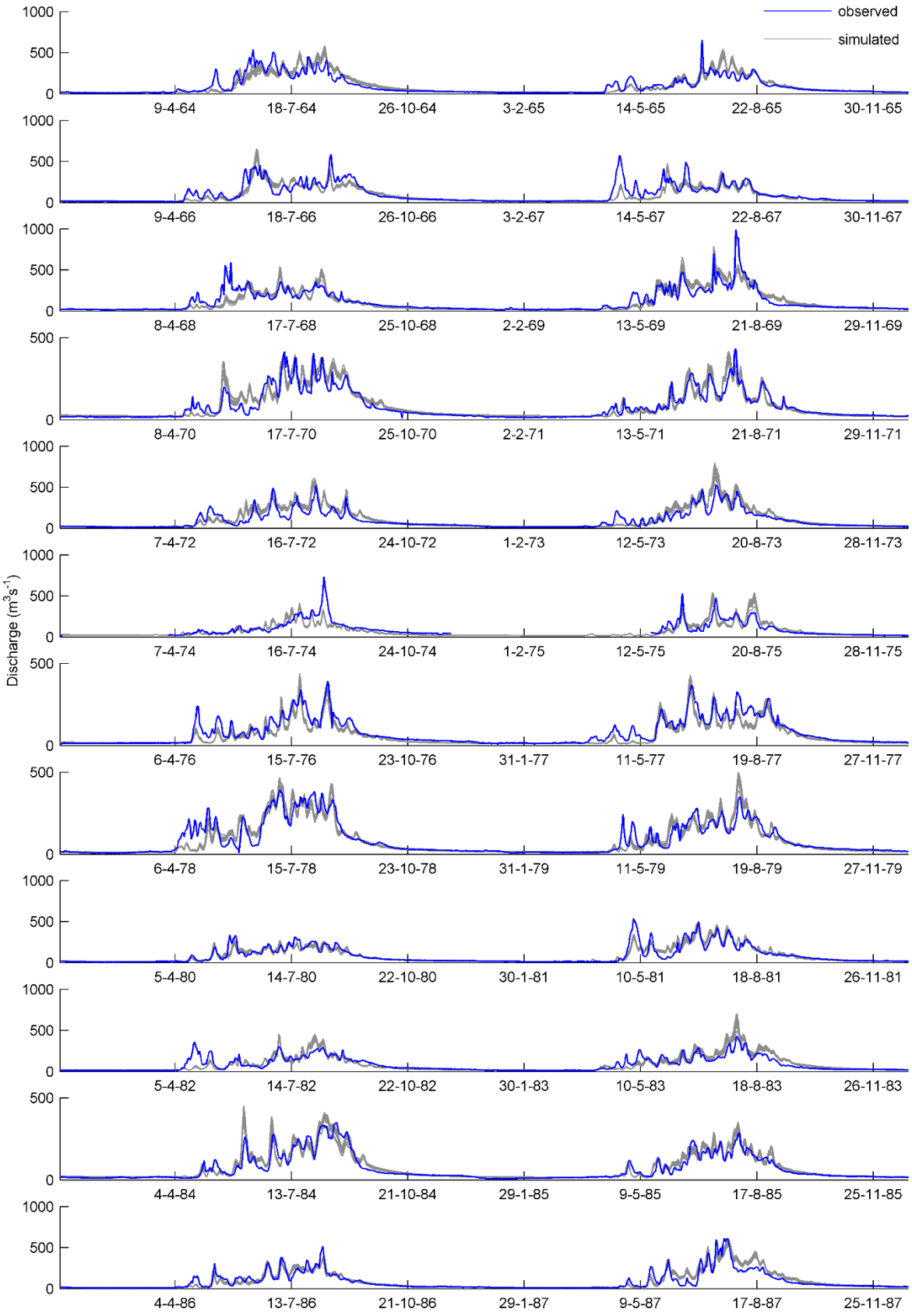

Fig. 4.5: Simulated and observed daily runoff for the period 1964–1987 for the Kakshaal catchment.

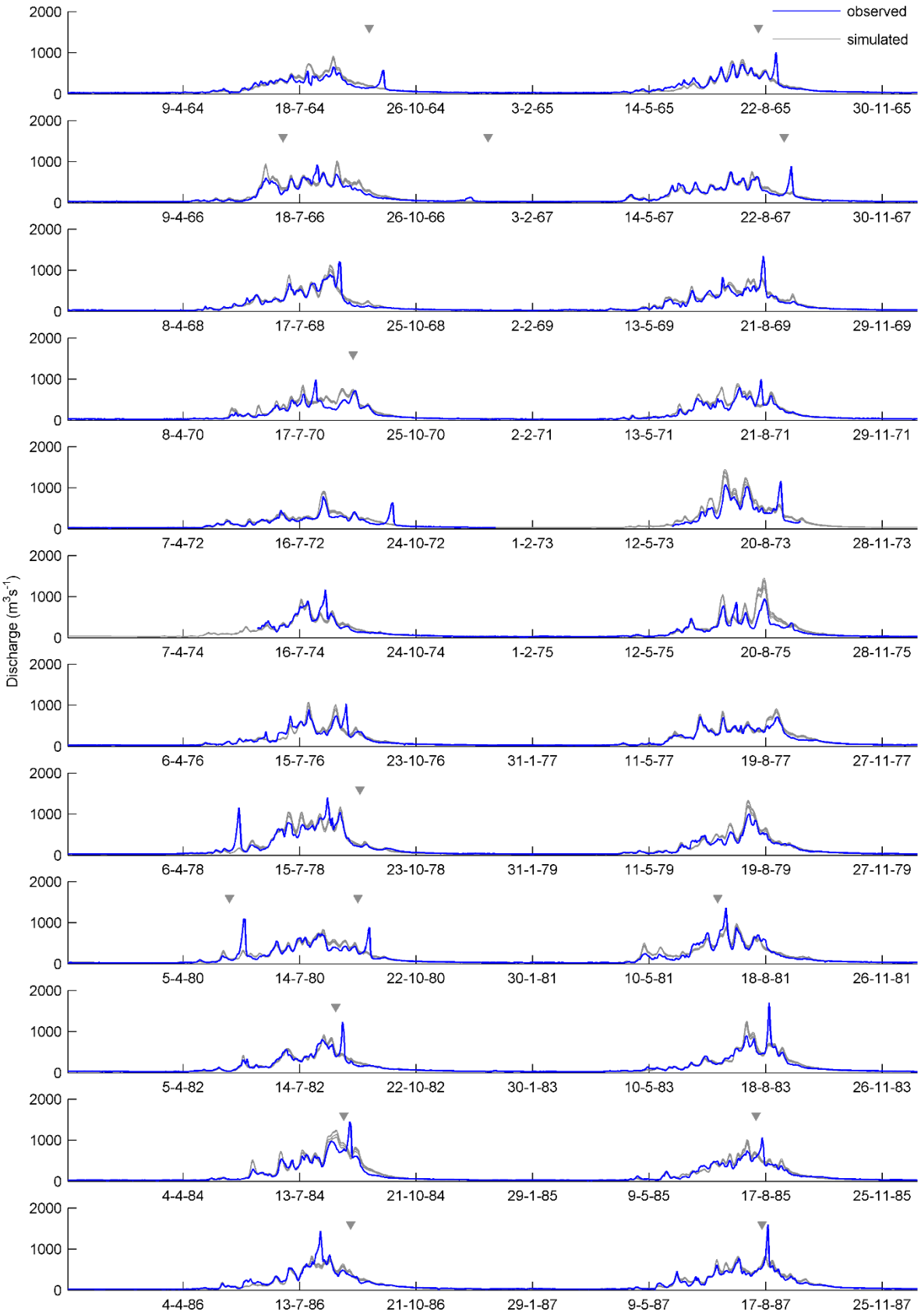

Fig. 4.6: Simulated and observed daily runoff for the period 1964–1987 for the Sari-Djaz catchment. Gray triangles mark the dates of reported GLOFs as listed in Glazirin (2010).

Table 4.5: Range of Nash-Sutcliffe efficiency (*NSE*), Nash-Sutcliffe efficiency calculated for logarithmic discharge values (*logNSE*), and volume bias for daily and monthly discharge during the calibration and validation period of the WASA model.

	NSE		*logNSE*		Volume bias (%)	
	Kakshaal	Sari-Djaz	Kakshaal	Sari-Djaz	Kakshaal	Sari-Djaz
			Daily			
calibration	0.77–0.82	0.85–0.87	0.84–0.87	0.95–0.96	−5 – 8	1 – 9
validation	0.67–0.73	0.80–0.84	0.82–0.86	0.94–0.95	−7 – 8	7 – 14
			Monthly			
calibration	0.81–0.86	0.91–0.94	0.88–0.91	0.96–0.96	−6 – 7	−2 – 6
validation	0.80–0.86	0.86–0.92	0.85–0.90	0.95–0.95	−7 – 6	10 – 18

influence of precipitation in this catchment, which is generally associated with higher uncertainties than temperature. For some flood peaks at the Sari-Djaz gauge, large deviations occur (Fig. 4.6). These peaks may be related to GLOFs from Merzbacher Lake, which cannot be simulated by the WASA model.

The dynamics of the simulated catchment average glacier mass balance is generally similar to the observed mass balance of Karabatkak glacier (Fig. 4.7). The low correlation value between the time series at Karabatkak and the simulated glacier mass balance in the Sari-Djaz Basin (Fig. 4.4, Fig. 4.7) can to a large extent be explained by an underestimation during the last five years. For the period 1976–1999, for which geodetic glacier mass balances were available, the mass balance was simulated at −0.85 to −0.25 m w.e. a^{-1} and −0.83 to −0.51 m w.e. a^{-1} for the Kakshaal and Sari-Djaz catchment, respectively, in agreement with the range derived from the geodetic estimates applied for model calibration. Over the complete simulation period 1957–2004, the simulated mass balance was estimated between −0.76 and −0.13 m w.e. a^{-1} for the Kakshaal and between −0.77 and −0.40 m w.e. a^{-1} for the Sari-Djaz catchment.

The simulated fraction of runoff from glacier melt (defined as melt water from glacier ice/firn, without snowmelt or rain from glacier areas) may be approximated as the simulated glacier melt divided by the simulated runoff, assuming that evaporation from glacier melt is small. This results in glacier melt contributions to runoff of 9–24% for the Kakshaal catchment, and 35–48% for the Sari-Djaz catchment.

The calibrated precipitation bias factor (see section 3.4.1) allowed us to estimate the areal precipitation from the simulated and observed discharge. The catchment average precipitation was estimated about 50% higher than the original estimates based on interpolated precipitation data (Table 4.1). It reached values of 372–399 mm a^{-1} for the Kakshaal and 450–526 mm a^{-1} for the Sari-Djaz catchment.

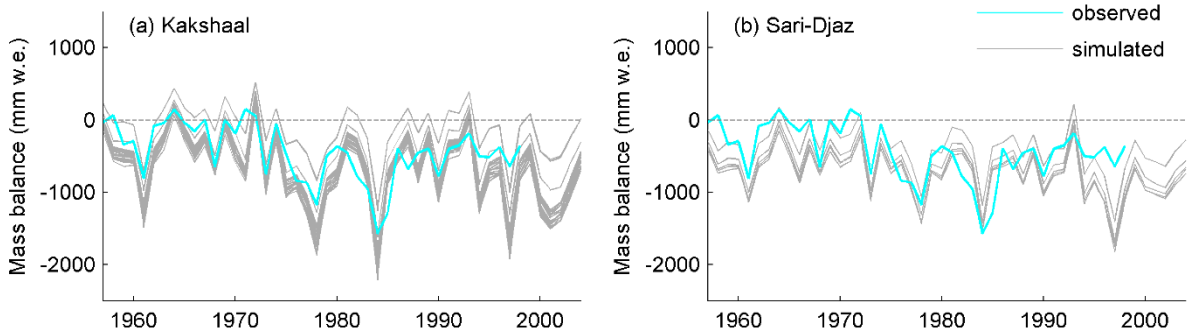

Fig. 4.7: Interannual variation of the observed glacier mass balance at Karabatkak glacier, and simulated catchment average glacier mass balance for (a) the Kakshaal and (b) the Sari-Djaz catchment.

4.4.2.2 Model performance with respect to interannual variability of discharge

While the model performance is high for daily or monthly time series, it is considerably lower when looking at interannual variations of annual or seasonal flow. This is because seasonal variations between winter low flow and summer high flow are much more distinct than interannual variations. For the Kakshaal catchment, R^2-values between simulated and observed series of average annual flow over the whole study period are between 0.61 and 0.64 (Table 4.6). Comparable fits were also achieved for annual series of seasonal flow, except for spring where discharge and its interannual variation are underestimated by the model. The simulated streamflow follows the long-term variations in observed discharge over the 48 year period with low discharge values at the beginning of the time period and around 1980, and higher discharges in the late 1960s and particularly after 1995 (Fig. 4.8 a). In the Sari-Djaz catchment, model performance for interannual variation of annual or seasonal values is lower than for the Kakshaal catchment (Table 4.6). This is mostly due to larger deviations during the last ~10 years (Fig. 4.8), and generally low performance in winter and spring. A reason for the larger deviations of the hydrological model for the last ~10 years could be the lower availability of precipitation data during that period. In both catchments, the observed runoff shows increasing trends over both annual and seasonal time scales (Table 4.6). In the Kakshaal catchment, trends in simulated annual and seasonal discharge were close to the observed ones. In the Sari-Djaz catchment, the hydrological model showed an underestimation of trends in summer, and an overestimation in autumn. This might be related to a tendency of the GLOFs from Lake Merzbacher occurring earlier in the year (Glazirin, 2010).

For the Kakshaal catchment, the hydrological simulations resulted for all seasons in higher R^2-values than the regression model, and the hydrological model also showed a better performance with respect to the deviation between trends in modeled and observed discharge. For the Sari-Djaz catchment the two models were comparable; the hydrological model resulted in slightly higher R^2 for winter and autumn, and the regression model showed a higher performance in summer. In summer and autumn, the regression model achieved better results with respect to trends.

4.4.2.3 Effect of changes in glacier area and elevation

The effect of changes in glacier surface geometry was estimated by a set of simulations in which glacier area and glacier surface elevation were kept constant (Table 4.7). These simulations showed that changes in the glacier geometry had little effect on discharge trends in the less glacierized Kakshaal catchment. In the Sari-Djaz catchment, with a glacierization of 21%, the simulations showed that not considering changes in glacier geometry results in higher discharge increases. Estimated from these simulations, the increases in annual discharge would have been 14–23% higher if there had been no changes in glacier geometry (uncertainty range based on parameter uncertainties; Table 4.7). Changes in glacier area, which led to lower discharge increases, outweighed the changes in glacier surface elevation, which led to additional increases in discharge. Overall, glacier changes therefore resulted in lower discharge increases as compared to the hypothetical situation of unchanged glacier geometries.

4.4.2.4 Simulation with precipitation and temperature time series without trend

In order to investigate to what extent precipitation and temperature changes contributed to the observed changes in discharge, simulations with time series of these variables without trend were performed. This included three simulation experiments: (1) precipitation and temperature input without trend, (2) precipitation without trend and original temperature; (3) temperature without trend and original precipitation. For these simulation experiments, temperature or precipitation time series were generated which contained no changes over time, i.e. no trends nor changes of the daily or interannual variability. In order to achieve this, the same one-year daily time series was applied to all years in the simulation period. The daily time series of an arbitrary year in the middle of the study period was selected (the year 1980 was used), and a linear scaling was applied to the daily values so that the monthly means corresponded to the climatological means of the study period 1957–2004. The simulations with these temperature or precipitation time series were performed by keeping the glacier geometries unchanged.

The simulations with no-trend time series for both precipitation and temperature resulted in negligible streamflow trends (Table 4.8). This corroborates that the discharge trends for the model with the original precipitation/temperature data can largely be explained by the changes in these forcing data. The results show that both increases in temperature and precipitation contributed to the discharge increases in the Kakshaal catchment, with a higher influence of the precipitation increases (Table 4.8). The simulated trends with no-trend temperature or no-trend precipitation input were compared to the trends of the simulations with the original time

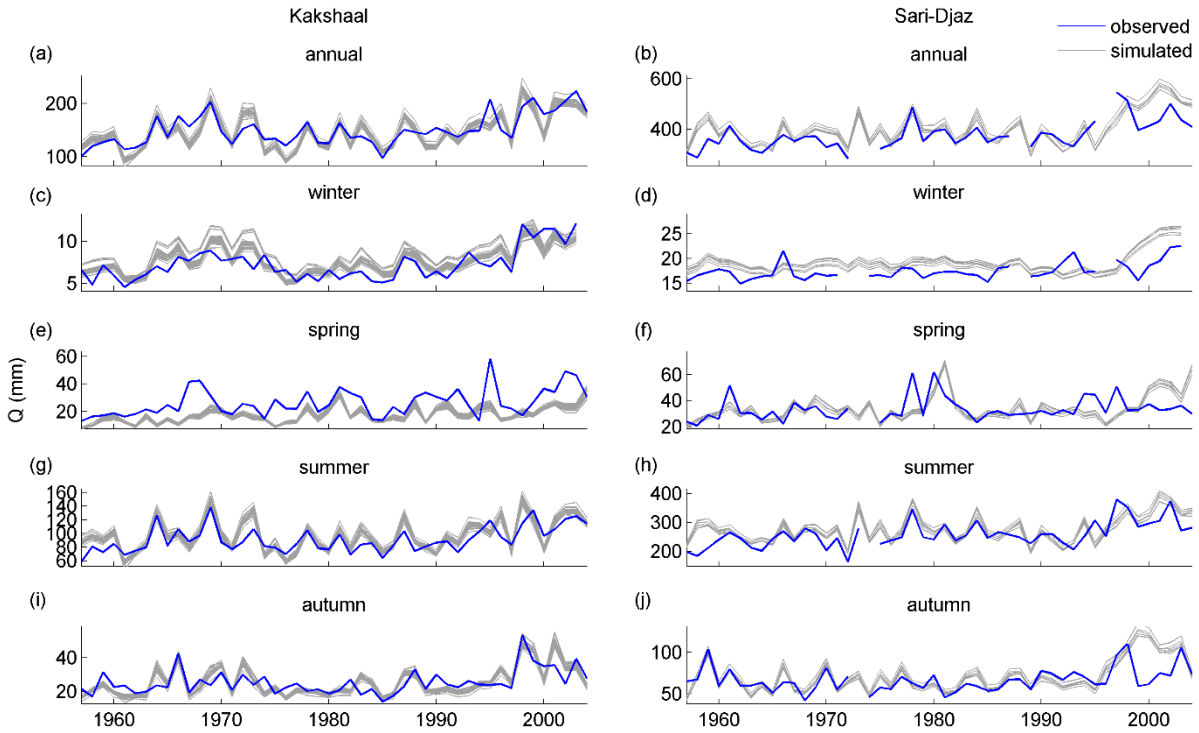

Fig. 4.8: Observed and simulated (a-b) annual, and (c–j) seasonal discharge for the Kakshaal and the Sari-Djaz catchment.

Table 4.6: Coefficient of determination (R^2) between annual and seasonal simulated (*Qsim*) and observed (*Qobs*) discharge, as well as corresponding trends over the period 1957–2004. For model performance and trends of simulated discharges the table gives ranges over the selected model parameterizations.

		Qobs		*Qsim*	
	R^2	Trend	*p*-value	Trend	*p*-value
	-	(mm)	-	(mm)	-
		Kakshaal			
Annual	0.61–0.64	47	0.00	47–53	0.00–0.03
Winter	0.57–0.65	3	0.00	2–3	0.00–0.02
Spring	0.27–0.29	14	0.00	9–12	0.00–0.00
Summer	0.62–0.65	27	0.00	22–27	0.04–0.09
Autumn	0.45–0.62	6	0.02	7–9	0.00–0.05
		Sari-Djaz			
Annual	0.43–0.49	102	0.00	89–101	0.00–0.01
Winter	0.19–0.22	3	0.01	1–1	0.03–0.09
Spring	0.10–0.12	7	0.01	8–10	0.01–0.02
Summer	0.41–0.47	74	0.00	50–57	0.02–0.03
Autumn	0.34–0.36	12	0.10	25–30	0.00–0.01

Table 4.7: Trends over the period 1957–2004 and *p*-values of the Mann-Kendall test for time series of annual and seasonal discharge, as simulated by different WASA model versions with and without considering changes in glacier area and elevation (ranges over the selected model parameterizations). The simulations were performed with the original precipitation and temperature time series.

	Model 1 Change in glacier area and surf. elevation		Model 2 Constant glacier geometry	
	Trend (mm)	*p*-value	Trend (mm)	*p*-value
		Kakshaal		
Annual	47–53	0.00–0.03	47–54	0.00–0.02
Winter	2–3	0.00–0.02	2–3	0.00–0.01
Spring	9–12	0.00–0.00	9–12	0.00–0.00
Summer	22–27	0.04–0.09	23–28	0.04–0.09
Autumn	7–9	0.00–0.05	8–10	0.00–0.03
		Sari-Djaz		
Annual	89–101	0.00–0.01	105–116	0.00–0.00
Winter	1–1	0.03–0.09	1–1	0.01–0.02
Spring	8–10	0.01–0.02	8–10	0.01–0.02
Summer	50–57	0.02–0.03	60–71	0.01–0.01
Autumn	25–30	0.00–0.01	29–33	0.00–0.00

Table 4.8: Trends over the period 1957–2004 and *p*-values of the Mann-Kendall test for simulated time series of annual and seasonal discharge, as ranges over the selected model parameterizations. The simulations were performed with different combinations of original (orig.) and no-trend precipitation and temperature time series, using a model which assumes a constant glacier geometry.

	orig. P, orig. T		no-trend P, no-trend T		orig. P, no-trend T		no-trend P orig. T	
	Trend (mm)	*p*-value	Trend (mm)	*p*-value	Trend (mm)	*p*-value	Trend (mm)	*p*-value
				Kakshaal				
Annual	47–54	0.00–0.02	0–3	0.04–0.78	35–45	0.07–0.12	11–22	0.01–0.05
Winter	2–3	0.00–0.01	0–0	0.00–0.31	2–3	0.00–0.02	0–1	0.00–0.09
Spring	9–12	0.00–0.00	−0–0	0.27–0.88	7–10	0.00–0.00	6–7	0.01–0.03
Summer	23–28	0.04–0.09	0–2	0.08–0.91	15–23	0.25–0.40	1–7	0.22–0.65
Autumn	8–10	0.00–0.03	0–1	0.01–0.32	4–7	0.08–0.15	3–6	0.01–0.17
				Sari-Djaz				
Annual	105–116	0.00–0.00	3–6	0.17–0.59	2–6	0.81–0.92	92–107	0.00–0.01
Winter	1–1	0.01–0.02	0–1	0.00–0.02	−1–−1	0.17–0.22	4–4	0.00–0.00
Spring	8–10	0.01–0.02	−0–0	0.32–0.53	1–2	0.70–0.95	10–11	0.00–0.01
Summer	60–71	0.01–0.01	1–3	0.12–0.45	2–3	0.87–0.99	50–59	0.01–0.01
Autumn	29–33	0.00–0.00	1–2	0.38–0.49	−4–−3	0.33–0.38	27–31	0.00–0.00

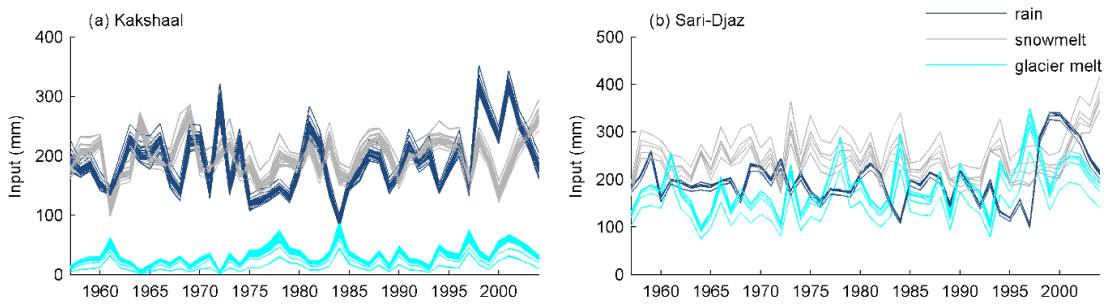

Fig. 4.9: Annual variation of the input from rain, snowmelt, and glacier melt, as represented in the WASA model for the selected model parameterizations for (a) the Kakshaal and (b) the Sari-Djaz catchment.

series in order to estimate the relative influence of temperature or precipitation changes on the discharge trends. Not considering the temperature changes reduced the annual discharge trend in the Kakshaal catchment by 10–28%, while not considering the precipitation changes reduced the annual discharge trend by 56–78%. In the Sari-Djaz catchment, changes in precipitation had only very small effects and increases in discharge were mainly caused by the increases in temperature (Table 4.8). Not considering changes in temperature reduced the annual discharge trend in the Sari-Djaz catchment by 94–98%, while not considering changes in precipitation reduced the annual discharge trend by only 7–12%.

4.4.2.5 Changes of rainfall, snowmelt, and glacier melt contributions

We analyzed changes in simulated contributions of snowmelt, liquid precipitation, and glacier melt to the overall input to the hydrological system. This is slightly different from the contributions to discharge, which would be reduced by evapotranspiration and shifted in time by storage delays. In both catchments, the model shows significant increases of glacier melt ($p < 0.01$) and insignificant increases in rainfall and snowmelt ($p > 0.1$) (Table 4.9, Fig. 4.9). In the Kakshaal catchment, the model indicates that 13–29% of the changes in water input were due to increased glacier melt, and 71–87% due to increased precipitation, while for the Sari-Djaz catchment, increased glacier melt accounted for 56–85% of the changes in water input and changes in precipitation for 15–44%.

Table 4.9: Trends over the period 1957–2004 in the simulated inputs to the hydrological system by rain, snowmelt, and glacier melt for the two study catchments, calculated with the Sen's slope estimator and the Mann-Kendall test.

	Trend (mm)	*p*-value of trend	Contribution to overall trend in simulated water input (%)
		Kakshaal	
Rain	35–42	0.13–0.25	42–59
Snowmelt	16–24	0.26–0.46	22–30
Glacier melt	9–24	0.00–0.01	13–29
		Sari-Djaz	
Rain	12–15	0.82–0.91	14–17
Snowmelt	−1–24	0.41–0.99	−1–27
Glacier melt	49–73	0.00–0.01	56–85

4.5 Discussion

4.5.1 Glacier melt contribution and trend attribution for two headwater catchments of the Aksu River

Our estimates of the glacier melt contribution to runoff of 9–24% for the Kakshaal and 35–48% for the Sari-Djaz are similar to those of Zhao et al. (2013), who derived glacier melt contributions of 23% and 43.8% for the Kakshaal and Sari-Djaz, respectively. Dikikh (1993) estimated glacier melt contributions of 24% for the Kakshaal, and 33% for the Sari-Djaz Basin (the latter is called Aksu in their study; glacier contributions calculated from the values for ice melt and total runoff in Table 7 of their study). Including rainfall, snowmelt, and glacier melt from the glacier area, Aizen and Aizen (1998) estimated that the glacier areas contributed 36–38% to the total runoff in the Sari-Djaz Basin (called Aksu in their study).

Our results shed new light on the attribution of streamflow trends for the Aksu headwater catchments. Several other studies also indicated that the increase of annual runoff of these two catchments was caused by increases in temperature and precipitation, based on correlations between runoff time series of the Kakshaal and Sari-Djaz catchments with precipitation and temperature time series of meteorological stations (Jiang et al., 2007; Xu et al., 2010), but did not quantify the role of precipitation or temperature increases. Based on a hydrological simulation approach, a study by Zhao et al. (2013) found that discharge increases over the period 1970–2007 were (a) due to precipitation increases alone in the Kakshaal catchment, and (b) by 96% due to increases in precipitation and 4% due to increases in glacier melt in the Sari-Djaz catchment. In contrast, our study found that precipitation and temperature played both a role for the discharge increases in the Kakshaal catchment, and that temperature increases were the dominant driver in the Sari-Djaz catchment. While the analyzed time periods do not completely overlap, important differences of the current study in comparison to Zhao et al. (2013), which may explain these differences, are (1) the additional use of calibration criteria based on glacier mass balance for model calibration, (2) the attention paid to data consistency during temperature and precipitation interpolation, (3) the representation of glacier geometry changes, (4) the model evaluations for interannual variations in discharge and representation of observed discharge trends, and (5) the consideration of uncertainties resulting from different model parameterizations.

4.5.2 Data-based versus simulation-based trend attribution

So far, only few studies applied simulation-based approaches for the attribution of discharge trends in snow/glacier melt influenced catchments. Engelhardt et al. (2014) applied a hydrological model to three highly glacierized catchments in Norway over a time period of 52 years. The study differs from our study in that trend analysis methods were not applied, and trend attribution was not a primary aim of the study. A similar approach as in our study was used by Hamlet et al. (2005, 2007). Using a hydrological model and simulations with fixed temperature and precipitation time series, the authors investigated the influence of precipitation and temperature changes on changes in snow water equivalents, and the timing of runoff and soil moisture recharge over the western United States during the 20th century.

A higher number of studies attempts to explain streamflow trends in mountainous regions by investigating concurrent trends in meteorological variables. For example, Birsan et al. (2005) related the mostly increasing trends of annual runoff in their study on 48 catchments in Switzerland to increased temperature as the observed changes in precipitation were not sufficient to explain the streamflow changes. For glacierized basins, the analysis of concurrent trends in temperature and precipitation may also be complemented by considering glacier mass balance time series (Pellicciotti et al., 2010). Approaches based on regression analyses allow a more quantitative relation between trends and their possible drivers than those based on concurrent trends. For example, Aguado et al. (1992) and Dettinger and Cayan (1995) applied multiple linear regressions to investigate the relative contribution of temperature and precipitation changes on trends in fractional winter and spring runoff for mountain catchments in California. Moore and Demuth (2001) related variability in summer streamflow in a small glacierized catchment in British Columbia, Canada, to climate variability using a multiple linear regression approach with the possible explanatory variables temperature, temperature of previous months, winter mass balance, net mass balance of the previous year, and time.

This study applied both, a simulation-based and a regression-based approach. We see the two methods as complementary; as they are based on basically different approaches, it is expected that this increases the reliability of the results. Both approaches indicated that the streamflow changes were consistent with the changes in temperature and precipitation. Comparing the evidence provided by both methods, the evidence from the data-based approach was regarded as weaker than the one from the simulation-based approach, due to the following

reasons: (i) it was hampered by correlations between precipitation and temperature, (ii) it is not directly based on cause-effect relationships, (iii) it showed a slightly weaker performance for the representation of seasonal streamflow and trends, and (iv) other information (e.g. information on glacier mass balances from measured time series or geodetic estimates) cannot easily be integrated. The simulation approach can also give further insights, for example on the changes in rainfall, snowmelt, and glacier melt.

4.5.3 Multiobjective model calibration

For the simulation-based analysis, this study applied a multicriteria approach for model calibration. In addition to the daily discharge time series, interannual discharge variations, and two criteria based on glacier mass balance were considered. Taking into account mass balance observations is particularly important if the uncertainties in the areal precipitation are large and inputs from snowmelt, rain, and glacier melt overlap in time. Both factors hinder the identification of the glacier melt parameters from the discharge time series alone (Schaefli et al., 2005). Including objective functions on different variables can highlight discrepancies between the model and the different data sets, which can give directions where further research should focus. The trade-off between the model performance with respect to discharge and the glacier mass balance time series in the Sari-Djaz Basin indicates possible problems in the model or the data. The reason is currently unclear. The temporal dynamics of Karabatkak glacier may not be representative for the temporal dynamics of a larger number of glaciers, but this could also be caused by errors in the measured mass balance time series, errors in other model calibration or input data, or model structural errors. A new study indicates geodetic glacier mass balance estimates of −0.35 ± 0.34 m w.e. a^{-1} for the time period 1974–1999 for the Sari-Djaz Basin (Pieczonka and Bolch, in revision), which largely overlaps with the glacier mass balance criterion applied in this study. Geodetic mass balance estimates derived specifically for the Kakshaal catchment could reduce the uncertainties resulting from the use of mass balance estimates of neighboring regions.

Evaluating the model behavior with respect to the observed variability in the past is also a good model test if it is intended to apply the model to estimate streamflow changes in response to climate scenarios. This could for example reveal problems of the model which are related to instabilities of the model parameters over time (Merz et al., 2011). Studies which apply a hydrological model for climate change scenarios often evaluate the model performance only by using classical criteria like the Nash-Sutcliffe criterion on a daily or monthly time scale (e.g., Liu et al., 2010; Hagg et al., 2013). As mountain catchments often have a strong seasonal regime with high flows from snow and glacier melt during summer and low flows during the winter half year, this results in apparently high model performance for all models which can replicate this behavior (Schaefli and Gupta, 2007). However, if a model is planned to be applied for climate change scenarios, it is crucial that also the performance with respect to interannual variability and possible long term trends is included as this is the critical part of the model with respect to scenario analyses.

4.5.4 Limitations

The temporal variability of temperature and precipitation was derived from the available data and may not be representative for all subregions. Temperature or precipitation trends may be different for different subregions, but this is only captured by the analysis if it is reflected in the measured data. Precipitation data of the station Tien Shan were left in the analysis, even though it is unclear whether they are influenced by the relocation in 1998/99. Based on our previous analyses performed without including the station Tien Shan, it is assumed that omitting this station from the interpolation of precipitation data would mainly influence results of the Sari-Djaz catchment. The analysis would likely attribute a larger part of the discharge increases also to increases in precipitation, but temperature increases would probably still play the dominant role for increases in discharge.

The initial glacier elevation in the hydrological model was based on the SRTM DEM, as a DEM for the entire study area at the beginning of the study period was not available. Absolute glacier elevations are therefore underestimated. This is partly compensated by the calibrated parameters and is assumed to have little effect on the estimated trends in glacier melt and discharge. It is to note that the lowering of the glacier surface elevation during the study period, which has an influence on the trends, is represented in the model.

This study considered direct effects on evapotranspiration, for example through increasing temperatures. In contrast, indirect effects through possible changes in vegetation and in the length of the growing period were not accounted for. Thus, the increase in evapotranspiration may have been underestimated. Due to the generally low temperatures at high altitudes, this effect is, however, assumed not to be very large in the considered catchments.

While our approach accounted for uncertainties in model parameters, uncertainties from different

model structures were not considered. These could be explored using an ensemble of hydrological models.

4.6 Conclusions

This study used two approaches based on regression analysis and hydrological modeling in order to understand the recent increases in discharge in two headwater catchments of the Tarim River. For the more strongly glacierized Sari-Djaz catchment, both approaches agree on the dominant role of increasing temperatures as a cause for the discharge changes. Simulated changes in the input contributions from rain, snowmelt, and glacier melt show that increases in the overall input to the hydrological system were dominated by increases in glacier melt, which contributed 56–85% to the overall increases in water input from rain, snowmelt, and glacier melt. In the Kakshaal catchment, increasing discharge over the last decades was related to both increasing temperatures and increasing precipitation. The data-based approach suggests that a larger fraction of the discharge increase was due to changes in temperature rather than to changes in precipitation. In contrast, the simulation-based approach revealed a higher influence of the precipitation increases. These results, derived from the simulations with input time series without trends, are also reflected by the simulated changes in the input contributions from rain, snowmelt, and glacier melt. Increases in glacier melt in the Kakshaal catchment contributed 13–29% to the overall increases in water input from rain, snowmelt, and glacier melt, while the major part of the increases to the water input was from increased precipitation.

The simulations indicated glacier melt contributions to runoff of 9–24% for the Kakshaal, and 35–48% for the Sari-Djaz catchment. The analysis highlighted large uncertainties of areal precipitation in data sparse mountain regions of Central Asia. With values of 372–399 mm a^{-1} for the Kakshaal and 450–526 mm a^{-1} for the Sari-Djaz catchment, the areal precipitation estimates adjusted by model calibration indicated that the original estimates based on interpolation of station data underestimated by about 50%.

Using calibration criteria based on glacier mass balances and the interannual variation of discharge in addition to the performance with respect to daily discharge led to more credible models. This is particularly important in such data sparse catchments. In strongly seasonal discharge regimes, high performance with respect to monthly or daily streamflow can result in overestimation of the skill of the model. An evaluation with respect to the interannual variability of, e.g., seasonal flow can be a good measure to indicate whether the model responds adequately to interannual changes in temperature and precipitation, which is a prerequisite for models set up for climate change impact analyses.

While data-based approaches for trend attribution are simpler to implement and thus particularly suited to give insights for a large number of catchments, simulation-based approaches have the advantage that the cause-effect relationships can be represented in a process-based way, which suggests that their results can provide stronger evidence for trend attribution. Where possible, a complimentary application of both approaches is recommended.

Acknowledgments

The simulation results of this study can be obtained from the first author upon request. This study was conducted within the project SuMaRiO (Sustainable Management of River Oases along the Tarim River; http://www.sumario.de/), funded by the German Federal Ministry of Education and Research (BMBF grants 01LL0918A and 01LL0918B). T. Bolch was supported by the Deutsche Forschungsgemeinschaft (DFG, Code BO3199/2-1). We would like to thank Vladimir Aizen for providing the digital elevation models for the Ak-Shirak massif, and the Hydrometeorological Service of Kyrgyzstan for providing meteorological data.

5 Discussion and conclusions

5.1 Main results

Hydrological modeling and studying past hydrological changes can improve our understanding of the hydrology in mountain catchments in Central Asia. However, due to the scarcity of hydrometeorological data, hydrological modeling in mountain regions of Central Asia is challenging and involves large uncertainties. In this thesis, I first investigated which approaches and additional data can be used to improve hydrological modeling in data sparse mountain catchments. Second, I studied the possible causes for observed discharge increases in two snow and glacier melt influenced mountain catchments in Central Asia. This chapter summarizes the main achievements with respect to the research questions formulated in chapter 1.5.

Can monthly precipitation fields from downscaled reanalysis data be used for interpolating gauge observations?

Adequately estimating areal precipitation is an important problem for hydrological modeling in data sparse mountain catchments. Using downscaled reanalysis data for deriving the spatial distribution of precipitation can be a suitable alternative to purely statistical methods in such regions, particularly when it is not possible to derive the relation with the orography from the station data. This method of spatially interpolating precipitation data was applied to six headwater catchments of the Karadarya Basin in Central Asia, using reanalysis data, which had been downscaled to a resolution of 12 km with the regional climate model WRF (Weather Research and Forecasting Model; Skamarock et al., 2008). It was compared to other methods based on inverse distance weighting, multilinear regression against topographical variables, direct usage of the downscaled reanalysis data, and APHRODITE, an existing gridded data set (Yatagai et al., 2012). Results from the comparison by hydrological modeling showed that in five out of the six catchments the suggested method performed equally well as the best performing method for the study region based on multilinear regression. In the catchment where the suggested method behaved less well, this is possibly related to the insufficient representation of the orography by the regional climate model. The suggested method clearly outperformed the direct usage of WRF data, in particular with respect to the temporal variations of precipitation. Cross-validation further indicated that the suggested method is more robust against omitting individual stations than multilinear regression.

How can areal precipitation estimates be compared by hydrological modeling?

As areal precipitation estimates for sparsely gauged mountain catchments can show very large differences, methods for evaluating these estimates are important. However, in catchments with few precipitation gauges this is not straightforward, as the traditionally applied cross-validation can be misleading if the precipitation gauges are not in representative locations. The evaluation of precipitation estimates by hydrological modeling and a comparison of simulated and observed discharge can be a well-suited alternative. This approach has been further advanced in this thesis in order to differentiate between errors in the precipitation volume and errors in the temporal dynamics. Parameter uncertainties were considered within the calibration framework and were taken into account for evaluating the different precipitation data sets. The application of this approach to headwater catchments of the Karadarya Basin showed that the investigated precipitation estimates mainly differed in the volume bias. Considering parameter uncertainties in the calibration approach was essential to demonstrate that the differences between the investigated precipitation estimates were mostly larger than the parameter uncertainties.

What are the trade-offs between good model performance with respect to discharge and snow cover?

Using snow cover data in addition to discharge data for model calibration can be a good way to increase the internal consistency of a hydrological model. Trade-offs between good model performance with respect to discharge and with respect to snow cover were investigated in six headwater catchments of the Karadarya Basin using a multiobjective optimization algorithm. The results revealed only small trade-offs between the two calibration criteria. This means that in the investigated catchments good model performance for snow cover and discharge could be achieved with the same parameterization. However, small trade-offs do not mean that the snow cover observations were not important for constraining the model parameters. Additional Monte Carlo simulations demonstrated that good discharge simulations could also be obtained with large snow cover errors. Including snow cover

observations hence increased the internal consistency of the model.

Furthermore, this study showed that the need for threshold values for the comparison between simulated snow water equivalent and observed snow cover areas can be avoided by adding a parameterization for the calculation of snow cover area to the hydrological model (Liston, 2004). The parameterization assumes a lognormal distribution of snow water equivalents and introduces one additional calibration parameter. This approach for evaluating semidistributed models with respect to remotely sensed snow cover data is also recommended for other hydrological models which are to be calibrated to snow cover data.

How many snow cover scenes should be used for model calibration?

In this thesis, it was investigated how the added information content of additional snow cover scenes changes with an increasing number of scenes in the calibration period. My results showed that it was necessary to include 10–16 snow cover scenes for model calibration. When using too few snow cover scenes, this bears the risk that the selected scenes show an atypical deviation between observation and model, which results in the selection of parameters not representative for the catchment. This is the first time the added value of increasing the number of snow cover scenes in the calibration period has been analyzed. The results are practically relevant as an indication of how many snow cover scenes should be included for model calibration. Particularly when the snow cover information first needs to be derived from satellite raw data, this information can be very valuable.

To what extent can streamflow increases in two headwater catchments of the Aksu River be attributed to increases in temperature and precipitation?

This study investigated possible causes of recent streamflow increases in two headwater catchments of the Aksu River using two approaches based on regression analysis and hydrological modeling. For the Sari-Djaz catchment, the two approaches agreed that the increase in temperature is the dominant driver for the observed discharge increase. For the Kakshaal catchment, the study indicated that both increases in temperature and increases in precipitation contributed to the increases in discharge. However, while the regression-based approach pointed toward a higher influence of temperature compared to precipitation on the increasing discharge, the simulation-based approach indicated a slightly higher influence of the precipitation increases. The simulation-based approach additionally enabled an analysis of the changes in simulated water sources from glacier melt, snowmelt, and rainfall. The results indicated that increases in glacier melt accounted for 56–85% of the increases in water input to the hydrological system for the Sari-Djaz catchment. In the less strongly glacierized Kakshaal catchment, a larger fraction of the increases in water inputs to the hydrological system was from precipitation than from glacier melt, which contributed 13–29% to the overall increases in water input.

5.2 Discussion and future research questions

5.2.1 Interpolation of station-based precipitation time series with precipitation fields from downscaled reanalysis data

The method of interpolating station-based precipitation time series with precipitation fields from downscaled reanalysis data depends on the assumption that the precipitation field from atmospheric modeling represents the spatial precipitation field in a realistic way. When applying this method to a new region, it is therefore necessary to check this assumption. As a first point, it can be tested whether the resolution of the atmospheric model is sufficient to resolve the main topographic features of the studied region. If relevant mountain ranges are disregarded by the model, the regional climate model will not be able to simulate corresponding precipitation features, and one should first seek a higher resolution of the regional climate model before the data can be used for interpolation. In a next step, the simulated precipitation data can be evaluated quantitatively as well as qualitatively. Quantitative evaluation involves comparing the simulated precipitation with observed precipitation time series from available stations, even though this comparison is compromised by the scale difference between the grid-based model data and the point-based station data. Further qualitative evaluation of the spatial precipitation patterns may be performed considering documented precipitation features in the region or more general knowledge about precipitation patterns in mountainous terrain.

Using precipitation fields from high-resolution reanalysis data was also pursued for the study in the Aksu catchment. However, the study area was only covered by the first model domain of the WRF model with a resolution of 36 km × 36 km, but did not fall inside the second model domain with a resolution of 12 km × 12 km. Since the topography

was only represented in a relatively crude way, it was doubtful whether the precipitation fields from the relatively coarse downscaled reanalysis data could provide a realistic precipitation field for the Aksu region. The evaluation with measured precipitation time series revealed larger biases and lower correlations to observations than in the Karadarya Basin. As further downscaled data with a higher resolution were not available, downscaled reanalysis data were not applied for interpolating station-based precipitation time series for the study in the Aksu catchment.

Due to increasing computer power, downscaled reanalysis data at high resolutions become increasingly available, offering a good potential for the wider use of such data. For example, data sets of WRF-downscaled reanalysis data are available for the Arctic (Arctic System Reanalysis; Polar Meterology Group/Byrd Polar Research Center, 2012) and for a region in high Asia between c. 25–40°N and 72–100°E (High Asia Refined Analysis; Maussion et al., 2014). The direct use of downscaled reanalysis precipitation data in a hydrological model can be problematic, as also shown in this thesis (chapter 2). A method which combines spatial fields from downscaled reanalysis data with measured precipitation time series, like the one suggested in this study, can be a useful alternative and may therefore also be a suitable method for deriving precipitation estimates in other data sparse mountain regions.

5.2.2 Evaluating areal precipitation estimates by hydrological modeling

Evaluating areal precipitation estimates by hydrological modeling and the comparison of simulated and observed discharge is particularly well suited for mountain areas with sparse data and a high variability of precipitation, as results from cross-validation may not be meaningful in such regions. However, the method cannot be transferred to regions where the water balance cannot be closed due to significant groundwater in- or outflows or water abstractions for which the data is not available. Furthermore, the method may be inconclusive if the differences between precipitation data sets are only small. While the evaluation procedure may be influenced by errors in the observed discharge data as well as the model structure and input data, in data sparse mountain catchments, these uncertainties are likely smaller than the differences between different precipitation estimates. In many mountain regions precipitation is considerably higher than evapotranspiration, which means that possibly large errors in evapotranspiration only result in small errors of the precipitation estimate. This makes approaches based on the water balance well suited for many mountain regions (Müller-Lemans et al., 1997).

Using hydrological modeling and observed discharge enabled a correction of the original areal precipitation estimates for the investigated catchments. Since it was later discovered that the catchment Gulcha in the Karadarya Basin is influenced by abstractions, the precipitation estimates for this catchment are likely too low. If data on the abstractions become available, an improved areal precipitation estimate can be made using naturalized streamflow data for model calibration.

In this thesis, the introduction of a precipitation bias factor in order to differentiate between deviations in the precipitation volume and in the temporal performance was suggested. The precipitation bias factor was used as a temporally constant factor, which was included in the model calibration. This enabled us to specify by how much the original areal precipitation estimate over- or underestimated precipitation, and it was possible to separately evaluate the temporal performance of the precipitation data set. While this study used a temporally constant factor, in reality this factor is likely to change over time. The value of the precipitation bias factor may change depending on, e.g., weather pattern, wind direction, precipitation type (solid or liquid), precipitation intensity, or wind speed during the event. Temporally variable rainfall errors have been considered using storm-dependent rainfall multipliers (Kavetski et al., 2002; Kavetski et al., 2006; Vrugt and Robinson, 2007), or time-dependent stochastic rainfall multipliers (Reichert and Mieleitner, 2009). However, a temporally varying precipitation factor may increase parameter uncertainties and involves the risk that such a factor compensates for other errors in the model. Further research should investigate in what cases a temporally varying precipitation factor can be introduced. For example, fast reacting catchments, where a large part of the rainfall directly leads to runoff, offer a high potential for the identification of a temporally variable precipitation factor. In contrast, the temporal variation of the precipitation factor cannot be resolved during times when precipitation is stored in the catchment, e.g. as snow.

5.2.3 Satellite-derived snow cover for hydrological model calibration

The study in the Karadarya Basin clearly showed that using satellite-derived snow cover in addition to discharge data increased the model consistency. If the model is only calibrated to discharge, the model performance with respect to discharge was similar, but the snow cover error could be consider-

ably higher. In the investigated catchments, significant trade-offs between good model performance regarding discharge and regarding snow cover were only found for the catchment Gulcha. These larger trade-offs were likely caused by deficiencies of the discharge data, which are impaired by abstractions upstream of the discharge gauge. While, apart from this catchment, trade-offs for the investigated catchments were only small, larger trade-offs may be found in other catchments, or using other hydrological models. Such trade-offs can indicate problems with the calibration data, other input data, the model structure, or a combination of these, and the reasons should be investigated.

Despite the advantages of including snow cover data for model calibration, this approach was not applied in the Aksu Basin due to particularities of the snow regime that the model was not able to represent. In the Aksu Basin, the winter season is characterized by very low precipitation. Based on MODIS snow cover data, one can observe that snowfall events during winter only result in a short-lasting snow cover, which disappears during the following days, even though interpolated temperatures are clearly below 0°C. It is therefore unlikely that the disappearance of the snow cover can be explained by snowmelt, and probably other processes played a role. Such processes could be snow redistribution by wind, which might accumulate the thin snow cover in small wind-sheltered parts not captured by the binary satellite data, or sublimation. Since the current version of the WASA model cannot represent these processes, including these snow cover data would likely shift the model parameters toward wrongly generating snowmelt at very low temperatures in order to reproduce the disappearance of the snow cover.

While these snow cover data could not be applied for calibrating the current version of the WASA model in the Aksu Basin, they might be helpful for implementing these processes into the model. In order to model sublimation, one could replace the temperature-index method for snowmelt by an energy-balance model which also includes sublimation. However, moving toward such a more physics-based approach also increases the data requirements. The current version of the WASA model does not include snow redistribution by wind before the onset of snowmelt. The variability of the snow water equivalent is represented by a parameterization, which comes into effect only after the beginning of snowmelt. As the variability of wind speed and directions cannot easily be captured, the physical redistribution of snow by wind is difficult to describe. A further challenge is to quantify the respective role of sublimation and wind transport. Field observations could give more insights into this question.

In this thesis, it was analyzed how the added information content of additional snow cover scenes in the calibration period changes with an increasing number of scenes. The study included six catchments and was performed for two calibration periods. The results are therefore expected to be transferable to other catchments with similar physiographic settings and snowfall regimes. However, further research is required for a more general understanding of this question. The following hypotheses emerged from my research and could be used to guide further investigations. The first hypothesis is that a lower number of snow cover scenes is required for catchments with a large elevation range. The study in the Karadarya Basin indicated that a higher number of snow cover scenes was required for the catchments Tosoi and Donguztoo, which are the two smallest catchments and also the two catchments with the smallest elevation range. As the data is summarized by catchment and elevation zone, a snow cover scene results in fewer data points in a catchment with a smaller elevation range than in a catchment with a larger elevation range. Furthermore, data points within or close to the transition elevation range between snow-free and snow-covered conditions are expected to be most informative. For a catchment with a smaller elevation range, the time period over which such most informative data points can occur is shorter than in catchments with a larger elevation range, where the disappearance of snow can be tracked in a higher number of elevation zones and thus over a longer time period. A second hypothesis is that a higher number of snow cover scenes is required with an increasing snow cover error of the calibrated model. If the snow cover error is large, more scenes are required to avoid an unrepresentative selection of snow cover scenes. For example, it was observed that the model had more difficulties in representing the snow cover during periods of intermittent snowfall events in the melting phase. This was related to uncertainties in the precipitation input, the precipitation phase (solid or liquid), and in the assumptions of the snow cover parameterization for small snowfall events during the melting phase. More snow cover scenes are likely needed if they are from such periods where the snow cover error is higher. More insights into these questions could be gained by evaluating the added information of an increasing number of snow cover scenes in the calibration period for further catchments. These catchments should have different characteristics than the catchments investigated in this thesis (e.g., other snowfall regimes) to investigate the transferability of the results from this study to different study areas, and they should include catchments with high and low elevation ranges.

This study used satellite-derived snow cover data as an additional variable for model calibration and

showed that these data delivered useful information for increasing the internal consistency of the model. A disadvantage of snow cover data is that they do not contain information on the amount of water stored in the snow cover. It has been shown that snow heights derived from airborne laser scanning are more powerful than snow cover data in constraining model parameters (Schöber et al., 2014). However, these data depend on flight missions and are therefore only available for small to medium size catchments and associated with high costs. Satellite-derived products on snow heights or snow water equivalents may be derived from active or passive microwave sensors. However, uncertainties of these data are large, particularly in mountainous areas (Botteron et al., 2013). Active microwave sensors like Synthetic Aperture Radar are affected by large errors resulting from inhomogeneities of the snowpack and vegetation, and furthermore only have low recurrence frequencies. Passive microwave sensors are limited by their low spatial resolution (e.g. AMSR-E 25 km), errors caused by varying snow densities and grain sizes, and the fact that the data are not reliable for snow water equivalents of >15cm (Hancock et al., 2013). These problems currently hamper the application of satellite-derived snow water equivalent products for the calibration of hydrological models in mountainous areas.

5.2.4 Further data for multiobjective model calibration

Besides snow cover and discharge data, also glacier mass balance data were used for model calibration in this thesis. At the end of this section, further data with a high potential for increasing model consistency, such as remote sensing data on soil moisture, evapotranspiration, and total water storage variations, as well as stable isotope and geochemical tracers, will be discussed.

Glacier mass balance data were crucial for the calibration of the hydrological model in the catchments of the Aksu Basin. Due to the large uncertainties in precipitation there is otherwise a high risk of compensating an underestimation of glacier melt runoff by too high runoff from snowmelt and rainfall, and vice versa. Since glacier mass balance data were included in the model calibration, the hydrological model enabled not only investigating whether changes in discharge were consistent with the changes in temperature and precipitation, but also whether these changes are also consistent with the glacier mass balance observations.

Mass balances representative for a larger area, as derived by geodetic methods, are particularly useful for calibrating hydrological models at larger scales. Ideally, the area of the geodetic estimate matches with the area of the investigated catchment. For the Sari-Djaz catchment, a mass balance estimate which covers the complete catchment is now available (Pieczonka and Bolch, in revision). It largely confirms the previous estimate based on two close-by regions. The Kakshaal catchment is geographically at a farther distance to the available geodetic estimates so that differences to the available estimates might be larger. A geodetic glacier mass balance estimate specifically for the Kakshaal catchment would be very useful to increase the confidence in the hydrological model for this catchment. Uncertainties of these geodetic estimates are currently still rather large. These uncertainties are related to the accuracy of the DEM, the penetration of the radar signal in snow and ice, and the assumptions for missing pixels in the accumulation areas (Surazakov and Aizen, 2006; Pieczonka et al., 2013). If the precipitation amount is corrected based on observed discharge and hydrological modeling, uncertainties of the glacier mass balance can result in large uncertainties of the precipitation bias factor in highly glacierized catchments with low precipitation. For example, the precipitation factor in the Sari-Djaz catchment varied between 1.35 and 1.58 for catchment average mass balances between −0.40 and −0.77 m w.e. a^{-1} (for the period 1957–2004). A reduced uncertainty range of the geodetic glacier mass balance estimates is crucial for being able to better constrain model parameters, particularly the melt parameters and the precipitation bias factor.

In addition to the geodetic estimates, a mass balance time series of a close-by glacier, measured using the glaciological method, was applied for model calibration. Despite possibly large differences in absolute mass balances values, temporal mass balance variations are regarded as well correlated at the scale of mountain ranges (Huss, 2012), and therefore the correlation between the measured time series and the catchment average time series was used as calibration criterion. This way, the advantages of both data sources could be combined. The criterion on the cumulative glacier mass loss was based on the geodetic estimates, which are representative for a larger area but not resolved in time, while the criterion on the temporal variation was based on glaciological measurements, which reflect the interannual variations but might not be representative in terms of absolute mass balance values.

Particularly for the Sari-Djaz catchment, model parameterizations which resulted in good performance with respect to discharge resulted in low correlation between the simulated catchment average mass balance time series and the mass balance time series at Karabatkak glacier. Such trade-offs can point to possible problems of the model or the data and can be used as indication where further research should focus. The fact that by evaluating

the temporal variation of a single glacier with the average behavior of a large number of glaciers different entities are compared could be a possible cause for the trade-off. Even though typically temporal mass balance variations are assumed to correlate well at the scale of mountain ranges (Huss, 2012), the spatial variations may in this case be larger. Another possible cause for this trade-off are errors in the glacier mass balance data itself. The accumulation area of Karabatkak glacier is difficult to access so that measurements were probably only performed over part of the glacier and extrapolated to the whole glacier. Changes in the methodology or accessibility of the glacier in individual years could have caused biased results for some years. Further possible causes for this trade-off are errors in other model input or calibration data, or errors in the model structure.

In this study, the geodetic glacier mass balances were used as lumped estimates. Further research may also compare the spatial variation of the simulated glacier mass balances with those provided by the geodetic estimates, as a further aspect for evaluating the internal consistency of the model. When investigating to what extent the model is able to represent spatially distributed mass balances, some modifications of the model might be considered. For example, differences in the energy input by aspect could be taken into account, either within the current approach by varying model parameters by aspect, or by changing to an enhanced temperature-index (Hock, 1999; Pellicciotti et al., 2005), or an energy-balance approach. In case the behavior of individual glaciers is mostly influenced by local variations in temperature and precipitation which are not captured by the sparse station network, or if glacier surges play a larger role, it may not be possible for the model to reflect the spatial variability in mass balances.

While glacier melt is expected to be lower under thick debris cover (Östrem, 1959; Reid et al., 2012), it is still not clear whether these lower melt rates for glaciers with debris cover are also valid at the regional scale (Gardelle et al., 2012; Kääb et al., 2012). In this thesis, different melt rates under clean ice and debris-covered glaciers were not considered. Since the cumulative glacier mass loss was calibrated against geodetic estimates which include glaciers with and without debris cover, the effect of debris cover is implicitly taken into account. In the future, the model might be modified to consider varying melt parameters as a function of the debris cover (Juen et al., 2014). Since the effect of the debris cover on ablation strongly depends on the thickness of the debris layer, with thin debris layers resulting in enhanced melt and thick debris layers resulting in reduced melt (Östrem, 1959), this also requires estimates of the debris layer thickness. These might be derived from thermal band satellite imagery (Foster et al., 2012). Spatially distributed geodetic mass balances could aid in estimating melt parameters as a function of debris cover thickness.

The internal consistency of hydrological models for data sparse mountain catchments may be further increased by considering data on other variables than snow or glacier mass balance. Remotely sensed data are, for example, also available for soil moisture, evaporation, or total water storage, and further information can also be gained from stable isotopes or geochemical tracers.

Satellite-derived soil moisture data can be obtained from different types of microwave sensors (Wagner et al., 2007). These data however only reflect soil moisture of the top few centimeters, while most hydrological models represent deeper soil layers. In order to enable a comparison between the simulated and satellite-derived soil moisture data, either a surface soil layer needs to be introduced into the hydrological model (Parajka et al., 2009), or an exponential filter may be applied to the remote sensing data so that the data can serve as an indicator for soil moisture of a deeper soil layer (Wagner et al., 1999). For mountainous areas, comparisons of satellite-derived to observed (Brocca et al., 2011) or simulated soil moisture showed, however, large deviations (Laguardia and Niemeyer, 2008; Parajka et al., 2009; Rüdiger et al., 2009). These deviations are caused by higher noise in the remote sensing soil moisture data in high relief terrain, but possibly also by higher errors of the hydrological models (e.g. due to higher uncertainties in soil depths, or larger uncertainties in the meteorological input data) (Rüdiger et al., 2009). A further advancement of remote sensing methods for soil moisture in high relief terrain will be of advantage for integrating these data with hydrological modeling in mountain catchments, and could contribute to increasing the internal model consistency of hydrological models in mountain catchments in Central Asia.

Different methods have been developed to calculate evapotranspiration based on remote sensing data. Evapotranspiration may be estimated as a residual of the surface energy balance, like in the frequently applied Surface Energy Balance Algorithm for Land (SEBAL) (Bastiaanssen et al., 1998). It estimates evapotranspiration based on land surface temperatures derived from remote sensing data in the thermal-infrared band, further satellite data on albedo, the Normalized Difference Vegetation Index, and vegetation height, and additional meteorological data on wind speed, air temperature, and the water vapor pressure (Gao et al., 2008). Evapotranspiration estimates derived with SEBAL were for example used for constraining parameters of a hydrological model applied in a data sparse river basin in Zambia, where they enabled the estimation of spatially variable parameters (Winsemius et al.,

2008). Due to the large uncertainties in the areal precipitation estimates in the catchments investigated in this thesis, a calibrated precipitation bias factor was introduced. The calculated evapotranspiration therefore is particularly important since it directly influences the water balance and thus also the precipitation bias factor. In future research, it would therefore be desirable to compare the evapotranspiration calculated by the hydrological model also with evapotranspiration estimates based on remote sensing data.

For large catchments, total water storage variations from GRACE (Gravity Recovery And Climate Experiment) can be a further data source for hydrological model calibration (Lo et al., 2010; Werth and Güntner, 2010). GRACE observes variations in the gravitation field, from which variations in water mass can be deduced. These estimates include storage variations in snow, glacier ice, groundwater, soil moisture, and lakes. However, the data has a maximum resolution of 400 km (Wahr et al., 2004), with a better performance observed at larger scales (Güntner, 2008). The catchments investigated in this thesis are therefore too small for a comparison with GRACE data. However, such an evaluation will be useful if in a future step the simulated area is extended to a larger region.

When applying hydrological models in catchments influenced by snow and glacier melt, an important question is whether the runoff contributions from rainfall, snowmelt, and glacier melt are simulated correctly. Hydrograph separation based on tracer methods may be used to evaluate the simulated runoff contributions. Stable isotopes of water (^{2}H and ^{18}O) are ideal conservative tracers of water sources. Hindshaw et al. (2011) evaluated the simulated runoff source contributions of their hydrological model by simulating the oxygen isotopic composition in streamflow with their hydrological model. The simulated oxygen isotopic composition could then be compared to the observed values. For cases where only groundwater and glacier melt water contribute to streamflow, the glacier melt contribution may be estimated using a two-component mixing model with glacier melt and groundwater as end members (Nolin et al., 2010). In many cases, however, snowmelt and rainfall overlap with glacier melt so that more end members need to be considered. This requires a multi-component mixing model and additional tracers (e.g., two tracers are needed for a three-component mixing model). Geochemical tracers may be used in combination with isotope tracers (Klaus and McDonnell, 2013). It would be desirable to apply these methods also for evaluating runoff contributions in mountain catchments in Central Asia. First steps toward a regular sampling program have recently been started in the Ala Archa catchment, located in the northern Tien Shan.

5.2.5 Attribution of streamflow trends in two high mountain catchments using a data-based and a simulation-based approach

In this thesis, a data-based and a simulation-based method for trend attribution in snow and glacier melt influenced catchments were applied. This also illustrated the advantages and disadvantages of the two approaches. Data-based approaches, like multilinear regression in this thesis, are typically relatively fast and easy to apply. An important disadvantage is that it is not clear whether the identified relationships are due to physical cause-effect relationships or statistical artifacts. Furthermore, the interpretation may be hampered by correlations between possible drivers, as also shown in my analyses. Using a hydrological modeling approach has the advantage that changes in the processes are linked to changes in drivers via causal relationships. Furthermore, other data, e.g. glacier mass balance data in this study, may be integrated with the hydrological model so that it is also possible to evaluate whether the assumed causes for the changes are also consistent with these data. The simulation-based approach also enables investigating changes in variables which are typically not measured, for example changes in glacier melt. Disadvantages of the hydrological modeling approach are the greater effort for setting up the model, and the introduction of model parameter and structural uncertainties.

As recommended by Merz et al. (2012), trend attribution studies should provide evidence of consistency with the assumed drivers of change and of inconsistency with alternative drivers, as well as a confidence statement. The attribution study in this thesis abstained from giving a quantitative confidence statement. A quantitative confidence interval could, for example, be based on the different model parameterizations, but the ensemble of model parameterizations in this study was regarded as too small for such an analysis, and the so derived confidence interval would cover only one aspect of the uncertainties. However, a qualitative evaluation of factors influencing the confidence in the attribution study can be provided (Table 5.1).

Generally, the confidence in attribution studies may be increased by applying several, possibly basically different approaches instead of relying on only one approach. The regression and simulation-based approaches can be regarded as independent. Even when using different approaches, these, however, still rely on the same data, which might contain errors or may not be able to capture spatially variable trends. For example, in this study, the same

runoff, temperature, and precipitation time series were used in both approaches.

For a reliable trend attribution study, spatially well distributed homogeneous meteorological time series should be available. Spatially variable trends in meteorological variables can only be captured by a sufficiently dense station network. Furthermore, the existence of a high number of meteorological stations facilitates the identification of inhomogeneities in the time series.

Trend attribution studies in data sparse catchments are therefore associated with higher uncertainties. A varying number of available stations over time can generate inhomogeneities in the interpolated data set if this fact is not considered appropriately during the interpolation of the meteorological data. In the present study, this problem was minimized by interpolating anomalies with respect to a reference period, instead of directly interpolating the meteorological time series.

Results of regression-based analyses should be questioned if no physical explanation for the identified statistical relationships can be found. For the regression models in the investigated catchments the identified relationships were physically plausible. The identification of the regression model, i.e. explanatory variables and regression coefficients, may however still have been impaired by correlations between the explanatory variables.

Like the regression based approach, the simulation-based approach is influenced by the quality of the available data, which in this case also includes further model input data. Parameter uncertainties were considered by selecting a number of parameter sets from the multiobjective calibration for the trend attribution analysis. Uncertainties from different model structures are likely larger than the parameter uncertainties. The resulting model structural uncertainties can be analyzed by including an ensemble of different hydrological models in the analysis, which was, however, outside the scope of the current analysis.

Different model structures can represent a different level of physical realism. For an attribution analysis, the model should consider all relationships which are relevant for the observed changes. For example, if glacier geometry changes are not represented in the model, their effect cannot be investigated, possibly leading to a wrong attribution statement.

Table 5.1: Factors for evaluating the credibility of a trend attribution study.

	Factor	Evaluation for the present study	
General	Application of several, basically different approaches	✓	Application of regression-based and simulation-based approach
	Inspection of various alternative drivers	✓	No evidence for a significant influence of water management; therefore unlikely a possible alternative driver
		-	Possible changes in wind speed, radiation, or humidity were not investigated due to lack of data
Data situation	Spatially well distributed, homogeneous time series	-	Meteorological data were only sparse, an inhomogeneity at Tien Shan station could not be excluded
		-	Measured time series for radiation, humidity, and wind speed not available
	Dealing with a variable number of meteorological inputs in an appropriate way	✓	Interpolation of anomalies instead of direct interpolation of the data
Regression-based approach	No correlation between the explanatory variables	-	Correlations were observed
	Selection of explanatory variables plausible	✓	Yes, see section 4.4.1
Simulation-based approach	Incorporation of other data besides discharge into model calibration /evaluation.	✓	Multiobjective calibration taking into account a measured glacier mass balance and geodetic mass balance estimates
	Evaluation of model parameter uncertainties	✓	28 and 6 model parameterizations were considered for the attribution analysis
	Physical realism of the model	✓	Glacier geometry changes were explicitly considered
		-	Temperature index approach for snow and glacier melt not suited for investigating the specific influence of changes in radiation
	Evaluation of model structural uncertainties	-	No, only one model was applied

The fact that the model applied in this thesis performed well with respect to interannual variations and long-term trends in discharge indicates that the model responds well to interannual changes in temperature and precipitation and can reproduce the changes observed over the study period. Evaluating the model with respect to interannual variations and long-term trends can also be a useful test if it is intended to apply the model for investigating the impacts of future climate change. Models and parameterizations which passed this test are more likely to adequately represent the physical processes and can therefore be expected to provide more reliable simulations under climate change. Simulations for future climate projections are however very challenging, since future changes are likely greater than the observed changes, and the model performance for these future conditions cannot be tested. In the present study, a number of model parameterizations was excluded because of their poor performance with respect to the observed trends over the complete study period. This would not have been possible if only data for the calibration period were available. A number of studies showed that calibrating a hydrological model to periods with contrasting climate conditions may result in a loss of model performance when parameters calibrated to one period are applied to another period (Vaze et al., 2010; Merz et al., 2011; Coron et al., 2012; Seiller et al., 2012). For example, Merz et al. (2011) calibrated a hydrological model to 273 catchments in Austria for periods with different climate conditions, and found that, in response to observed increases in precipitation and temperature, the model reacted with higher streamflow, while measured streamflow in contrast remained mostly stable. Similarly, Coron et al. (2012) found a low transferability between periods with different levels of precipitation in their study on 216 catchments in southeast Australia. The nature of the parameter nonstationarity may help to identify which processes are not adequately represented by the model and may guide model improvement (Westra et al., 2014). More research is required to assess which model structures and which calibration strategies are well suited for investigating past and future hydrologic changes.

5.3 Concluding remarks

The approaches developed in this thesis for mountain catchments in Central Asia are also transferrable to other data sparse mountain regions, such as the Andes or the Himalaya. For example, using precipitation patterns from downscaled reanalysis data for the interpolation of station-based precipitation data can be a good solution for interpolating precipitation data in regions where the precipitation pattern cannot be derived from the station data. The demonstration of larger snow cover errors when neglecting snow cover data for model calibration hopefully encourages other modelers to integrate snow cover data in the model calibration/evaluation process for achieving a higher model consistency. Due to the low density of precipitation gauges in many mountain regions, the comparison of precipitation data sets using hydrological modeling and observed discharge should be the preferred way of comparing precipitation data sets also in other mountain headwater catchments (if there are no significant influences of water management).

Based on hydrological modeling, this study improved our knowledge on the hydrological cycle in the Karadarya and Aksu catchments, including a better understanding of the water balance, hydrological processes, and past hydrological changes. It could be shown that global precipitation data sets strongly underestimate precipitation in the Karadarya headwater catchments. The additional use of snow cover data for model calibration enabled us to better constrain when and where snowmelt occurred. Results of the study in the Aksu headwater catchments indicated that the discharge increase in these catchments strongly relates to glacier shrinkage and increased glacier melt. This means that the additional water resources stem to a large extent from depletion of glacier storage. Once these storages are depleted, a decrease of the water resources must be expected. In the next step, the hydrological models will be applied in combination with climate change scenarios in order to estimate hydrologic impacts of projected climate change in the future. This contributes to improving the scientific basis for the planning of water management.

Hydrological modeling in mountain catchments of Central Asia is challenging, due to the low density of the measurement networks and the large uncertainties involved. However, water resources of these mountain regions are important for a large number of people, and hydrological modeling in these regions is therefore very relevant to gain an improved understanding of the regional hydrology. This thesis demonstrated that by including additional data derived from atmospheric modeling and remote sensing the credibility of hydrological models can be increased also in data sparse mountain regions.

References

Abbink, K., L.C. Moller, and S. O'Hara (2010): Sources of mistrust: An experimental case study of a Central Asian water conflict, *Environmental and Resource Economics*, *45*(2), 283-318.

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

Aguado, E., D. Cayan, L. Riddle, and M. Roos (1992): Climatic fluctuations and the timing of West Coast streamflow, *J. Clim.*, *5*(12), 1468-1483.

Aizen, V.B., E.M. Aizen, and J.M. Melack (1995): Climate, snow cover, glaciers, and runoff in the Tien-Shan, Central-Asia, *Water Resour. Bull.*, *31*(6), 1113-1129.

Aizen, V.B., E.M. Aizen, J. Dozier, J.M. Melack, D.D. Sexton, and V.N. Nesterov (1997): Glacial regime of the highest Tien Shan mountain, Pobeda-Khan Tengry massif, *J. Glaciol.*, *43*(145), 503-512.

Aizen, V.B. and E.M. Aizen (1998): Estimation of glacial runoff to the Tarim River, central Tien Shan, in: IAHS Publ. - Series of Proceedings and Reports, Hydrology, Water Resources and Ecology in Headwaters, Meran/Merano, Italy, 191-198.

Aizen, V.B., V.A. Kuzmichenok, A.B. Surazakov, and E.M. Aizen (2006): Glacier changes in the central and northern Tien Shan during the last 140 years based on surface and remote-sensing data, *Annals of Glaciology*, *43*(1), 202-213.

Anderson, E. (2006): Snow Accumulation and Ablation Model – SNOW-17, NOAA, Silver Spring, MD., NOAA's National Weather Service Hydrology Laboratory NWSRFS User Manual, 61 pp, (Available from http://www.nws.noaa.gov/oh/hrl/nwsrfs/users_manual/part2/_pdf/22snow17.pdf).

Andreadis, K.M. and D.P. Lettenmaier (2006): Assimilating remotely sensed snow observations into a macroscale hydrology model, *Adv. Water Resour.*, *29*(6), 872-886.

Artan, G., H. Gadain, J.L. Smith, K. Asante, C.J. Bandaragoda, and J.P. Verdin (2007): Adequacy of satellite derived rainfall data for stream flow modeling, *Nat. Hazards.*, *43*(2), 167-185.

Backhaus, K., B. Erichson, W. Plinke, and R. Weiber (2003): Multivariate Analysemethoden: Eine anwendungsorientierte Einführung, Springer, 583 pp.

Barnett, T.P., J.C. Adam, and D.P. Lettenmaier (2005): Potential impacts of a warming climate on water availability in snow-dominated regions, *Nature*, *438*(7066), 303-309.

Basist, A., G.D. Bell, and V. Meentemeyer (1994): Statistical relationships between topography and precipitation patterns, *J. Clim.*, *7*(9), 1305-1315.

Bastiaanssen, W.G.M., M. Menenti, R.A. Feddes, and A.A.M. Holtslag (1998): A remote sensing surface energy balance algorithm for land (SEBAL) - 1. Formulation, *J. Hydrol.*, *212*(1-4), 198-212.

Bates, B.C., R.E. Chandler, S.P. Charles, and E.P. Campbell (2010): Assessment of apparent nonstationarity in time series of annual inflow, daily precipitation, and atmospheric circulation indices: A case study from southwest Western Australia, *Water Resour. Res.*, *46*, W00H02.

Beek, T.A.D., F. Voss, and M. Florke (2011): Modelling the impact of Global Change on the hydrological system of the Aral Sea basin, *Physics and Chemistry of the Earth*, *36*(13), 684-695.

Behrangi, A., B. Khakbaz, T.C. Jaw, A. AghaKouchak, K.L. Hsu, and S. Sorooshian (2011): Hydrologic evaluation of satellite precipitation products over a mid-size basin, *J. Hydrol.*, *397*(3-4), 225-237.

Bengtsson, L., K.I. Hodges, and S. Hagemann (2004): Sensitivity of the ERA-40 reanalysis to the observing system: determination of the global atmospheric circulation from reduced observations, *Tellus A*, *56*(5), 456-471.

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

Birsan, M.-V., P. Molnar, P. Burlando, and M. Pfaundler (2005): Streamflow trends in Switzerland, *J. Hydrol.*, *314*(1), 312-329.

Bitew, M.M. and M. Gebremichael (2011a): Evaluation of satellite rainfall products through hydrologic simulation in a fully distributed hydrologic model, *Water Resour. Res.*, *47*, W06526.

Bitew, M.M. and M. Gebremichael (2011b): Assessment of satellite rainfall products for streamflow simulation in medium watersheds of the Ethiopian highlands, *Hydrol. Earth Syst. Sci.*, *15*(4), 1147-1155.

Bothe, O., K. Fraedrich, and X.H. Zhu (2012): Precipitation climate of Central Asia and the large-scale atmospheric circulation, *Theor. Appl. Climatol.*, *108*(3-4), 345-354.

Botteron, C., N. Dawes, J. Leclere, J. Skaloud, S.V. Weijs, and P.A. Farine (2013): Soil moisture and snow properties. Determination with GNSS in Alpine environments: challenges, status, and perspectives, *Remote Sensing*, *5*(7), 3516-3543.

Braun, L. and M. Aellen (1990): Modelling discharge of glacierized basins assisted by direct measurements of glacier mass balance, *IAHS Publ.*, *193*, 99-106.

Brocca, L., F. Melone, T. Moramarco, W. Wagner, V. Naeimi, Z. Bartalis, and S. Hasenauer (2010): Improving runoff prediction through the assimilation of the ASCAT soil moisture product, *Hydrol. Earth Syst. Sci.*, *14*(10), 1881-1893.

Brocca, L., S. Hasenauer, T. Lacava, F. Melone, T. Moramarco, W. Wagner, W. Dorigo, P. Matgen, J. Martinez-Fernandez, P. Llorens, J. Latron, C. Martin, and M. Bittelli (2011): Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe, *Remote Sens. Environ.*, *115*(12), 3390-3408.

Brown, D.P. and A.C. Comrie (2002): Spatial modeling of winter temperature and precipitation in Arizona and New Mexico, USA, *Climate. Res.*, *22*(2), 115-128.

Chen, Y., C. Xu, X. Hao, W. Li, Y. Chen, C. Zhu, and Z. Ye (2009): Fifty-year climate change and its effect on annual runoff in the Tarim River Basin, China, *Quaternary International*, *208*(1), 53-61.

Cheval, S., M. Baciu, and T. Breza (2003): An investigation into the precipitation conditions in Romania using a GIS-based method, *Theor. Appl. Climatol.*, *76*(1-2), 77-88.

Cole, S.J. and R.J. Moore (2008): Hydrological modelling using raingauge- and radar-based estimators of areal rainfall, *J. Hydrol.*, *358*(3-4), 159-181.

Corbari, C., G. Ravazzani, J. Martinelli, and M. Mancini (2009): Elevation based correction of snow coverage retrieved from satellite images to improve model calibration, *Hydrol. Earth Syst. Sci.*, *13*(5), 639-649.

Coron, L., V. Andreassian, C. Perrin, J. Lerat, J. Vaze, M. Bourqui, and F. Hendrickx (2012): Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments, *Water Resour. Res.*, *48*, W05552.

Dahlke, H.E., S.W. Lyon, J. Stedinger, G. Rosqvist, P. Jansson, and M. Weiler (2012): Contrasting trends in floods for two sub-arctic catchments in northern Sweden - does glacier presence matter?, *Hydrol. Earth Syst. Sci.*, *16*(7), 2123-2141.

Daly, C., R.P. Neilson, and D.L. Phillips (1994): A statistical topographic model for mapping climatological precipitation over mountainous terrain, *J. Appl. Meteorol.*, *33*(2), 140-158.

De Lannoy, G.J.M., R.H. Reichle, K.R. Arsenault, P.R. Houser, S. Kumar, N.E.C. Verhoest, and V.R.N. Pauwels (2012): Multiscale assimilation of Advanced Microwave Scanning Radiometer-EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado, *Water Resour. Res.*, *48*, W01522.

Deb, K., A. Pratap, S. Agarwal, and T. Meyarivan (2002): A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. Evol. Comput.*, *6*(2), 182-197.

Dee, D.P., S.M. Uppala, A.J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M.A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A.C.M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A.J. Geer, L. Haimberger, S.B. Healy, H. Hersbach, E.V. Holm, L. Isaksen, P. Kallberg, M. Kohler, M. Matricardi, A.P. McNally, B.M. Monge-Sanz, J.J. Morcrette, B.K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.N. Thepaut, and F. Vitart (2011): The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. Roy. Meteor. Soc.*, *137*(656), 553-597.

Delgado, J.M., B. Merz, and H. Apel (2014): Projecting flood hazard under climate change: an alternative approach to model chains, *Nat. Hazards Earth Syst. Sci.*, *14*(6), 1579-1589.

Dettinger, M.D. and D.R. Cayan (1995): Large-scale atmospheric forcing of recent trends toward early snowmelt runoff in California, *J. Clim.*, *8*(3), 606-623.

Dikikh, A. (1993): Lednikovyi stok rek Tyan-Shanya i ego rol'v formirovanii obshego stoka (Glacier runoff in the rivers of Tien Shan and its role in total runoff formation), *Materialy Glaciologicheskikh Issledovanii*, *77*, 41-50.

Donald, J.R., E.D. Soulis, N. Kouwen, and A. Pietroniro (1995): A land cover-based snow cover representation for distributed hydrological models, *Water Resour. Res.*, *31*(4), 995-1009.

Duethmann, D., J. Zimmer, A. Gafurov, A. Güntner, D. Kriegel, B. Merz, and S. Vorogushyn (2013): Evaluation of areal precipitation estimates based on downscaled reanalysis and station data by hydrological modelling, *Hydrol. Earth Syst. Sci.*, *17*(7), 2415-2434.

Duethmann, D., J. Peters, T. Blume, S. Vorogushyn, and A. Güntner (2014): The value of satellite-derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia, *Water Resour. Res.*, *50*(3), 2002-2021.

Dyurgerov, M.B. and M.F. Meier (2005): Glaciers and the changing earth system: a 2004 snapshot,

Institute of Arctic and Alpine Research, University of Colorado Boulder, CO, 117 pp.

Engelhardt, M., T.V. Schuler, and L.M. Andreassen (2014): Contribution of snow and glacier melt to discharge for highly glacierised catchments in Norway, *Hydrol. Earth Syst. Sci.*, *18*(2), 511-523.

Engeset, R.V., H.C. Udnaes, T. Guneriussen, H. Koren, E. Malnes, R. Solberg, and E. Alfnes (2003): Improving runoff simulations using satellite-observed time-series of snow covered area, *Nord. Hydrol.*, *34*(4), 281-294.

Fan, Y.T., Y.N. Chen, W.H. Li, H.J. Wang, and X.G. Li (2011): Impacts of temperature and precipitation on runoff in the Tarim River during the past 50 years, *Journal of Arid Land*, *3*(3), 220-230.

FAO (2013): Irrigation in Central Asia in figures: AQUASTAT survey - 2012, Food and Agriculture Organization of the United Nations, Rome, 228 pp.

FAO/IIASA/ISRIC/ISS-CAS/JRC (2012): Harmonized World Soil Database (version 1.2), edited by: FAO and IIASA, Rome, Italy and Laxenburg, Austria, available from http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/.

Farinotti, D., S. Usselmann, M. Huss, A. Bauder, and M. Funk (2012): Runoff evolution in the Swiss Alps: projections for selected high-alpine catchments based on ENSEMBLES scenarios, *Hydrol. Process.*, *26*(13), 1909-1924.

Feddes, R.A. and P.A.C. Raats (2004): Parameterizing the soil - water - plant root system, Kluwer Academic Publishers, http://edepot.wur.nl/35358, access: 17.09.2012.

Finger, D., F. Pellicciotti, M. Konz, S. Rimkus, and P. Burlando (2011): The value of glacier mass balance, satellite snow cover images, and hourly discharge for improving the performance of a physically based distributed hydrological model, *Water Resour. Res.*, *47*, W07519.

Foster, L.A., B.W. Brock, M.E.J. Cutler, and F. Diotri (2012): A physically based method for estimating supraglacial debris thickness from thermal band remote-sensing data, *J. Glaciol.*, *58*(210), 677-691.

Francke, T., A. Güntner, G. Mamede, E.N. Müller, and A. Bronstert (2008): Automated catena-based discretization of landscapes for the derivation of hydrological modelling units, *Int. J. Geogr. Inf. Sci.*, *22*(2), 111-132.

Franks, S.W., P. Gineste, K.J. Beven, and P. Merot (1998): On constraining the predictions of a distributed model: The incorporation of fuzzy estimates of saturated areas into the calibration process, *Water Resour. Res.*, *34*(4), 787-797.

Frei, C. and C. Schär (1998): A precipitation climatology of the Alps from high-resolution rain-gauge observations, *Int. J. Climatol.*, *18*(8), 873-900.

Frei, C., J.H. Christensen, M. Déqué, D. Jacob, R.G. Jones, and P.L. Vidale (2003): Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps, *Journal of Geophysical Research: Atmospheres (1984–2012)*, *108*(D3), 4124.

Frey, H., H. Machguth, M. Huss, C. Huggel, S. Bajracharya, T. Bolch, A. Kulkarni, A. Linsbauer, N. Salzmann, and M. Stoffel (2014): Estimating the volume of glaciers in the Himalayan–Karakoram region using different methods, *The Cryosphere*, *8*(6), 2313-2333.

Friedl, M.A., D.K. McIver, J.C.F. Hodges, X.Y. Zhang, D. Muchoney, A.H. Strahler, C.E. Woodcock, S. Gopal, A. Schneider, A. Cooper, A. Baccini, F. Gao, and C. Schaaf (2002): Global land cover mapping from MODIS: algorithms and early results, *Remote Sens. Environ.*, *83*(1-2), 287-302.

Gafurov, A., D. Kriegel, S. Vorogushyn, and B. Merz (2013): Evaluation of remotely sensed snow cover product in Central Asia, *Hydrology Research*, *44*(3), 506-522.

Gardelle, J., E. Berthier, and Y. Arnaud (2012): Slight mass gain of Karakoram glaciers in the early twenty-first century, *Nature Geoscience*, *5*(5), 322-325.

Garen, D.C., G.L. Johnson, and C.L. Hanson (1994): Mean areal precipitation for daily hydrologic modeling in mountainous regions, *Water Resour. Bull.*, *30*(3), 481-491.

Giese, E., D. Mamatkanov, and R. Wang (2005): Wasserressourcen und deren Nutzung im Flussbecken des Tarim (Autonome Region Xinjiang/VR China), Discussion Papers/Zentrum für internationale Entwicklungs- und Umweltforschung, 57 pp.

Glazirin, G.E. (2010): A century of investigtations on outbursts of the ice-dammed Lake Merzbacher (Central Tien Shan), *Austrian Journal of Earth Sciences*, *103*(2), 171-179.

Goodale, C.L., J.D. Aber, and S.V. Ollinger (1998): Mapping monthly precipitation, temperature, and solar radiation for Ireland with polynomial regression and a digital elevation model, *Climate. Res.*, *10*(1), 35-49.

Goovaerts, P. (2000): Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, *J. Hydrol.*, *228*(1-2), 113-129.

Gourley, J.J. and B.E. Vieux (2005): A method for evaluating the accuracy of quantitative precipitation estimates from a hydrologic modeling perspective, *J. Hydrometeorol.*, *6*(2), 115-133.

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

Güntner, A., S. Uhlenbrook, J. Seibert, and C. Leibundgut (1999): Multi-criterial validation of TOPMODEL in a mountainous catchment, *Hydrol. Process.*, *13*(11), 1603-1620.

Güntner, A. (2002): Large-scale hydrological modelling in the semi-arid North-East of Brazil, Potsdam Institute for Climate Impact Research, Potsdam, Germany, PIK-Report 77, 128 pp, http://www.pik-potsdam.de/research/publications/pikreports/.files/pr77.pdf.

Güntner, A. and A. Bronstert (2004): Representation of landscape variability and lateral redistribution processes for large-scale hydrological modelling in semi-arid areas, *J. Hydrol.*, *297*(1-4), 136-161.

Güntner, A. (2008): Improvement of Global Hydrological Models Using GRACE Data, *Surveys in Geophysics*, *29*(4-5), 375-397.

Haberlandt, U. and G.W. Kite (1998): Estimation of daily space-time precipitation series for macroscale hydrological modelling, *Hydrol. Process.*, *12*(9), 1419-1432.

Haberlandt, U., A. Schumann, and U. Büttner (2005): Räumliche Niederschlagsschätzung aus Punktmessungen und Radar am Beispiel des Elbehochwassers 2002, *Hydrologie und Wasserbewirtschaftung*, *49*(2), 56-68.

Hagg, W., M. Hoelzle, S. Wagner, E. Mayr, and Z. Klose (2013): Glacier and runoff changes in the Rukhk catchment, upper Amu-Darya basin until 2050, *Glob. Planet. Change*, *110*, 62-73.

Hall, D.K., G.A. Riggs, V.V. Salomonson, N.E. DiGirolamo, and K.J. Bayr (2002): MODIS snow-cover products, *Remote Sens. Environ.*, *83*(1-2), 181-194.

Hamlet, A.F., P.W. Mote, M.P. Clark, and D.P. Lettenmaier (2005): Effects of temperature and precipitation variability on snowpack trends in the western United States, *J. Clim.*, *18*(21), 4545–4561.

Hamlet, A.F., P.W. Mote, M.P. Clark, and D.P. Lettenmaier (2007): Twentieth-century trends in runoff, evapotranspiration, and soil moisture in the western United States, *J. Clim.*, *20*(8), 1468–1486.

Hancock, S., R. Baxter, J. Evans, and B. Huntley (2013): Evaluating global snow water equivalent products for testing land surface models, *Remote Sens. Environ.*, *128*, 107-117.

Hao, X., Y. Chen, and W. Li (2009): Impact of anthropogenic activities on the hydrologic characters of the mainstream of the Tarim River in Xinjiang during the past 50 years, *Environmental Geology*, *57*(2), 435-445.

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

Heistermann, M. and D. Kneis (2011): Benchmarking quantitative precipitation estimation by conceptual rainfall-runoff modeling, *Water Resour. Res.*, *47*, W06514.

Hevesi, J.A., J.D. Istok, and A.L. Flint (1992): Precipitation estimation in mountainous terrain using multivariate geostatistics: I Structural analysis, *J. Appl. Meteorol.*, *31*(7), 661-676.

Hidalgo, H.G., T. Das, M.D. Dettinger, D.R. Cayan, D.W. Pierce, T.P. Barnett, G. Bala, A. Mirin, A.W. Wood, C. Bonfils, B.D. Santer, and T. Nozawa (2009): Detection and attribution of streamflow timing changes to climate change in the western United States, *J. Clim.*, *22*(13), 3838-3855.

Hindshaw, R.S., E.T. Tipper, B.C. Reynolds, E. Lemarchand, J.G. Wiederhold, J. Magnusson, S.M. Bernasconi, R. Kretzschmar, and B. Bourdon (2011): Hydrological control of stream water chemistry in a glacial catchment (Damma Glacier, Switzerland), *Chemical Geology*, *285*(1-4), 215-230.

Hock, R. (1999): A distributed temperature-index ice- and snowmelt model including potential direct solar radiation, *J. Glaciol.*, *45*(149), 101-111.

Hock, R. (2003): Temperature index melt modelling in mountain areas, *J. Hydrol.*, *282*(1-4), 104-115.

Höppner, E. and N. Prechtel (2002): Snow cover mapping with NOAA-AVHRR images in the scope of an environmental GIS project for the Russian Altai (south Siberia), Proceedings of the Commission on Mountain Cartography of the International Cartographic Association (ICA), Mt. Hood, Oregon, May, 15-19, 2002, 26 pp.

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

Hurkmans, R., H. De Moel, J. Aerts, and P.A. Troch (2008): Water balance versus land surface model in the simulation of Rhine River discharges, *Water Resour. Res.*, *44*(1), W01418.

Huss, M., G. Jouvet, D. Farinotti, and A. Bauder (2010): Future high-mountain hydrology: a new parameterization of glacier retreat, *Hydrol. Earth Syst. Sci.*, *14*(5), 815-829.

Huss, M. (2012): Extrapolating glacier mass balance to the mountain-range scale: the European Alps 1900-2100, *Cryosphere*, *6*(4), 713-727.

Huss, M., R. Hock, A. Bauder, and M. Funk (2012): Conventional versus reference-surface mass balance, *J. Glaciol.*, *58*(208), 278-286.

Jarsjo, J., S.M. Asokan, C. Prieto, A. Bring, and G. Destouni (2012): Hydrological responses to climate change conditioned by historic alterations of land-use and water-use, *Hydrol. Earth Syst. Sci.*, *16*(5), 1335-1347.

Jarvis, A., H.I. Reuter, A. Nelson, and E. Guevara (2008): Holefilled seamless SRTM data, 4 ed., International Centre for Tropical Agriculture (CIAT), Cali, Colombia, available from http://srtm.csi.cgiar.org.

Jiang, Y., C.H. Zhou, and W.M. Cheng (2007): Streamflow trends and hydrological response to climatic change in Tarim headwater basin, *Journal of Geographical Sciences*, *17*(1), 51-61.

Jinshi, L. (1992): Jökulhlaups in the Kunmalike River, southern Tien Shan mountains, China., *Annals of Glaciology*, *16*, 85-88.

Johansson, B. and D.L. Chen (2003): The influence of wind and topography on precipitation distribution in Sweden: Statistical analysis and modelling, *Int. J. Climatol.*, *23*(12), 1523-1535.

Johansson, B. and D.L. Chen (2005): Estimation of areal precipitation for runoff modelling using wind data: a case study in Sweden, *Climate. Res.*, *29*(1), 53-61.

Jost, G., R.D. Moore, B. Menounos, and R. Wheate (2012): Quantifying the contribution of glacier runoff to streamflow in the upper Columbia River Basin, Canada, *Hydrol. Earth Syst. Sci.*, *16*(3), 849-860.

Juen, M., C. Mayer, A. Lambrecht, H. Han, and S. Liu (2014): Impact of varying debris cover thickness on ablation: a case study for Koxkar Glacier in the Tien Shan, *Cryosphere*, *8*(2), 377-386.

Juston, J., J. Seibert, and P.O. Johansson (2009): Temporal sampling strategies and uncertainty in calibrating a conceptual hydrological model for a small boreal catchment, *Hydrol. Process.*, *23*(21), 3093-3109.

Kääb, A., E. Berthier, C. Nuth, J. Gardelle, and Y. Arnaud (2012): Contrasting patterns of early twenty-first-century glacier mass change in the Himalayas, *Nature*, *488*(7412), 495-498.

Kavetski, D., S.W. Franks, and G. Kuczera (2002): Confronting input uncertainty in environmental modelling, in: Calibration of watershed models, edited by: Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., and Turcotte, R., American Geophysical Union, 49-68.

Kavetski, D., G. Kuczera, and S.W. Franks (2006): Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, *Water Resour. Res.*, *42*, W03408.

Kendall, M.G. (1975): Rank correlation methods, 4 ed., Charles Griffin, London, 196 pp.

Klaus, J. and J.J. McDonnell (2013): Hydrograph separation using stable isotopes: Review and evaluation, *J. Hydrol.*, *505*, 47-64.

Klemeš, V. (1986): Dilettantism in hydrology: transition or destiny?, *Water Resour. Res.*, *22*(9S), 177S-188S.

Koboltschnig, G.R., W. Schoner, M. Zappa, C. Kroisleitner, and H. Holzmann (2008): Runoff modelling of the glacierized Alpine Upper Salzach basin (Austria): multi-criteria result validation, *Hydrol. Process.*, *22*(19), 3950-3964.

Kollat, J.B. and P.M. Reed (2006): Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design, *Adv. Water Resour.*, *29*(6), 792-807.

Kollat, J.B., P.M. Reed, and T. Wagener (2012): When are multiobjective calibration trade-offs in hydrologic models meaningful?, *Water Resour. Res.*, *48*, W03520.

Konak, A., D.W. Coit, and A.E. Smith (2006): Multi-objective optimization using genetic algoithms: A tutorial, *Reliability engineering & system safety*, *91*, 992-1007.

Konz, M. and J. Seibert (2010): On the value of glacier mass balances for hydrological model calibration, *J. Hydrol.*, *385*(1-4), 238-246.

Kormann, C., T. Francke, M. Renner, and A. Bronstert (2014): Attribution of high resolution streamflow trends in Western Austria–an approach based on climate and discharge station data, *Hydrol. Earth Syst. Sci. Discuss.*, *11*(6), 6881-6922.

Körner, C. (1994): 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.

Kriegel, D., C. Mayer, W. Hagg, S. Vorogushyn, D. Duethmann, A. Gafurov, and D. Farinotti (2013): Changes in glacierisation, climate and runoff in the second half of the 20th century in the Naryn basin, Central Asia, *Glob. Planet. Change*, *110*, 51-61.

Krysanova, V., M. Wortmann, T. Bolch, B. Merz, D. Duethmann, J. Walter, S. Huang, T. Jiang, B. Su, and Z. Kundzewicz (2014): Analysis of current trends in climate parameters, river discharge and glaciers in the Aksu River basin (Central Asia), *Hydrological Sciences Journal*.

Kundzewicz, Z., B. Merz, S. Vorogushyn, H. Hartmann, D. Duethmann, M. Wortmann, S. Huang, B. Su, T. Jiang, and V. Krysanova (2015): Analysis of changes in climate and river discharge with focus

on seasonal runoff predictability in the Aksu River Basin, *Environmental Earth Sciences*, *73*(2), 501-516.

Kutuzov, S. and M. Shahgedanova (2009): Glacier retreat and climatic variability in the eastern Terskey-Alatoo, inner Tien Shan between the middle of the 19th century and beginning of the 21st century, *Glob. Planet. Change*, *69*(1-2), 59-70.

Laguardia, G. and S. Niemeyer (2008): On the comparison between the LISFLOOD modelled and the ERS/SCAT derived soil moisture estimates, *Hydrol. Earth Syst. Sci.*, *12*(6), 1339-1351.

Lamb, R., K. Beven, and S. Myraboe (1998): Use of spatially distributed water table observations to constrain uncertainty in a rainfall-runoff model, *Adv. Water Resour.*, *22*(4), 305-318.

Legates, D.R. and C.J. Willmott (1990): Mean seasonal and spatial variability in gauge-corrected, global precipitation, *Int. J. Climatol.*, *10*(2), 111-127.

Li, X.G. and M.W. Williams (2008): Snowmelt runoff modelling in an arid mountain watershed, Tarim Basin, China, *Hydrol. Process.*, *22*(19), 3931-3940.

Linsbauer, A., F. Paul, and W. Haeberli (2012): Modeling glacier thickness distribution and bed topography over entire mountain ranges with GlabTop: Application of a fast and robust approach, *Journal of Geophysical Research-Earth Surface*, *117*, F03007.

Liston, G.E. (2004): Representing subgrid snow cover heterogeneities in regional and global models, *J. Clim.*, *17*(6), 1381-1397.

Liu, S., Y. Ding, D. Shangguan, Y. Zhang, J. Li, H. Han, J. Wang, and C. Xie (2006): Glacier retreat as a result of climate warming and increased precipitation in the Tarim river basin, northwest China, *Annals of Glaciology*, *43*(1), 91-96.

Liu, Y.Q., C.D. Peters-Lidard, S. Kumar, J.L. Foster, M. Shaw, Y.D. Tian, and G.M. Fall (2013): Assimilating satellite-based snow depth and snow cover products for improving snow predictions in Alaska, *Adv. Water Resour.*, *54*, 208-227.

Liu, Z.F., Z.X. Xu, J.X. Huang, S.P. Charles, and G.B. Fu (2010): Impacts of climate change on hydrological processes in the headwater catchment of the Tarim River basin, China, *Hydrol. Process.*, *24*(2), 196-208.

Lloyd, C.D. (2005): Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain, *J. Hydrol.*, *308*(1-4), 128-150.

Lo, M.H., J.S. Famiglietti, P.J.F. Yeh, and T.H. Syed (2010): Improving parameter estimation and water table depth simulation in a land surface model using GRACE water storage and estimated base flow data, *Water Resour. Res.*, *46*, W05517.

Luce, C.H. and D.G. Tarboton (2004): The application of depletion curves for parameterization of subgrid variability of snow, *Hydrol. Process.*, *18*(8), 1409-1422.

Lutz, A.F., W.W. Immerzeel, A. Gobiet, F. Pellicciotti, and M.F.P. Bierkens (2013): Comparison of climate change signals in CMIP3 and CMIP5 multi-model ensembles and implications for Central Asian glaciers, *Hydrol. Earth Syst. Sci.*, *17*(9), 3661-3677.

Ly, S., C. Charles, and A. Degre (2011): 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*(7), 2259-2274.

Madsen, H. (2000): Automatic calibration of a conceptual rainfall–runoff model using multiple objectives, *J. Hydrol.*, *235*(3-4), 276-288.

Mamatkanov, D.M., L.V. Bazhanova, and V.V. Romanovskiy (2006): Vodnye resursy Kyrgyzstana na sovremennom etape (Water resources of Kyrgyzstan in the recent period), Bishkek (in Russian).

Mann, H. (1945): Non-parametric test against trend, *Econometrica*, *13*, 245-259.

Marquinez, J., J. Lastra, and P. Garcia (2003): Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis, *J. Hydrol.*, *270*(1-2), 1-11.

Martinez-Cob, A. (1996): Multivariate geostatistical analysis of evapotranspiration and precipitation in mountainous terrain, *J. Hydrol.*, *174*(1-2), 19-35.

Maurer, E.P., J.D. Rhoads, R.O. Dubayah, and D.P. Lettenmaier (2003): Evaluation of the snow-covered area data product from MODIS, *Hydrol. Process.*, *17*(1), 59-71.

Maussion, F., D. Scherer, T. Mölg, E. Collier, J. Curio, and R. Finkelnburg (2014): Precipitation seasonality and variability over the Tibetan Plateau as resolved by the High Asia Reanalysis, *J. Clim.*, *27*(5), 1910-1927.

Mayr, E., W. Hagg, C. Mayer, and L. Braun (2013): Calibrating a spatially distributed conceptual hydrological model using runoff, annual mass balance and winter mass balance, *J. Hydrol.*, *478*, 40-49.

Merz, B., S. Vorogushyn, S. Uhlemann, J. Delgado, and Y. Hundecha (2012): "More efforts and scientific rigour are needed to attribute trends in flood time series.", *Hydrol. Earth Syst. Sci.*, *16*(5), 1379-1387.

Merz, R., J. Parajka, and G. Blöschl (2011): Time stability of catchment model parameters: Implications for climate impact analyses, *Water Resour. Res.*, *47*, W02531.

Micklin, P. (2007): The Aral Sea disaster, *Annu. Rev. Earth Planet. Sci.*, *35*, 47-72.

Mitchell, T.D. and P.D. Jones (2005): An improved method of constructing a database of monthly climate observations and associated high-resolution grids, *Int. J. Climatol.*, *25*(6), 693-712.

Mizukami, N. and M.B. Smith (2012): Analysis of inconsistencies in multi-year gridded quantitative precipitation estimate over complex terrain and its impact on hydrologic modeling, *J. Hydrol.*, *428*, 129-141.

Molnar, P., P. Burlando, and F. Pellicciotti (2011): Streamflow trends in mountainous regions, *Encyclopedia of Snow, Ice and Glaciers*, 1084-1089.

Monteith, J.L. (1965): Evaporation and environment, *Symo. Soc. Exp. Biology*, *19*, 205-234.

Moore, R. and M. Demuth (2001): Mass balance and streamflow variability at Place Glacier, Canada, in relation to recent climate fluctuations, *Hydrol. Process.*, *15*(18), 3473-3486.

Mroczkowski, M., G.P. Raper, and G. Kuczera (1997): The quest for more powerful validation of conceptual catchment models, *Water Resour. Res.*, *33*(10), 2325-2335.

Mueller, E., A. Güntner, T. Francke, and G. Mamede (2010): Modelling sediment export, retention and reservoir sedimentation in drylands with the WASA-SED model, *Geoscientific Model Development*, *3*(1), 275-291.

Müller-Lemans, H., M. Aellen, L. Braun, G. Kappenberger, and U. Steinegger (1997): Niederschlagsverteilung im Tödigebiet: Messungen und Überprüfung mit der Wasserhaushaltsgleichung, *Beitr. Hydrol. Schweiz*, *36*, 7-43.

Myneni, R.B., S. Hoffman, Y. Knyazikhin, J.L. Privette, J. Glassy, Y. Tian, Y. Wang, X. Song, Y. Zhang, G.R. Smith, A. Lotsch, M. Friedl, J.T. Morisette, P. Votava, R.R. Nemani, and S.W. Running (2002): Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data, *Remote Sens. Environ.*, *83*(1-2), 214-231.

National Climate Data Center (1998): http://www.ncdc.noaa.gov/oa/pod-guide/ncdc/docs/podug/index.htm, access: 29.01.2013.

NCDC (2005): Daily and sub-daily precipitation for the former USSR (DSI-9813) 1.0 ed., edited by: National Climatic Data Center NESDIS NOAA U.S. Department of Commerce, available from http://www.ncdc.noaa.gov

Ninyerola, M., X. Pons, and J.M. Roure (2000): A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques, *Int. J. Climatol.*, *20*(14), 1823-1841.

Ninyerola, M., X. Pons, and J.M. Roure (2007): Monthly precipitation mapping of the Iberian Peninsula using spatial interpolation tools implemented in a Geographic Information System, *Theor. Appl. Climatol.*, *89*(3-4), 195-209.

Nolin, A.W., J. Phillippe, A. Jefferson, and S.L. Lewis (2010): Present-day and future contributions of glacier runoff to summertime flows in a Pacific Northwest watershed: Implications for water resources, *Water Resour. Res.*, *46*, W12509.

Osmonov, A., T. Bolch, C. Xi, A. Kurban, and W. Guo (2013): Glacier characteristics and changes in the Sary-Jaz River Basin (Central Tien Shan, Kyrgyzstan)–1990–2010, *Remote Sensing Letters*, *4*(8), 725-734.

Östrem, G. (1959): Ice melting under a thin layer of moraine, and the existence of ice cores in moraine ridges, *Geografiska Annaler*, 228-230.

Parajka, J. and G. Blöschl (2006): Validation of MODIS snow cover images over Austria, *Hydrol. Earth Syst. Sci.*, *10*(5), 679-689.

Parajka, J., R. Merz, and G. Blöschl (2007): Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments, *Hydrol. Process.*, *21*(4), 435-446.

Parajka, J. and G. Blöschl (2008): The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models, *J. Hydrol.*, *358*(3-4), 240-258.

Parajka, J., V. Naeimi, G. Blöschl, and J. Komma (2009): Matching ERS scatterometer based soil moisture patterns with simulations of a conceptual dual layer hydrologic model over Austria, *Hydrol. Earth Syst. Sci.*, *13*(2), 259-271.

Pellicciotti, F., B. Brock, U. Strasser, P. Burlando, M. Funk, and J. Corripio (2005): An enhanced temperature-index glacier melt model including the shortwave radiation balance: development and testing for Haut Glacier d'Arolla, Switzerland, *J. Glaciol.*, *51*(175), 573-587.

Pellicciotti, F., A. Bauder, and M. Parola (2010): Effect of glaciers on streamflow trends in the Swiss Alps, *Water Resour. Res.*, *46*(10), W10522.

Pellicciotti, F., C. Buergi, W.W. Immerzeel, M. Konz, and A.B. Shrestha (2012): Challenges and uncertainties in hydrological modeling of remote

Hindu Kush-Karakoram-Himalayan (HKH) basins: Suggestions for calibration strategies, *Mt. Res. Dev.*, *32*(1), 39-50.

Perrin, C., L. Oudin, V. Andreassian, C. Rojas-Serna, C. Michel, and T. Mathevet (2007): Impact of limited streamflow data on the efficiency and the parameters of rainfall-runoff models, *Hydrological Sciences Journal*, *52*(1), 131-151.

Perry, M. and D. Hollis (2005): The generation of monthly gridded datasets for a range of climatic variables over the UK, *Int. J. Climatol.*, *25*(8), 1041-1054.

Phillips, D.L., J. Dolph, and D. Marks (1992): A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain, *Agr. Forest Meteorol.*, *58*(1-2), 119-141.

Pieczonka, T., T. Bolch, W. Junfeng, and L. Shiyin (2013): Heterogeneous mass loss of glaciers in the Aksu-Tarim Catchment (Central Tien Shan) revealed by 1976 KH-9 Hexagon and 2009 SPOT-5 stereo imagery, *Remote Sens. Environ.*, *130*, 233-244.

Pieczonka, T. and T. Bolch (in revision): Region-wide glacier mass budgets and area changes for the Central Tien Shan since the 1970s using declassified stereoscopic imagery, *Glob. Planet. Change*.

Polar Meterology Group/Byrd Polar Research Center, T.O.S.U. (2012): Arctic Systems Reanalysis Model Output Data, edited by: Research Data Archive at the National Center for Atmospheric Research - Computational and Information Systems Laboratory, available from http://rda.ucar.edu/datasets/ds631.4/.

Prudhomme, C. and D.W. Reed (1998): Relationships between extreme daily precipitation and topography in a mountainous region: A case study in Scotland, *Int. J. Climatol.*, *18*(13), 1439-1453.

Raleigh, M.S., K. Rittger, C.E. Moore, B. Henn, J.A. Lutz, and J.D. Lundquist (2013): Ground-based testing of MODIS fractional snow cover in subalpine meadows and forests of the Sierra Nevada, *Remote Sens. Environ.*, *128*, 44-57.

Rao, C.R.N. and J. Chen (1995): Inter-satellite calibration linkages for the visible and near-infared channels of the Advanced Very High Resolution Radiometer on the NOAA-7, -9, and -11 spacecraft, *International Journal of Remote Sensing*, *16*(11), 1931–1942.

Refsgaard, J.C. (1997): Parameterisation, calibration and validation of distributed hydrological models, *J. Hydrol.*, *198*, 69-97.

Reichert, P. and J. Mieleitner (2009): Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters, *Water Resour. Res.*, *45*, W10402.

Reid, T.D., M. Carenzo, F. Pellicciotti, and B.W. Brock (2012): Including debris cover effects in a distributed model of glacier ablation, *J. Geophys. Res.-Atmos.*, *117*, D18105.

Rodell, M. and P.R. Houser (2004): Updating a land surface model with MODIS-derived snow cover, *J. Hydrometeorol.*, *5*(6), 1064-1075.

Rüdiger, C., J.C. Calvet, C. Gruhier, T.R.H. Holmes, R.A.M. de Jeu, and W. Wagner (2009): An intercomparison of ERS-Scat and AMSR-E soil moisture observations with model simulations over France, *J. Hydrometeorol.*, *10*(2), 431-447.

Schaefli, B., B. Hingray, M. Niggli, and A. Musy (2005): A conceptual glacio-hydrological model for high mountainous catchments, *Hydrol. Earth Syst. Sci.*, *9*(1-2), 95-109.

Schaefli, B. and H.V. Gupta (2007): Do Nash values have value?, *Hydrol. Process.*, *21*(15), 2075-2080.

Schaefli, B. and M. Huss (2011): Integrating point glacier mass balance observations into hydrologic model identification, *Hydrol. Earth Syst. Sci.*, *15*(4), 1227-1241.

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

Schiemann, R., D. Luthi, P.L. Vidale, and C. Schär (2008): The precipitation climate of Central Asia - intercomparison of observational and numerical data sources in a remote semiarid region, *Int. J. Climatol.*, *28*(3), 295-314.

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

Schöber, J., K. Schneider, K. Helfricht, P. Schattan, S. Achleitner, F. Schöberl, and R. Kirnbauer (2014): Snow cover characteristics in a glacierized catchment in the Tyrolean Alps-Improved spatially distributed modelling by usage of Lidar data, *J. Hydrol.*, *519*, 3492–3510.

Seibert, J. (2000): Multi-criteria calibration of a conceptual runoff model using a genetic algorithm, *Hydrol. Earth Syst. Sci.*, *4*(2), 215-224.

Seibert, J. and J.J. McDonnell (2002): On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resour. Res.*, *38*(11), 1241.

Seiller, G., F. Anctil, and C. Perrin (2012): Multimodel evaluation of twenty lumped hydrological models under contrasted climate

conditions, *Hydrol. Earth Syst. Sci.*, *16*(4), 1171-1189.

Sen, P.K. (1968): Estimates of the regression coefficient based on Kendall's tau, *Journal of the American Statistical Association*, *63*(324), 1379-1389.

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

Shrestha, M., L. Wang, T. Koike, H. Tsutsui, Y. Xue, and Y. Hirabayashi (2013): Correcting basin-scale snowfall in a mountainous basin using a distributed snowmelt model and remote sensing data, *Hydrology and Earth System Sciences Discussions*, *10*(9), 11711-11753.

Shuttleworth, W.J. and J.S. Wallace (1985): Evaporation from sparse crops - an energy combination theory, *Q. J. Roy. Meteor. Soc.*, *111*(469), 839-855.

Siegfried, T. and T. Bernauer (2007): 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*(11), W11406.

Siegfried, T., T. Bernauer, R. Guiennet, S. Sellars, A.W. Robertson, J. Mankin, P. Bauer-Gottwein, and A. Yakovlev (2012): Will climate change exacerbate water stress in Central Asia?, *Clim. Change*, *112*(3-4), 881-899.

Skamarock, W.C., J.B. Klemp, J. Dudhia, D.O. Gill, D.M. Barker, M.G. Duda, X. Huang, W. Wang, and J.G. Powers (2008): A description of the advanced research WRF Version 3, technical report NCAR/TN–475+STR, 124 pp.

Sorg, A., T. Bolch, M. Stoffel, O. Solomina, and M. Beniston (2012): Climate change impacts on glaciers and runoff in Tien Shan (Central Asia), *Nature Climate Change*, *2*(10), 725-731.

Sorg, A., M. Huss, M. Rohrer, and M. Stoffel (2014): The days of plenty might soon be over in glacierized Central Asian catchments, *Environ. Res. Lett.*, *9*(10), 104018.

Sorman, A.A., A. Sensoy, A.E. Tekeli, A.U. Sorman, and Z. Akyurek (2009): Modelling and forecasting snowmelt runoff process using the HBV model in the eastern part of Turkey, *Hydrol. Process.*, *23*(7), 1031-1040.

Stahl, K. and R.D. Moore (2006): Influence of watershed glacier coverage on summer streamflow in British Columbia, Canada, *Water Resour. Res.*, *42*(6), W06201.

Stahl, K., R.D. Moore, J.M. Shea, D. Hutchinson, and A.J. Cannon (2008): Coupled modelling of glacier and streamflow response to future climate scenarios, *Water Resour. Res.*, *44*(2), W02422.

Stisen, S. and I. Sandholt (2010): Evaluation of remote-sensing-based rainfall products through predictive capability in hydrological runoff modelling, *Hydrol. Process.*, *24*(7), 879-891.

Suklitsch, M., A. Gobiet, H. Truhetz, N.K. Awan, H. Gottel, and D. Jacob (2011): Error characteristics of high resolution regional climate models over the Alpine area, *Climate Dynamics*, *37*(1-2), 377-390.

Sun, R.H., B.P. Zhang, and J. Tan (2008): A multivariate regression model for predicting precipitation in the Daqing Mountains, *Mt. Res. Dev.*, *28*(3-4), 318-325.

Surazakov, A.B. and V.B. Aizen (2006): Estimating volume change of mountain glaciers using SRTM and map-based topographic data, *Geoscience and Remote Sensing, IEEE Transactions on*, *44*(10), 2991-2995.

Tang, Q.H., H.P. Hu, T.K. Oki, and F.Q. Tian (2007a): Water balance within intensively cultivated alluvial plain in an arid environment, *Water Resour. Manag.*, *21*(10), 1703-1715.

Tang, Y., P. Reed, and T. Wagener (2006): How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration?, *Hydrol. Earth Syst. Sci.*, *10*(2), 289-307.

Tang, Y., P.M. Reed, and J.B. Kollat (2007b): Parallelization strategies for rapid and robust evolutionary multiobjective optimization in water resources applications, *Adv. Water Resour.*, *30*(3), 335-353.

Tao, H., M. Gemmer, Y.G. Bai, B.D. Su, and W.Y. Mao (2011): Trends of streamflow in the Tarim River Basin during the past 50 years: Human impact or climate change?, *J. Hydrol.*, *400*(1-2), 1-9.

Tobin, C., L. Nicotina, M.B. Parlange, A. Berne, and A. Rinaldo (2011): Improved interpolation of meteorological forcings for hydrologic applications in a Swiss Alpine region, *J. Hydrol.*, *401*(1-2), 77-89.

Tolson, B.A. and C.A. Shoemaker (2008): Efficient prediction uncertainty approximation in the calibration of environmental simulation models, *Water Resour. Res.*, *44*(4), W04411.

Udnaes, H.C., E. Alfnes, and L.M. Andreassen (2007): Improving runoff modelling using satellite-derived snow covered area?, *Nord. Hydrol.*, *38*(1), 21-32.

Unger-Shayesteh, K., S. Vorogushyn, D. Farinotti, A. Gafurov, D. Duethmann, A. Mandychev, and B. Merz (2013): What do we know about past changes

in the water cycle of Central Asian headwaters? A review, *Glob. Planet. Change*, *110*, 4-25.

Uppala, S.M., P.W. Kallberg, A.J. Simmons, U. Andrae, V.D. Bechtold, M. Fiorino, J.K. Gibson, J. Haseler, A. Hernandez, G.A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R.P. Allan, E. Andersson, K. Arpe, M.A. Balmaseda, A.C.M. Beljaars, L. Van De Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Holm, B.J. Hoskins, L. Isaksen, P. Janssen, R. Jenne, A.P. McNally, J.F. Mahfouf, J.J. Morcrette, N.A. Rayner, R.W. Saunders, P. Simon, A. Sterl, K.E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo, and J. Woollen (2005): The ERA-40 re-analysis, *Q. J. Roy. Meteor. Soc.*, *131*(612), 2961-3012.

van der Linden, P. and J.F.B. Mitchell (2009): ENSEMBLES: climate change and its impacts: summary of research and results from the ENSEMBLES project, 160 pp.

Vaze, J., D.A. Post, F.H.S. Chiew, J.M. Perraud, N.R. Viney, and J. Teng (2010): Climate non-stationarity - Validity of calibrated rainfall-runoff models for use in climate change studies, *J. Hydrol.*, *394*(3-4), 447-457.

Vieux, B.E. and P.B. Bedient (2004): Assessing urban hydrologic prediction accuracy through event reconstruction, *J. Hydrol.*, *299*(3-4), 217-236.

Villarini, G. and A. Strong (2014): Roles of climate and agricultural practices in discharge changes in an agricultural watershed in Iowa, *Agriculture Ecosystems & Environment*, *188*, 204-211.

Viviroli, D., H.H. Durr, B. Messerli, M. Meybeck, and R. Weingartner (2007): Mountains of the world, water towers for humanity: Typology, mapping, and global significance, *Water Resour. Res.*, *43*(7), W07447.

Viviroli, D., D.R. Archer, W. Buytaert, H.J. Fowler, G.B. Greenwood, A.F. Hamlet, Y. Huang, G. Koboltschnig, M.I. Litaor, J.I. Lopez-Moreno, S. Lorentz, B. Schadler, H. Schreier, K. Schwaiger, M. Vuille, and R. Woods (2011): Climate change and mountain water resources: overview and recommendations for research, management and policy, *Hydrol. Earth Syst. Sci.*, *15*(2), 471-504.

Voigt, S., M. Koch, and M.F. Baumgartner (1999): A multichannel threshold technique for NOAA AVHRR data to monitor the extent of snow cover in the Swiss Alps, Interactions between the cryosphere, climate and greenhouse gases, Proceedings of IUGG 99 Symposium HS2, Birmingham, UK, 35-43.

Vrugt, J.A. and B.A. Robinson (2007): Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging, *Water Resour. Res.*, *43*(1), W01411.

Wagner, W., G. Lemoine, and H. Rott (1999): A method for estimating soil moisture from ERS scatterometer and soil data, *Remote Sens. Environ.*, *70*(2), 191-207.

Wagner, W., G. Blöschl, P. Pampaloni, J.C. Calvet, B. Bizzarri, J.P. Wigneron, and Y. Kerr (2007): Operational readiness of microwave remote sensing of soil moisture for hydrologic applications, *Nord. Hydrol.*, *38*(1), 1-20.

Wahr, J., S. Swenson, V. Zlotnicki, and I. Velicogna (2004): Time-variable gravity from GRACE: First results, *Geophys. Res. Lett.*, *31*(11), L11501.

Wang, X.W., H.J. Xie, T.G. Liang, and X.D. Huang (2009): Comparison and validation of MODIS standard and new combination of Terra and Aqua snow cover products in northern Xinjiang, China, *Hydrol. Process.*, *23*(3), 419-429.

Wang, Y. (2006): Local records of the Akesu River basin, Fangzi Publisher, Beijing, 719 pp.

Weedon, G.P., S. Gomes, P. Viterbo, W.J. Shuttleworth, E. Blyth, H. Osterle, J.C. Adam, N. Bellouin, O. Boucher, and M. Best (2011): Creation of the WATCH Forcing Data and its use to assess global and regional reference crop evaporation over land during the twentieth century, *J. Hydrometeorol.*, *12*(5), 823-848.

Weedon, G.P., G. Balsamo, N. Bellouin, S. Gomes, M.J. Best, and P. Viterbo (2014): The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, *Water Resour. Res.*, *50*(9), 7505-7514.

Weingartner, R., M. Barben, and M. Spreafico (2003): Floods in mountain areas - an overview based on examples from Switzerland, *J. Hydrol.*, *282*(1-4), 10-24.

Werth, S. and A. Güntner (2010): Calibration analysis for water storage variability of the global hydrological model WGHM, *Hydrol. Earth Syst. Sci.*, *14*(1), 59-78.

Westra, S., M. Thyer, M. Leonard, D. Kavetski, and M. Lambert (2014): A strategy for diagnosing and interpreting hydrological model nonstationarity, *Water Resour. Res.*, *50*(6), 5090-5113.

WGMS (2012): Fluctuations of Glaciers 2005–2010, edited by: Zemp, M., Frey, H., Gärtner-Roer, I., Nussbaumer, S., Hoelzle, M., Paul, F., and Haeberli, W., ICSU(WDS)/IUGG(IACS)/UNEP/UNESCO/ WMO, World Glacier Monitoring Service, Zurich, Switzerland, 366 pp.

Winsemius, H.C., H.H.G. Savenije, and W.G.M. Bastiaanssen (2008): Constraining model parameters on remotely sensed evaporation:

justification for distribution in ungauged basins?, *Hydrol. Earth Syst. Sci.*, *12*(6), 1403-1413.

Wortmann, M., V. Krysanova, Z. Kundzewicz, B. Su, and X. Li (2013): Assessing the influence of the Merzbacher Lake outburst floods on discharge using the hydrological model SWIM in the Aksu headwaters, Kyrgyzstan/NW China, *Hydrol. Process.*, *28*(26), 6337–6350.

Xu, H., B. Zhou, and Y. Song (2010): Impacts of climate change on headstream runoff in the Tarim River Basin, *Hydrology Research*, *42*(1), 20-29.

Xu, J.H., Y.N. Chen, W.H. Li, Y. Yang, and Y.L. Hong (2011): An integrated statistical approach to identify the nonlinear trend of runoff in the Hotan River and its relation with climatic factors, *Stochastic Environmental Research and Risk Assessment*, *25*(2), 223-233.

Yang, D.Q., B.E. Goodison, J.R. Metcalfe, V.S. Golubev, E. Elomaa, T. Gunther, R. Bates, T. Pangburn, C.L. Hanson, D. Emerson, V. Copaciu, and J. Miklovic (1995): Accuracy of Tretyakov precipitation gauge: Result of WMO intercomparison, *Hydrol. Process.*, *9*(8), 877-895.

Yatagai, A., K. Kamiguchi, O. Arakawa, A. Hamada, N. Yasutom, and A. Kitoh (2012): 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.

Yatheendradas, S., C.D.P. Lidard, V. Koren, B.A. Cosgrove, L.G.G. De Goncalves, M. Smith, J. Geiger, Z.T. Cui, J. Borak, S.V. Kumar, D.L. Toll, G. Riggs, and N. Mizukami (2012): Distributed assimilation of satellite-based snow extent for improving simulated streamflow in mountainous, dense forests: An example over the DMIP2 western basins, *Water Resour. Res.*, *48*(9), W09557.

Yilmaz, K.K., T.S. Hogue, K.L. Hsu, S. Sorooshian, H. Gupta, and T. Wagener (2005): Intercomparison of rain gauge, radar, and satellite-based precipitation estimates with emphasis on hydrologic forecasting, *J. Hydrometeorol.*, *6*(4), 497-517.

Yue, S., P. Pilon, B. Phinney, and G. Cavadias (2002): The influence of autocorrelation on the ability to detect trend in hydrological series, *Hydrol. Process.*, *16*(9), 1807-1829.

Yue, S., P. Pilon, and B. Phinney (2003): Canadian streamflow trend detection: impacts of serial and cross-correlation, *Hydrological Sciences Journal*, *48*(1), 51-63.

Zaitchik, B.F. and M. Rodell (2009): Forward-looking assimilation of MODIS-derived snow-covered area into a land surface model, *J. Hydrometeorol.*, *10*(1), 130-148.

Zhang, Q., C.-Y. Xu, H. Tao, T. Jiang, and Y. Chen (2010): Climate changes and their impacts on water resources in the arid regions: a case study of the Tarim River Basin, China, *Stochastic Environmental Research and Risk Assessment*, *24*(3), 349-358.

Zhao, H.X. and R. Fernandes (2009): Daily snow cover estimation from Advanced Very High Resolution Radiometer Polar Pathfinder data over Northern Hemisphere land surfaces during 1982-2004, *J. Geophys. Res.-Atmos.*, *114*(D5), D05113.

Zhao, Q., B. Ye, Y. Ding, S. Zhang, S. Yi, J. Wang, D. Shangguan, C. Zhao, and H. Han (2013): Coupling a glacier melt model to the Variable Infiltration Capacity (VIC) model for hydrological modeling in north-western China, *Environmental Earth Sciences*, *68*(1), 87-101.

Zhou, H., X. Zhang, H. Xu, H. Ling, and P. Yu (2012): Influences of climate change and human activities on Tarim River runoffs in China over the past half century, *Environmental Earth Sciences*, *67*(1), 231-241.

Acknowledgements

My first thanks go to my advisor Andreas Güntner for many important discussions, helpful feedback, and his support throughout this work.

For taking the time to read and review my thesis, I am grateful to Bruno Merz, Axel Bronstert, and Jan Seibert.

I want to thank Bruno Merz for valuable discussions and for giving me the possibility to pursue my PhD thesis at GFZ. Sergiy Vorogushyn deserves a great thank you for discussions on various topics concerning hydrology in Central Asia, for his support and engagement in the SuMaRiO and CAWa projects, and for always giving me the feeling that my questions are welcome.

Many thanks go to Abror Gafurov, Daniel Farinotti, David Kriegel, Katy Unger-Sayesteh, and Sergiy Vorogushyn for our good collaboration and discussions within the "CAWa group" at GFZ, which were important for this work. In the beginning of this work, Till Franke was a big help for any details related to the WASA model, and Bettina Schaefli supported me with encouraging discussions and email advice.

This study is part of the research projects CAWa (Regional Research Network "Water in Central Asia") and SuMaRiO (Sustainable Management of River Oases along the Tarim River), funded by the Federal Foreign Office and the Federal Ministry of Education and Research, respectively. Many thanks go to all partners of the projects in Germany, Switzerland, Kyrgyzstan, Uzbekistan, and China for the cooperation, and for sharing data, without which this study would not have been feasible.

A special thank you goes to my colleagues and former colleagues at the hydrology section at GFZ, for a very enjoyable working atmosphere and for their help with numerous problems during this time.

Finally, I want to thank my family for their support, and Holger for his understanding and encouragement.

Author's declaration

I prepared this dissertation without illegal assistance. The work is original except where indicated by special reference in the text and no part of the dissertation has been submitted for any other degree.

This dissertation has not been presented to any other University for examination, neither in Germany nor in another country.

Doris Düthmann

Potsdam, January 2015