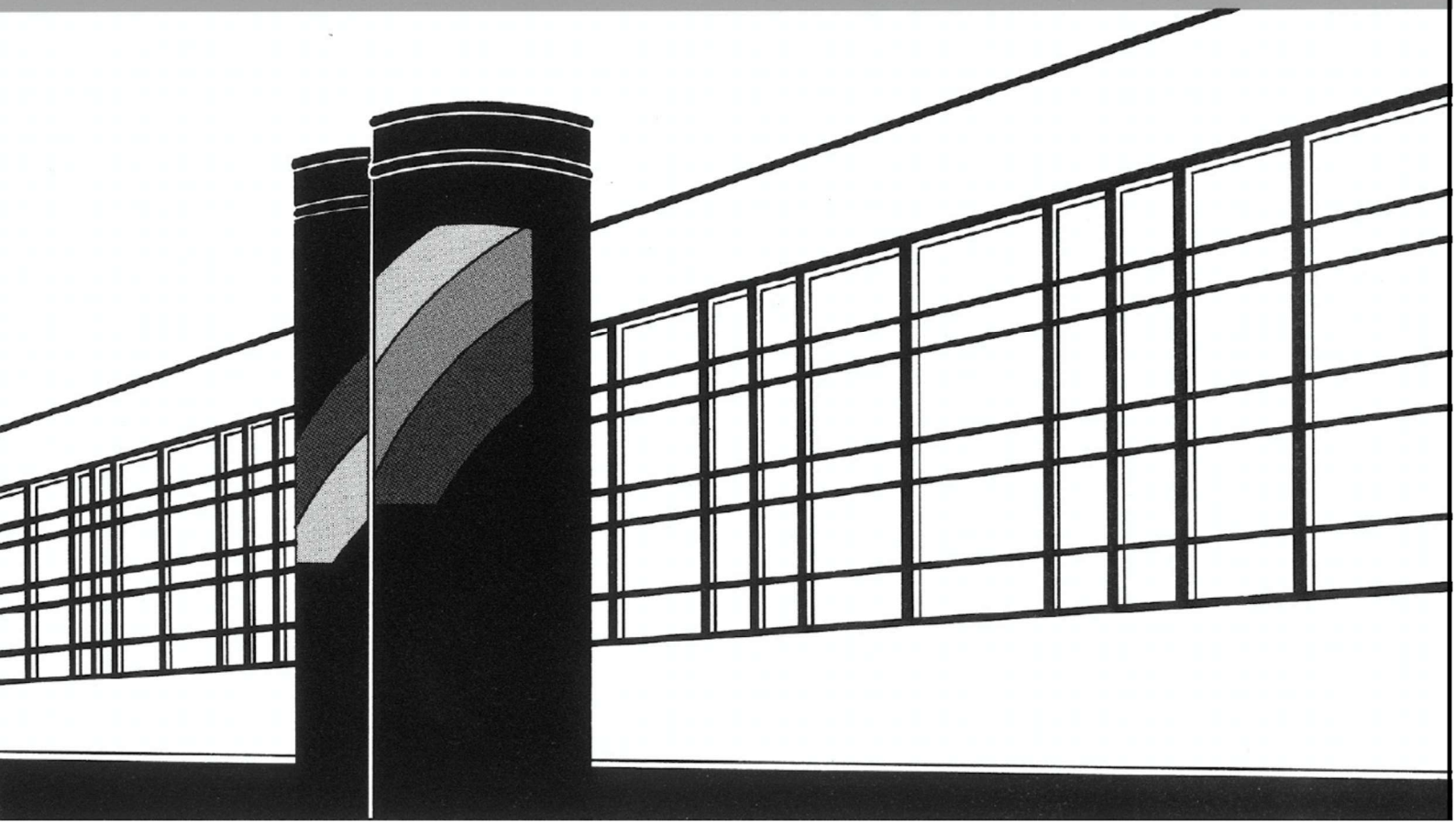
Universität Stuttgart

Institut für Wasser- und Umweltsystemmodellierung

Mitteilungen



Heft 296

Felix Herma

Data Processing and Model Choice for Flood Prediction

Data Processing and Model Choice for Flood Prediction

von der Fakultät Bau- und Umweltingenieurwissenschaften der
Universität Stuttgart zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.) genehmigte Abhandlung

vorgelegt von
Felix Herma
aus Freiburg im Breisgau, Deutschland

Hauptberichter:	Prof. Dr. rer. nat. Dr.-Ing. András Bárdossy
Mitberichter:	Prof. Dr.-Ing. Markus Disse

Tag der mündlichen Prüfung: 08. August 2022

Institut für Wasser- und Umweltsystemmodellierung
der Universität Stuttgart
2022

Heft 296 **Data Processing and Model Choice for Flood Prediction**

von
Dr.-Ing.
Felix Herma

Eigenverlag des Instituts für Wasser- und Umweltsystemmodellierung
der Universität Stuttgart

D93 Data Processing and Model Choice for Flood Prediction

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://www.d-nb.de abrufbar

Herma, Felix:
Data Processing and Model Choice for Flood Prediction, Universität Stuttgart. - Stuttgart: Institut für Wasser- und Umweltsystemmodellierung, 2022

(Mitteilungen Institut für Wasser- und Umweltsystemmodellierung, Universität Stuttgart: H. 296)
Zugl.: Stuttgart, Univ., Diss., 2022
ISBN 978-3-910293-00-7

NE: Institut für Wasser- und Umweltsystemmodellierung <Stuttgart>: Mitteilungen

Herausgegeben 2022 vom Eigenverlag des Instituts für Wasser- und Umweltsystem-modellierung

Druck: DCC Kästl e.K., Ostfildern

Danksagung

Ich möchte mich ganz herzlich bedanken bei:

Professor András Bárdossy für die Chance der Promotion. Ich danke ihm für seine Geduld, dass gewisse Extremereignisse manchmal auf sich warten lassen. Ich habe sehr großen Respekt vor seinem enormen hydrologischen und statistischen Wissen sowie den immer neuen wissenschaftlichen Ideen. Es erfüllt mich mit Stolz, in seiner Arbeitsgruppe mitgewirkt zu haben.

Professor Markus Disse für die Bereitschaft der Übernahme des Mitberichts. Seine sympathische und umgängliche Art wirkte sich positiv auf die Vorbereitung zur Prüfung und die Prüfung selbst aus.

Professor Jan Hofmann für die kurzfristige und unkomplizierte Übernahme des Prüfungsvorsitzes sowie seiner sehr angenehmen Art die Prüfung zu leiten.

Jochen Seidel für die große Unterstützung und Hilfe bei Fragen zur Messtechnik und der unbürokratischen Organisation, dass die erforderlichen Rahmenbedingungen die ganze Zeit zur Verfügung standen. Ich schätze es sehr, dass mehr als eine rein kollegiale Verbindung entstanden ist, was sich heute noch durch private Treffen im Schwabenland oder Südbaden auszeichnet.

Astrid Lemp, dass sie die ganze Zeit für einen stetigen Informationsfluss zum Lehrstuhl gesorgt hat. Sie hat mich damit immer wieder dezent animiert, den Faden wiederaufzunehmen und zum Ende zu kommen. Ihre große Hilfsbereitschaft und immer freundliche Art waren unerlässlich, um die administrativen Dinge aus der Ferne organisieren zu können.

Meinen über die lange Zeit doch sehr zahlreichen Kolleginnen und Kollegen (Jan, Christian, Shailesh, Claus, Thomas P., Ferdi, Philipp, Jhan, Alejandro, Magdalena, Dirk, Taki, Theresia, Yulizar, Thomas M., Tobias, Micha, Faizan, Julia und Zlatko) für die sehr positive, hilfsbereite und freundschaftliche Atmosphäre, welche einen immer noch gerne an das Institut zurückkehren lassen. Hier möchte ich besonders meinen großen Dank an Sebastian („Hörni") für seine Hilfsbereitschaft beim Programmieren, der Generierung unglaublicher Datenmengen, seiner fachlichen Expertise und dem Korrekturlesen aussprechen. Mit ihm durfte ich im „Austragsbüro" in der letzten Zeit am Institut viele heitere und motivierende Momente verbringen, welche noch heute bis Down Under anhalten.

Den Werkstätten der Versuchsanstalt für den Bau und die Umsetzung der zahlreichen Messeinrichtungen. Hier gilt mein besonderer Dank Steffen Hägele für die schier unmöglichen Modifikationen von Messgeräten, seine grenzenlose Hilfsbereitschaft und immer gute Laune, im wie außerhalb des E-Labors.

All meinen Hiwis und Studierenden (Maud, Meltem, Alex, Dr. Enkhan, Micha und Florian), welche mich in der Feld- und Projektarbeit tatkräftig unterstützt haben und teilweise mit an Grenzen (alpine Aufstiege mit Höhenangst, Abflussmessungen bei Schneefall und Dunkelheit, Dichtheitstests von Wathosen im Tiefbehälter) gegangen sind.

Dr. Fabrizio Fenicia von der Eawag für die Einführung und Bereitstellung des Modells „Batea Superflex" sowie dem Software-Update, selbst noch nachmittags an Heiligabend.

Meinen Eltern dafür, dass sie mir eine unbeschwerte Ausbildung und Studium ermöglicht haben sowie meinen Geschwistern Thorsten und Carolin für das gegenseitige Erlernen der erforderlichen Sozialkompetenzen im gesellschaftlichen Umgang miteinander.

Meiner Frau Maren für ihre jahrelange Geduld, der Hinnahme zahlreicher Entbehrungen bei Freizeitaktivitäten und dem unermüdlichen Antrieb mit der Promotion abzuschließen. Ich danke ihr für die hilfreichen Impulse und anregenden Diskussionen. Zuletzt freue ich mich unheimlich über unsere bezaubernde Tochter Sophie, welche unser Leben nun auf ganz neue Weise erfüllt und bereichert.

Contents

List of Figures

List of Tables

Abbreviations

$BATEA$	Bayesian total error analysis
BIS	GeoFachdatenAtlas of Bavarian Soil Information System, LfU
$BMBF$	German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung)
BMI	German Federal Ministry of the Interior (Bundesministerium des Innern)
BMU	German Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety (Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit)
$COVER$	Percentage of validation data included within the $95\,\%$ uncertainty bands
CWB	Climatic Water Balance
DEM	Digital elevation model
DWD	German Weather Service (Deutscher Wetterdienst)
EDK	External drift kriging
EDK_s	External drift kriging with smoothed DEM
$eHYD$	Electronic Hydrographic Service of Austria
EP	Relative peak error
ESM	External snow model
ET	Time to peak error
EV	Relative volume error
$EvaSim$	Coupled Traffic and Hydraulic Simulation to Aid in Emergency Response Planning
GKD	Bavarian Hydrological Service (Gewässerkundlicher Dienst Bayern)
HAD	Hydrological Atlas of Germany, BMU
HBV	Rainfall-runoff model developed at the Swedish Meteorological and Hydrological Institute
HQ	Highest observed flood discharge within a reference time period

HS	Hydrological summer
HW	Hydrological winter
HY	Hydrological year
IWS	Institute for Modelling Hydraulic and Environmental Systems at the University of Stuttgart
KGE_m	Modified Kling-Gupta efficiency
LfU	Bavarian Environment Agency (Bayerisches Landesamt für Umwelt)
$MCMC$	Markov chain Monte Carlo
MHQ	Mean flood discharge within a reference time period
MI	Modeling interval
MQ	Long-term mean discharge
NSE	Nash-Sutcliffe efficiency
ODK	Ordinary kriging
$PBIAS$	Percent bias
QC	Quality control
$Q-Q$	Quantile-quantile plot
R	Correlation coefficient
RM	Random mixing
$RMSE$	Root mean squared error
ROS	Rain-on-snow
R^2	Coefficient of determination
S	Standard deviation
$SHARPNESS$	Measure of the distance between the lower and upper uncertainty bands
SLS	Standard least square method
$SMHI$	Swedish Meteorological and Hydrological Institute
SRM	Snowmelt-runoff model
$SUPERFLEX$	Flexible model framework (*Fenicia et al.*, 2011)
$SWAT$	Soil and Water Assessment Tool
UE	Unusual event
WLS	Weighted least square method
$ZAMG$	Central Institution for Meteorology and Geodynamics (Zentralanstalt für Meteorologie und Geodynamik), Austria

Symbols

Symbol	Definition	Unit
α	Weights of linear equation system for RM	
β	HBV model parameter of runoff production	
ΔS	Catchment storage and release change	
γ	Variogram	
λ_i	Weight of observation at location i	
ν	Langrange multiplier	
ATI	Antecedent temperature index	$[^\circ C]$
C_{ET}	HBV model parameter for PE change	$[^\circ C^{-1}]$
c_w	Constant for DEM smoothing	$[-]$
d	Smoothing distance	$[m]$
DD	Degree-day factor	$[mm\,d^{-1}\,^\circ C^{-1}]$
DD_m	Modified degree-day factor	$[mm\,d^{-1}\,^\circ C^{-1}]$
DD_w	Rain-on-snow weight	$[d^{-1}\,^\circ C^{-1}]$
ET_A	Actual evapotranspiration	$[mm]$
FC	Field capacity	$[mm]$
h_A	Discharge depth	$[mm]$
h_{ij}	Distance between station pair i,j	$[m]$
h_{MAP}	Mean areal precipitation depth	$[mm]$
h_{SWE}	Mean areal snow water equivalent	$[mm]$
I	Interception	$[mm\,t^{-1}]$
k	Weighting factor of the Muskingum method	$[-]$
K_c	Crop coefficient	$[-]$
m	Slope of regression curve	$[-]$
$MAXBAS$	Base length of the triangular unit hydrograph	$[h]$
N_{dat}	Data quantity of total data set	$[\%]$
P	Precipitation depth	$[mm]$
P_{eff}	Effective precipitation	$[mm]$
PE	Potential evapotranspiration	$[mm]$

PE_a	Mean annual potential evapotranspiration	$[mm]$
PE_M	Monthly long-term mean potential evapotranspiration	$[mm]$
PE_{tot}	Total potential evapotranspiration	$[mm]$
PWP	Permanent wilking point	$[mm]$
Q	Discharge	$[mm\,t^{-1}]$, $[mm]$
Q_m	Mean discharge	$[mm\,t^{-1}]$, $[mm]$
Q_{max}	Maximum discharge	$[mm\,t^{-1}]$, $[mm]$
$Q_{obs,m}$	Mean observed discharge	$[mm\,t^{-1}]$, $[mm]$
Q_{sum}	Sum of discharge	$[mm]$
Q_{tot}	Discharge of the total time series	$[mm\,t^{-1}]$, $[mm]$
Q_w	Discharge of the hydrological winter half-year	$[mm\,t^{-1}]$, $[mm]$
R_A	Extraterrestrial radiation	$[mm\,t^{-1}]$
RK	Rank of analyzed quantity	$[-]$
SM	Actual soil moisture	$[mm]$
S_{melt}	Amount of snowmelt	$[mm]$
SWE	Snow water equivalent	$[mm]$
$TIPM$	Antecedent temperature index parameter	$[-]$
t	Temporal resolution	$[t]$
T	Temperature	$[°C]$
T_m	Mean temperature	$[°C]$
T_M	Monthly long-term mean temperature	$[°C]$
T_{max}	Maximum temperature	$[°C]$
T_{min}	Minimum temperature	$[°C]$
w	Weight for DEM smoothing	$[-]$
W	Water level	$[cm]$
x	Retention constant of the Muskingum method	$[s]$
$Y_i(x)$	Unconditional random field	
$Z(x)$	Spatial random field	

Model Configurations

The following table contains a brief description of the most important model configurations used in Chapters 6 and 7. It serves as a reference for better understanding the discussion of the results and conclusions.

ID	Description
$I(45)$, $I^*(59)$, $II(34)$, $III(61)$, $IV_d\,(24)$	Roman numerals represent the basic daily model configurations, which differ by the number of gauging stations (see parentheses) used for the spatial precipitation interpolations and simulations; see Tables 3.2 and 6.1.
ODK, EDK, EDK_s	Various kriging methods for interpolating precipitation (see Abbreviations and Section 4.3). In each case, the interpolations are performed with the same rain gauge configuration.
RM, $RM_{>700}$, $RM_{>1000}$	Spatial precipitation simulations using random mixing (RM) as well as accounting for *greater-equal* constraints at gauging stations above certain elevations ($RM_{>700}$ above $700\,m.a.s.l.$ and $RM_{>1000}$ above $1000\,m.a.s.l.$); see Section 4.4.
$C1_{WLS}$ - $C4_{WLS}$	Application of different complex reservoir compositions within the FLEX model (see Table 5.3). The subscripts WLS (weighted) or SLS (standard) indicate the used least square method for calibration.
HY, HS, HW, $HS_{<20mm}$	Different model runs for the hydrological year (HY) as well as the hydrological summer (HS, $HS_{<20mm}$) and winter half-years (HW); see Subsection 6.3.1.
MI_1 - MI_5	Time-dependent analysis with varying modeling intervals (MI) in terms of length of time period and swapped time intervals for calibration and validation; see Table 6.15.
UE_1 - UE_4	Application of different discharge time series determined by several combinations of unusual events (UE); see Subsection 6.3.2.
MI^*_{11} - MI^*_4	Varying modeling intervals (MI^*, analogous to MI) with single disregarded years for model calibration and validation; see Table 7.1.
I_d, III_d, III_h, $IV_{h,ESM}$, $III_{h,ESM}$	Different rain gauge configurations and *liquid water* time series with different temporal resolution (d = daily and h = hourly) and consideration of an external snow model (ESM); see Subsections 7.2 and 7.3.

Abstract

Flood risk management is becoming increasingly important due to more frequent and severe flood events. Powerful and reliable hydrological models are required for decision-makers as a useful prediction tool. Therefore, the general focus in hydrological modeling should not only be on improving absolute performance by developing progressively more complex model structures. Rather, it is about identifying and understanding where and why certain modeling approaches fail and to find solutions accordingly.

This thesis initially provides a comprehensive quality control of the model input to systematically analyze possible data deficiencies. In addition to classical geostatistical interpolation methods, a new conditional simulation method (*random mixing* (RM)) is used to generate precipitation fields. By implementing inequality constraints, this stochastic approach allows for a stronger consideration of measurement uncertainties. Two different model structures are applied to test whether and to what extent a comparatively simple lumped model approach (FLEX) is inferior to the more complex spatially distributed HBV model. One great advantage of FLEX is the straightforward adaptation of the model structure and application to different temporal resolutions as well as relatively fast computation times.

Through own previous modeling attempts, it has been shown that the generation of robust model outputs often failed regardless of model complexity. To further investigate the problem of unstable model results, certain modifications are made to the modeling approaches. Model calculations are performed in hourly resolution to represent the fast-changing discharge conditions in smaller watersheds. A spatial densification of the sparse hourly observation network with disaggregated daily data is achieved. An external spatially distributed snow model (ESM) is developed for simultaneous consideration of different aggregate states of precipitation and snow processes. The concept of *liquid water* series is introduced to be able to combine an ESM and the FLEX framework.

Another aspect of this thesis is a comprehensive, often neglected, assessment of data and model adequacy for a better understanding of the implications for hydrological modeling. A larger number of performance measures than usual are applied and evaluated. A time-dependent model evaluation addresses the influence of time series lengths and different modeling intervals, as well as separate modeling of hydrological seasons. The difficulty of determining an unique parameter set that describes different discharge conditions equally well is also discussed. In this context, an approach is tested which uses time series with reduced data input determined by unusual events. The optimized parameter set can then be applied selectively to model certain prevailing discharge conditions. In investigating the model parameterization, precipitation time series simulated using RM are used to consider the problem of equifinality. Finally, based on all these findings, the most promising modeling approaches are assembled and a conclusion on the predictive reliability is drawn.

Kurzfassung

Ein effizienter Hochwasserschutz und das dazugehörende Hochwasserrisikomanagement gewinnen aufgrund häufiger auftretender und schwerer Hochwasserereignisse zunehmend an Bedeutung. Es werden hierfür leistungsfähige und zuverlässige hydrologische Modelle benötigt, welche an die Vor-Ort-Situation angepasst sind und den Planern sowie den Entscheidungsträgern als Vorhersageinstrument dienen. Bisherige Studien zeigen, dass hydrologische Modelle oft noch nicht auf einem zufriedenstellenden und verlässlichen Niveau für den operativen Einsatz sind.

Die Motivation für diese Arbeit entstand aus einem Projekt, welches sich mit der Notfallplanung und Evakuierungsmaßnahmen für urbane Gebiete und Verkehrsinfrastruktur bei extremen Hochwasserereignissen befasste. Das ausgewählte Untersuchungsgebiet weist eine sehr heterogene und alpine Topographie auf. Die Aufgabe der Hydrologie war es einen robusten und verlässlichen Modellansatz zu finden, um bis dato unbeobachtete hydrometeorologische Extremereignisse generieren zu können. Aktuelle Entwicklungen in der hydrologischen Modellierung zeigen, dass der Trend zu immer komplexeren Modellstrukturen mit höherem Daten- und Parameterbedarf geht. Der damit einhergehende, oft einseitige Fokus auf der Leistungssteigerung von Kalibrierungsergebnissen und eine oftmals unzureichende Modellvalidierung, können die erforderlichen Voraussetzungen für zuverlässige Vorhersagen vortäuschen. Diese Studie konzentriert sich nicht nur auf die Verbesserung der absoluten Modellgüte durch die Anwendung verschiedener Modell- und Prozessmodifikationen. Vielmehr geht es darum, zu erkennen und zu verstehen, wo und warum bestimmte Modellierungsansätze scheitern, um anschließend Lösungen für Verbesserungen zu finden. Die gängige Bewertung der Modellgüte erfolgt anhand der modellierten Ausgabedaten. Es wird dabei häufig versucht die Abweichungen von beobachteten und modellierten Daten durch eine Änderung der Modellstruktur und Parameter zu korrigieren. Diese Vorgehensweise basiert auf der weit verbreiteten Annahme, dass die Fehlervarianz grundsätzlich durch die modellierten Daten hervorgerufen wird und die Messwerte als fehlerfrei angenommen werden. Aus diesem Grund wird in dieser Arbeit ein besonderes Augenmerk auf die Eingangsdaten und deren Verarbeitung gerichtet, bevor sie an die verwendeten Modelle übergeben werden.

Zu Beginn der Untersuchungen konnten zwei wesentliche Schwächen der verwendeten Modellkonfigurationen identifiziert werden. Einerseits haben alle Konfigurationen, unabhängig von ihrer Komplexität, erhebliche Defizite bei der Generierung robuster Modellergebnisse. Dies macht sich durch einen deutlichen Unterschied der Modellgüte zwischen Kalibrierung und Validierung bemerkbar. Andererseits sind deutliche Probleme bei der korrekten Wiedergabe bestimmter Abflussbedingungen zu erkennen. Der Wasserhaushalt von alpinen Einzugsgebieten wird stark durch die Speicherung und Abgabe von Wasser in Form

von Schnee und Eis beeinflusst. Aus hydrologischer Sicht ist die Schneeschmelze ebenfalls von besonderem Interesse, da sie im Frühjahr mit überlagernden intensiven und lang anhaltenden Niederschlägen ein hohes Hochwasserpotenzial aufweist. Anfängliche Modellierungsversuche ergaben inakzeptable Ergebnisse für die definierte Schneeschmelzperiode. Aufgrund der Aufgabenstellung und damit verbundenen Ziele ist die Darstellung der markanten Hochwasserabflüsse entscheidend. Viele Modellierungsansätze haben gezeigt, dass die Scheitelabflüsse bei der Kalibrierung und insbesondere bei der Validierung deutlich unterschätzt werden. Die Scheitelabflüsse werden bei der Validierung oft nicht einmal von den Unsicherheitsbändern erfasst. Solche Modellergebnisse sind keine Grundlage für zuverlässige Vorhersagen von bisher unbeobachteten Hochwasserereignissen.

Um die potentiellen Ursachen für die vorgenannten Defizite systematisch zu erfassen, wird zunächst eine umfassende Qualitätskontrolle der Eingangsdaten durchgeführt. Dabei werden geostatistische Interpolationsmethoden unterschiedlicher Komplexität miteinander verglichen. Es wird eine neue bedingte Simulationsmethode (Random Mixing (RM)) für die räumliche Interpolation von Niederschlagsfeldern angewendet. Dieser stochastische Ansatz ermöglicht durch die Einbindung von linearen Ungleichheitsbedingungen (größer-gleich Bedingungen) eine stärkere Berücksichtigung von Messunsicherheiten. Die Datenanalyse zeigt, dass unabhängig von den verwendeten Konfigurationen der Niederschlagsstationen und der angewandten Interpolationsmethoden ein deutliches Defizit in der Wasserbilanz besteht. Die Verwendung der Konfiguration mit den meisten und räumlich am gleichmäßigsten verteilten Niederschlagsstationen führt aber bereits zu einer Verringerung des Wasserdefizits, insbesondere in den Winterhalbjahren. Die Berücksichtigung der Topographie als Zusatzinformation bei den Kriging-Verfahren wirkt sich zusätzlich positiv auf die Wasserbilanz aus. Ein weiterer Aspekt ist, dass akkumulierte räumliche Niederschlagsmengen in der Regel nicht durch die kleinräumige Variabilität der Topographie beeinflusst werden. Daher wird ein geglättetes digitales Höhenmodell (DEM) verwendet, um die unplausible Nachbildung der zugrunde liegenden kleinräumigen Geländestruktur in den Niederschlagsfeldern zu korrigieren. Mittels RM simulierte Niederschlagsfelder reduzieren das Wasserdefizit um mehr als die Hälfte in dieser Studie. Die Wasserbilanz kann nahezu ausgeglichen werden, wenn an den höchstgelegenen Niederschlagsstationen größer-gleich Bedingungen für die Simulationen berücksichtigt werden. Damit ist RM unter Einbindung von größer-gleich Bedingungen geeignet Messunsicherheiten, insbesondere bei schwierigen topographischen Randbedingungen, besser zu beschreiben.

Die Anwendung der beiden unterschiedlichen Modellstrukturen zeigt, dass ein vergleichsweise einfacher räumlich konzentrierter Modellansatz (FLEX, lumped) der komplexeren räumlich verteilten Struktur des HBV-Modells (distributed) nicht grundsätzlich unterlegen ist. Ein großer Vorteil des FLEX-Konzepts ist die relativ unkomplizierte Anpassung der Modellstruktur und Anwendung auf unterschiedliche zeitliche Auflösungen. Die verhältnismäßig schnellen Rechenzeiten des Modells sind ein weiterer Pluspunkt. Somit können mit vertretbaren Zeitaufwand Modellläufe in täglicher Auflösung mit hunderten simulierter Niederschlagszeitreihen durchgeführt werden. Unabhängig von der Modellkomplexität konnten bei vielen Modellierungsversuchen dennoch keine robusten Ergebnisse erzielt werden. Dies ist ein Hinweis darauf, dass dieses Problem nicht primär von der Modellstruktur oder der Parametrisierung verursacht wird. Das Gleiche gilt für die ungenügend dar-

gestellten Abflussbedingungen. Für die Modellierung der Schneeschmelzperioden erweist sich die räumlich verteilte Berücksichtigung der Eingangsdaten Niederschlag und Temperatur im HBV-Modell als vorteilhafter. Aus dieser Erkenntnis wird das Konzept der *liquid water* Zeitreihen entwickelt. Dazu wird die Flexibilität der FLEX-Struktur genutzt, indem der Schneeprozess im Modell deaktiviert wird. Stattdessen wird ein ausgelagertes Schneemodell (ESM) unter Anwendung des Gradtag-Verfahrens (Temperatur-Index-Methode) erstellt, welches die Akkumulations- und Schmelzprozesse in einem räumlich verteilten Raster berechnet. Damit wird die gleichzeitige Berücksichtigung verschiedener Aggregatzustände von Niederschlag und Schneeprozessen innerhalb des Einzugsgebiets möglich. Die räumlich verteilte Wasserverfügbarkeit aus verschiedenen Bezugsquellen wird anschließend zu einem räumlich konzentrierten Wert pro Zeitschritt aggregiert. Die resultierende *liquid water* Zeitreihe kann so an das FLEX-Modell übergeben werden. Diese Methode liefert auf Tagesbasis vielversprechende Ergebnisse, obwohl die Modellleistung für die Schneeschmelze besser, aber immer noch unbefriedigend ist. Daraus lässt sich schließen, dass für eine sehr heterogene und alpine Topographie eine räumlich konzentrierte Darstellung der Schneeprozesse allein nicht ausreichend ist. Die Kombination aus einem räumlich konzentrierten Grundgerüst zur Modellierung der Hauptabflusskomponenten und der individuellen Auslagerung von komplexen Prozessen erweist sich hier als geeignetes Werkzeug.

Das FLEX-Modell eignet sich aufgrund seiner verhältnißmäßig schnellen Rechenzeiten sehr gut für höher zeitlich aufgelöste Modellierungen. Daher wird im Rahmen dieser Arbeit auch der Einfluss der zeitlichen Auflösung eingehend untersucht. Die vorhandenen Niederschlagsstationen mit stündlichen Daten sind sehr spärlich, weshalb das Messnetz zusätzlich mit disaggregierten Daten von Tagesstationen verdichtet wird. Die täglichen Daten werden mit der Nächste-Nachbarn-Methode disaggregiert, welche es ermöglicht die stündliche Niederschlagsverteilung an jeder Tagesstation zu bestimmen. Die ausgewerteten stündlichen Modellläufe weisen für alle verwendeten Stationskonfigurationen robustere Modellergebnisse auf. Dies deutet darauf hin, dass in kleineren Einzugsgebieten eine hohe zeitliche Auflösung erforderlich ist, um die sich schnell ändernden Abflussbedingungen ausreichend gut darzustellen. Die zusätzliche räumliche Verdichtung des Messnetzes durch disaggregierte Tagesdaten wirkt sich somit positiv auf die Modellleistung und Robustheit aus.

Ein weiterer Aspekt dieser Arbeit ist eine umfassende, oft etwas vernachlässigte, Evaluierung der Daten- und Modelleignung, um ein besseres Verständnis für deren Auswirkungen auf die hydrologische Modellierung zu erlangen. Es wird zunächst eine größere Anzahl von Gütemaßen als üblich angewendet und ausgewertet. Einerseits sollte nicht außer Acht gelassen werden, dass Gütemaße sehr empfindlich auf einzelne, unzulänglich modellierte (Hochwasser-) Ereignisse reagieren können. Andererseits kann eine kontinuierliche Über- oder Unterschätzung der Abflussbedingungen sich kaum nachteilig auf das Leistungsniveau auswirken. Es wurde gezeigt, dass die Bewertung nur auf Basis eines Gütemaßes (z. B. NSE) zu falschen Bewertungen bezüglich der Qualität und Zuverlässigkeit der Modellausgabe führen kann. Andere Maße, wie z. B. die modifizierte Kling-Gupta Kennzahl (KGE_m), weisen in dieser Studie eine geringere Sensitivität zwischen den verschiedenen Modellansätzen auf, was bei der Auswertung ebenfalls schwierig zu interpretieren sein und zu falschen Entscheidungen führen kann. Gütemaße, wie der $PBIAS$, welche den Modellfehler anhand des Trends zur Über- oder Unterschätzung des Volumens beschreiben, erfor-

dern bei der Interpretation ebenfalls eine gewisse Vorsicht. Eine scheinbar ausgeglichene Volumenbilanz über den Gesamtmodellierungszeitraum kann unter Umständen nur durch Kompensationseffekte einzelner deutlich voneinander abweichender und relativ schwacher Ergebnisse verursacht werden, wie die jahresweise Auswertung der Modellgüte gezeigt hat.

Eine zeitabhängige Modellevaluierung kommt zu dem Ergebnis, dass die Länge der Zeitreihen, hydrologische Jahreszeiten oder unterschiedliche Modellierungsintervalle nur eine untergeordnete Rolle spielen. Hier sind vielmehr ausreichend häufige und möglichst unterschiedliche hydrologische Bedingungen, wie z. B. ausgeprägte Hochwasserereignisse und längere Niedrigwasserperioden, entscheidend. Diese sollten sich möglichst gleichmäßig über den gewählten Kalibrierungs- und Validierungszeitraum verteilen. Die saisonale Modellierung mit hydrologischen Sommer- und Winterhalbjahren zeigt, dass die erhaltenen Parametersätze jeweils unterschiedliche Informationen für das hydrologische Pendant liefern. Das ist ein Beispiel dafür, wie schwierig es ist, einen einzigen Parametersatz zu bestimmen, welcher unterschiedlichste Abflussbedingungen gleich gut beschreibt. Ein Ansatz könnte sein, dass verschiedene Parametersätze separat auf Basis unterschiedlicher Abflussbedingungen optimiert werden. Die so ermittelten Parametersätze können dann jeweils nach den vorherrschenden Bedingungen angewendet werden. Ein solcher Ansatz wird anhand von Zeitreihen mit reduziertem Dateneinsatz dargestellt, welche mit Hilfe von außergewöhnlichen hydrometeorologischen Ereignissen definiert werden. Es kann, mit Ausnahme der Schneeprozesse, eine robuste Parametrisierung mit deutlich weniger Eingangsdaten und geringerer Modelllaufzeiten für die restlichen Abflussbedingungen erreicht werden. Das Problem der Äquifinalität wird bei der Analyse der Modellparametrisierung thematisiert. Hierbei erweisen sich die mittels RM simulierten Niederschlagszeitreihen als geeignet, um ausreichend viele Modellläufe mit identischer Modellgüte sowohl bei der Kalibrierung als auch der Validierung zu erhalten. Damit lässt sich zeigen, dass die optimalen Parametersätze variieren können. Diese Erkenntnis kann eine klare Entscheidung für den besten Modellansatz erschweren und muss auch im Hinblick auf die Übertragbarkeit der Parameter auf andere Modellierungszwecke berücksichtigt werden.

Schließlich werden aus all diesen einzelnen Erkenntnissen die vielversprechendsten Modellierungsansätze zusammengestellt. Auf der Grundlage dieser schrittweisen Anpassung der Modellierungsansätze wird eine Schlussfolgerung über die Vorhersagesicherheit gezogen. Die drei wichtigsten Modifikationen, welche zu einer wesentlich verbesserten Entwicklung sowohl bei der Schneeschmelze als auch bei den Spitzenabflüssen beitragen, sind: Modellrechnungen in stündlicher Auflösung, eine räumliche Verdichtung des Messnetzes mit disaggregierten Tagesdaten sowie ein externes räumlich verteiltes Schneemodell (ESM). Die Hauptquellen von Unsicherheit können in dieser Studie auf diese drei Einflussfaktoren zurückgeführt werden. Es wird sehr deutlich, wie wichtig eine qualitativ hochwertige räumliche und zeitliche Datenerfassung als Grundlage für zuverlässige Modellierungsansätze ist. Es ist eine trügerische Annahme, dass immer komplexere Modellstrukturen die Verwendung von realen Daten ersetzen können. Das genaueste Modell nutzt nichts, wenn es nicht mit korrekten Daten versorgt und verifiziert wird. Die zeitunabhängige Auswertung mit Abflussdauerlinien bestätigt, dass die Konfiguration, welche die zuvor genannten verbessernden Modifikationen beinhaltet, die statistischen Abflusseigenschaften am besten wiedergibt. Die Auswertung der höchsten Abflussereignisse zeigt unter Berücksichtigung der

Unsicherheiten ebenfalls, dass nur dadurch ein zuverlässiger Modellierungsansatz erreicht wird, welcher für Hochwasservorhersagen anwendbar erscheint. Dennoch müssen Vorhersagen von bisher unbeobachteten extremen Hochwasserereignissen behutsam interpretiert und bei Entscheidungen angewandt werden, da sie eine Extrapolation ins Unbekannte mit zunehmenden Unsicherheiten darstellen.

Diese Arbeit zeigt die Eignung verschiedener Ansätze zur Datenverarbeitung und Modellauswahl, um von anfänglich inakzeptablen zu sehr guten und robusten Modellergebnissen zu gelangen. Anhand vieler verschiedener Evaluierungen lässt sich erkennen, welche Faktoren bei der hydrologischen Modellierung berücksichtigt werden müssen und welche offenbar eher eine untergeordnete Rolle spielen. Das Ziel dieser Arbeit ist es, die Einzugsgebietseigenschaften, die erforderliche Datengrundlage, das Prozess- und Modellverhalten sehr umfassend zu untersuchen. Daher wurden alle Untersuchungen an einem einzigen Einzugsgebiet durchgeführt. Die Ergebnisse dieser Studie sollen bei vergleichbaren Fragestellungen helfen, um beim Thema Datenverarbeitung und Modellauswahl zielgerichteter vorgehen zu können. Bei weiteren Anwendungen sind die vorgeschlagenen Lösungsansätze und ihre Gültigkeit an einer größeren Anzahl von Einzugsgebieten mit unterschiedlichen Eigenschaften sowie klimatischen Bedingungen zu untersuchen.

1 Introduction

More frequent and severe flood events demonstrate the importance of efficient flood protection and the corresponding flood risk management. In June 2013, one of the heaviest floods in recent times occurred in Central Europe and caused noteworthy damages along the huge rivers Danube and Elbe, as well as in smaller catchments in the Alps. The Federal Ministry of the Interior (BMI) estimated the preliminary amount of damages to be around 6.7 billion Euro (*BMI*, 2013). To reduce the risk of heavy damages in the future, hydrological modeling must be increasingly considered as an option for verifying predictions of extreme flood events. Thus, an important basic for the establishment of the reliability of simulation models is an active communication between model developer and its possible users (*Schlesinger et al.*, 1979; *Refsgaard and Henriksen*, 2004).

1.1 Motivation

In literature, the application possibilities and limitations of hydrological models as well as existing uncertainties are controversially discussed. It shows that the state of science is still going beyond a satisfactory level. *Savenije* (2009) argued that good models do not exist. Thus, the development of more appropriate models must be the challenge rather than the competition to find the *best* model. Finding an universal model that performs equally well independently of the application area and the boundary conditions is practically impossible (*Andréassian et al.*, 2009). In many studies, the enhancements focus only on increasing model complexity and the involved improvements in calibration, but an adequate validation and analysis of uncertainty is often missing.

"Difficulties in defining truly mechanistic model structures and difficulties of model calibration and validation suggest that the application of distributed hydrological models is more an exercise in prophecy than predictions."

(*Beven*, 1993)

This upper statement illustrates the dilemma of achieving reliable predictions in hydrological modeling. This is a good link to the motivation of this thesis which has been originated within the project $EvaSim$ (Coupled Traffic and Hydraulic Simulation to Aid in Emergency Response Planning). The research project was funded by the German Federal Ministry of Education and Research ($BMBF$). Within this project, different scientific disciplines such as

hydrological and hydraulic engineers, traffic planner and sociologist, as well as users from engineering companies and decision makers from public institutions worked together. The aim was to improve emergency response planning for urban areas and transportation infrastructure that can be affected by extreme flood events caused by heavy precipitation or dam failures. With the linkage of hydrological models, flood wave spreading simulation as well as the simulation and optimization of traffic flow evacuation scenarios in the case of emergency were developed on real investigation areas. As final result, a tool which is transferable to any other similar region was provided. The focus for hydrology in this project was on investigating extreme precipitation events and creating unobserved flood scenarios. The modeled hydrographs were used as input for the hydraulic models and the derivation of flood hazard maps.

The results of this project confirmed the difficulties of generating reliable hydrological model predictions in order to apply them to the creation of appropriate rescue scenarios. On the one hand, there is the need of preferably exact results for the users (e.g., hydraulic modelers or decision makers) as a prediction tool. Then, there is the different understanding that not only one solution exists but also a range of possible outcomes associated with uncertainties. On the other hand, the obtained results, providing a good model performance in calibration, could pretend a reliable solution. But, a careful evaluation of the results has shown that especially parts of interest, e.g., flood events are represented relatively poor. Thus, the simulation of synthetic flood events does not provide very reliable predictions. However, such extreme events can cause heavy damages or in the worst case even cost human life. In this context, the risk of an uncertain or even wrong prediction could have serious consequences. So, further investigations on data processing techniques and an appropriate model choice are important.

1.2 Scope of the Thesis

A major problem in hydrological modeling is dealing with model inconsistencies. *Beven* (2007) stressed that model rejection may have a more beneficial learning effect than the acceptance of a model. Identifying a model weakness or a process inadequately described by the model, can contribute to further development and improvement. A model rejection is often caused by inadequate input data or observations which are used for calibration and validation.

This study addresses these issues such as data or model inadequacy. The influence of data configuration and processing on the models is quantified. The appropriate model choice for certain modeling purposes is investigated on the basis of different complex models and process structures. In this study, the model assessments are always performed using various graphical, absolute and relative performance measures. Depending on the efficiency criterion, the evaluation may focus on fulfilling different hydrological constraints, such as discharge volume, high or low flow conditions. This approach is intended to show that some sensitivity on the part of the decision maker is required in the selection of performance measures and the resulting model evaluation in order to draw the correct conclusions from such

model outputs. The work presents approaches how the applied modifications can overcome some limitations based on data and model inadequacies including uncertainty. Finally, the purpose is to evaluate the predictive reliability, especially for flood events. The investigations performed are divided into the following three main parts:

Effects of Model and Process Complexity
Numerous hydrological models with a wide range of complexity are available. Within this study, a distributed and a lumped model structure are applied. In addition, the lumped model varies in its complexity by different configurations of the process describing reservoirs and corresponding parameter sets. Both models are used for the investigation of an alpine headwater catchment.

In this context, it is investigated if a distributed model is more appropriate than a simple lumped approach. Likewise, it is tested if within the lumped model structure different complex reservoir configurations result in different model performances. These model structures differ in various data requirements and set-ups. As a consequence, the time required to generate model output varies considerably. Modeling of snow processes is a sophisticated task and often handled with relative simple assumptions in hydrological models. Therefore, the question arises whether a lumped model structure is able to precisely describe snow processes in a topographically highly variable area? For comparison, a distributed external snow model with various modifications is applied and the concept of *liquid water* time series is introduced.

This section also investigates to which extent temporal aggregation or disaggregation can be an instrument for handling data or model limitations. This issue does not deal with complexity in the sense of complex process descriptions, but rather about the fact that aggregation or disaggregation require a higher effort for data preparation and processing as well as for model calibration and evaluation. Within this framework, the effect of modeling on a higher temporal resolution (e.g., hourly) and the following aggregation of the model results (e.g., daily) is analyzed. In addition, it is tested whether the incorporation of additional data (e.g., observations of precipitation) on a lower temporal resolution cause a benefit for high-resolution model calculations. For this purpose, an approach using the *nearest-neighbor* method for disaggregating daily data is developed.

Assessment of Data and Model Adequacy
The target of this section is the evaluation of model output for their performance level and their predictive reliability. These include aspects related to data quality, data processing, and model robustness. The temporal and spatial available observations of the required input data are often sparse or of limited quality. Hence, such an influence on model performance and robustness is analyzed using various rain gauge densities and distributions as well as modified temperature time series and possible data discrepancies.

In a further step, it is examined to what extent the model performance is influenced by the way the input data are processed. Several conventional spatial interpolation methods are used here, which are compared with a new conditional simulation method (*random mixing*). The purpose of applying such a simulation method is to investigate the influence of input uncertainty on the model calibration and its output. Measurement uncertainties due to the

alpine topography of the study area are more strongly considered with the implementation of linear inequality constraints.

A time-dependent performance analysis deals with various issues related to time periods and durations of calibration and validation. The goal is to determine if such factors have an impact on a high-performing and robust modeling approach. In addition, unusual hydrological sequences are used in an event-based modeling approach to determine if robust model parameterizations can be achieved with significantly reduced data and time requirements. In the context of determining robust parameter sets, the issue of equifinality is also addressed.

Model Reliability - Applicability for Flood Prediction
The aim of this section is the evaluation of model outputs for their predictive reliability and finally the prediction capability of severe flood events. For this purpose, the most promising model approaches are combined and evaluated based on the results of the previous findings. The extent to which different complex process descriptions (e.g., snow modeling) or temporal resolution are required for the correct representation of specific hydrological conditions is demonstrated. This topic concludes with a comparison of the development of hydrological conditions identified as weaknesses during the modeling process.

1.3 Outline of the Thesis

The thesis starts in **Chapter 2** with a general introduction to the state of science in hydrological modeling, followed by the presentation of the two applied models. Finally, the approaches to evaluate the model performance are summarized.

In **Chapter 3**, the study area and its properties are introduced. A brief description of the data availability, the selected time period and the different used station configurations is provided.

Chapter 4 deals with data validation and processing. First, the sources of uncertainty and the quality of the required data are surveyed. Conventional statistical methods such as kriging and a new conditional simulation method (*random mixing*) for processing meteorological input data are then applied and compared.

In the next sections, aspects concerning model complexity, model choice and model processing are addressed. These investigations are shown and discussed in **Chapter 5**.

An assessment of data and model adequacy is provided in **Chapter 6**. Inconsistencies of model results based on data inadequacy as well as modeling weakness are evaluated. In addition, the influence of different data properties, processing and calibration strategies are analyzed.

Finally, the model reliability and prediction capability for severe flood events are evaluated in **Chapter 7**.

2 Hydrological Modeling

Hydrological modeling has become a widely used discipline in hydrology. *Savenije* (2009) defines hydrology as the science explaining the occurrence and behavior of water above, over and through the Earth. In this context, a model is a simplified description of a complex natural system. A hydrological system includes several physical, chemical and biological processes which are used as input variables for a model (*Clarke*, 1973). It uses mathematical, statistical and numerical functions to express input variables and the processes of a hydrological system transforming them to output variables. Models in hydrology represent certain parts of the hydrological system like climate, surface and subsurface processes, as well as their interactions. *Andréassian et al.* (2009) describes a hydrological model as a set of equations that allow discharge simulations based on input data and a set of parameters. *Garrick et al.* (1978) have already pointed out that for a given catchment the rainfall-discharge relationship can be modeled comparatively well by simple conceptual models if proper parameter values are selected. In summary, a model is used to compensate the lack of sufficient temporal or spatial knowledge of processes of interest (*Oreskes et al.*, 1994).

In the following, a hydrological model is considered as rainfall-runoff model. The intended use of these models is divided into forecasts and predictions. A forecast model deals with the estimation of prospective discharge conditions which can also be operated in real-time. On the other hand, predictions evaluate the effects of physical or climatic changes on the hydrological system. Within the vast availability of hydrological models, the distinction between purely scientific and engineering models exists. A scientific model is a testing scheme where a theory or hypothesis of a real-world system is verified against empirical evidence (*Silberstein*, 2006). From the viewpoint of an engineer, a model is primarily a tool based on a theory, with the focus on the functional adequacy in terms of a particular application (*Savenije*, 2009; *Gupta et al.*, 2012). For the credibility of models, a proper evaluation within comprehensive benchmarks of terminology is necessary. Here, e.g., *Schlesinger et al.* (1979) and *Refsgaard and Henriksen* (2004) distinguish between model qualification (adequacy of the conceptual model), model verification (confirmation of the conceptual model by the computer-based model) and model validation (ability for site-specific application).

Several authors argue that models are generally false and that models are known to be false (*Morton*, 1993; *Savenije*, 2009). It is stated that a kind of incompatibility of models exists, in the way that different models can suggest very different predictions for the same conditions. This can cause a dilemma in decision making. *Oreskes et al.* (1994) even argued that it is impossible to verify and validate numerical models of natural systems. As, natural systems are always open the obtained parameter sets and model results are never unique. Also, fixed model structures are not always applicable for variable conditions. Models are often used for a specific practical, as well as for a scientific target and do not guarantee transferability

to similar conditions outside the considered domain. Thus, the use of flexible model structures is an alternative to traditional fixed approaches. Flexible components allow the user to customize the model for its individual requirements and catchment conditions (*Fenicia et al.*, 2011; *Kavetski and Fenicia*, 2011). The development of models with an appropriate model structure, good model performance and small parameter and predictive uncertainty should be the scientific purpose (*Savenije*, 2009). *Gupta et al.* (2012) arranged the three stages of formal model building into conceptual, mathematical and computational model. Within this terminology, a conceptual model simplifies a natural system by specifying the boundaries, input, output and state variables, incorporation of physical laws and consideration of uncertainties. In a next step, a mathematical model determines the mathematical execution of states, fluxes, parameters and boundary conditions within the system. Finally, the numerical implementation of the above requirements is achieved with a computational model.

Another issue is that the examined physical systems are mostly inherently unpredictable. So, the trustworthiness of simple obtainable predictions has to be regarded very critical (*Morton*, 1993). Additionally, the association of terms such as *verification* and *validation* is often misleading in the context of model predictions and reliability. A verified model suggests that its truth is demonstrated and thus its reliability. But, proving the truth of any statement is just possible for closed systems. A numerical model is never a closed system because the required observations and parameters are incompletely known and accompanied by inferences and assumptions (*Oreskes et al.*, 1994). The second term *validation* characterizes the generation of acceptance. In this way, models which have a consistent numerical structure and do not have identifiable errors can be assumed valid. Nevertheless, in the context of modeling the term of *validation* is often used in an arguable way. *Oreskes et al.* (1994) mentioned that current model results of specific realizations are misleadingly accepted as valid and that it would even be more wrong to interpret a model as an adequate description of physical reality.

In recent studies, more and more issues regarding adequate model structures and complexity, proper parameterization and equifinality are discussed (*Hawkins*, 2004; *Beven*, 2006; *Fenicia et al.*, 2011; *Hrachowitz et al.*, 2014; *Del Giudice et al.*, 2015). There are no error-free models in terms of complete knowledge about the parameters, model structure and input data. Therefore, the modelers' interest shifts more to the structural adequacy of the model (*Gupta et al.*, 2008, 2012). Structural inadequacy results from the imperfect representation of a natural system by a model (*Del Giudice et al.*, 2015). Many authors discuss the issue of finding an appropriate level of model complexity (*Jakeman and Hornberger*, 1993; *Lindström et al.*, 1997; *Perrin et al.*, 2003; *Schoups et al.*, 2008; *Fenicia et al.*, 2008; *Kavetski and Fenicia*, 2011; *Hrachowitz et al.*, 2014). Too complex models can cause model overfitting and parameter equifinality problems (*Beven*, 1993; *Gan and Biftu*, 1996; *Hawkins*, 2004; *Schoups et al.*, 2008; *Gharari et al.*, 2014). Equifinality discusses the fact that different reasonable parameter sets can be obtained for representing the behavior of a system. The estimation of prediction uncertainty should consider this aspect and not just refuse other acceptable representations (*Beven*, 2006, 2007). In this context, *Hawkins* (2004) points out the principle of parsimony which declares the use of models that incorporate all requirements for the modeling purpose but nothing more. Nevertheless, *Doherty and Welter* (2010) emphasizes that model structural noise can be diminished by including as many parameters as possible to the calibration process. They

recommend the addition of non-explanatory model parameters to the calibration process to achieve a better model fit of the output variables. On the one hand, achieving reliable and robust predictions is an interaction of finding an appropriate model complexity with a minimum of parameterization and structural uncertainty. On the other hand, an adequate functional representation of the real system properties should be kept (*Hrachowitz et al.*, 2014).

2.1 Model Complexity

Since the scientific community pays more attention to the development of models, also the complexity increases continuously. For this reason, the choice of an appropriate level of complexity of the model is crucial and must first be decided for each individual application. This affects the requirements for computer capacity, input data, and finally the model to be used (*Bergström*, 1991). Traditional conceptual rainfall-runoff models require relative few data (discharge, precipitation, evapotranspiration and temperature) as input. In this regard, one may inquire if the input data even contains adequate information to support more complex models (*Gupta et al.*, 2008; *Schoups et al.*, 2008). The choice of model and its complexity is dependent on the data availability and modeling purpose. Also, the case of overfitting should be avoided. Overfitting defines models which imply more capabilities than required or the use of approaches which are more complex than necessary (*Hawkins*, 2004). Two types of overfitting can be specified. The use of nonessential flexible model structures or the consideration of irrelevant elements (*Hawkins*, 2004). *Gan et al.* (1997) specified three issues for a proper model choice. The selected model must be able to represent the fundamental hydrological processes of the catchment. Thus, the specific catchment properties affect the selection of a suitable model complexity (*Kavetski and Fenicia*, 2011). Another issue is the choice of temporal resolution of the model calculations. The higher the temporal resolution the more computing time is necessary. On the other hand, reproducing the dynamics of hydrological processes requires sufficiently high temporal resolution. An appropriate time step must be selected depending on the purpose and application of the individual modeling. Thirdly, a suitable model calibration is essential (*Gan et al.*, 1997).

Model complexity is defined in two ways within this study. First, the complexity depends on the model type, e.g., lumped or fully-distributed, and thus on the varying data requirement. Furthermore, the complexity is considered by the varying number of free parameters within the same model.

2.1.1 Model Types

There are many different types of hydrological models presented in literature. A number of authors suggest classifications of models into different groups. *Clarke* (1973) distinguishes between four main groups of models (stochastic-conceptual, stochastic-empirical, deterministic-conceptual and deterministic-empirical). Further, he groups models regarding their spatial representation of the input variables into lumped, probability-distributed and geometrically-distributed. A deterministic model provides the same output for a given

input condition and does not take randomness into account. Stochastic models estimate a range of likely output values from a given input condition (*Gayathri et al.*, 2015). In *Refsgaard and Knudsen* (1996) hydrological models are divided in three main groups (empirical black box, lumped conceptual and distributed physically based). An empirical model is a data driven model and does not consider any physical processes. The unit hydrograph is an example for this kind of model (*Beven*, 2012; *Gayathri et al.*, 2015). Conceptual models describe the hydrological catchment behavior by functions, parameters and reservoirs which are abstractions of physical processes (*Refsgaard and Henriksen*, 2004). These processes are the interactions of precipitation, evapotranspiration, percolation, drainage and runoff generation. Parameters are estimated by model calibration, which requires time series of various meteorological and hydrological data. The HBV and SUPERFLEX models presented in the next chapter belong to this group of models. Physically based models contain more principles of physical elements which are represented by *real* measured parameters. These kind of models do not require much effort on model calibration because their parameters have a direct physical interpretation (*Gayathri et al.*, 2015). Normally, physically based models are used in scientific applications for the small scale. In this case, physical parameters are well controlled and have a relative small variability (*Bergström*, 1991). As an example, the *SWAT* model can be mentioned here. *Seibert* (2003) argues that for calibrating conditions that are outside the range of available observations, physically based models may be more appropriate than conceptual models. In summary, the spectrum of models ranges from lumped conceptual (simple) to distributed physically based (complex) structures (*Gharari et al.*, 2014).

In this study, only conceptual models are applied. Therefore, these model types are described in more detail regarding their spatial representation of the included hydrological processes in the section below. A shortcoming of conceptual model systems is the fact that physical processes are often explained by parameters that are not directly related and measurable (*Gharari et al.*, 2014). Thus, such models need runoff data for calibration and validation as well as no considerable changes of catchment properties should have occurred within the modeling period. This may result in limited applicability to gauged catchments only (*Refsgaard and Knudsen*, 1996). The calibration of conceptual models is a challenge and has to deal with many factors such as nonlinearity of the models, large number of unknown parameters, observation errors and low information content of data (*Gupta and Sorooshian*, 1985; *Bárdossy and Singh*, 2008). In addition, the quality of parameter sets is affected by the conceptual model formulation and structure, the efficiency and robustness of the optimization algorithm and not least by the selected objective functions used for optimization (*Gan and Biftu*, 1996). Thus, the usability and reliability of hydrological models depend on an accurate calibration procedure (*Gupta et al.*, 1999).

Lumped models do not account for spatial heterogeneity in input data, model parameters, and process representation. The model represents the catchment dynamics as a single unit (*Beven*, 2012). These comparatively simple models, transforming precipitation into runoff, are easy to apply and have low data requirements for parameter estimation (*Perrin et al.*, 2001; *Del Giudice et al.*, 2015). In this study, the applied SUPERFLEX model has a lumped structure.

Semi-distributed models account for spatial catchment heterogeneity with zones of similar properties. The attributes for partitioning a catchment into zones are the same elevation, soil or land use class (*Das et al.*, 2008). The input data (precipitation, temperature, etc.) are prepared and transferred to the model according to the particular zones. Hydrological process descriptions and parameter estimation can also be considered separately for each zone. This model type is an intermediate stage between a lumped and a distributed model structure. It is able to include basic spatial heterogeneity of a catchment but still remains less complex than a distributed model.

Distributed models subdivide a catchment into smaller spatial units like regular grid cells. Based on this division, input data and parameter values are required for each single cell. The demand of spatially distributed information increases noticeable and especially the definition of the parameters for every single cell can generate difficulties. A reduction of necessary information for the parameterization can be obtained if parts of the hydrological processes are replaced by simplified conceptual modules (*Brath et al.*, 2004). The used distributed conceptual HBV-IWS model belongs to this type of model. Increasing complexity of the model and the connected optimization process also demand more computational resources than the previously mentioned model types.

2.1.2 Data Requirement

The quantitative data requirement usually rises with model complexity. For a basic lumped model structure discharge, precipitation and potential evapotranspiration time series are typically needed. If snow processes have to be taken into account, temperature data are also required. In more complex model frameworks additional data like long term mean data, land use or soil information is essential. The advantage of hydrological models with low data requirement is their easy operational utilizability (*Perrin et al.*, 2001).

Another kind of data requirement is the temporal or spatial resolution of the observations and input information. A lumped model requires uniformly distributed input data such as mean areal precipitation or temperature, which means that there is only one value considered for the whole catchment per time step. In contrast, a distributed model structure assumes spatially variable input data for the corresponding hydrological unit (e.g., subcatchments, zones or grid cells), which allows a more realistic representation of the catchment properties. Choosing the appropriate spatial resolution depends on the spatial variability of precipitation, catchment topography and size. Large and mountainous catchments are potentially more affected by such conditions (*Gan and Biftu*, 1996).

2.1.3 Process Parameter

The amount of parameters used to describe the hydrological processes in conceptual models varies considerably. Finding an adequate number of free parameters and thus a level of complexity is an important challenge in this context. Too few parameters restrict a sufficient model flexibility and too many cause problems of parameter estimation and model robustness (*Perrin et al.*, 2003). The model parameters are usually not directly measurable and

are assumed to be constant in time (*Wagener*, 2003). Thus, model parameters are identified by optimization and the comparison of observed with modeled data. A more detailed and complex representation of the hydrological processes within a model lead to a more complicated determination of the parameter values, especially if these are independent (*Nash and Sutcliffe*, 1970). Increased computer power allows the implementation of powerful optimization algorithms to a greater extent and consequently a rising amount of model parameters (*Gan et al.*, 1997). Another assumption of more enhanced calibration procedures is the reduction of parameter uncertainty as well as the strengthening of consistent model structures (*Seibert*, 2003). The more complex a model is, the more parameters are generally needed. The increase of the amount of model parameters does not necessarily involve more input data supporting the determination of the values (*Beven*, 2006). In this context, overparameterization of hydrological models due to a significant increase of the number of parameters belongs to one of the major problems (*Jakeman and Hornberger*, 1993; *Perrin et al.*, 2003). *Perrin et al.* (2001) identified that models with more parameters and thus degrees of freedom perform better in the calibration mode. In validation mode, this trend is not recognizable anymore and less parameterized models achieve similar model results.

2.2 Model Description

In this study, the HBV and $SUPERFLEX$ model are applied. Both models belong to the group of conceptual rainfall-runoff models and differ in complexity due to data requirement, amount of parameters and model structure. The original version of the HBV model was modified at the Institute for Modelling Hydraulic and Environmental Systems at the University of Stuttgart (IWS). The main distinction between the two models is their spatial representation of hydrological processes in which the HBV is a distributed model and the SUPERFLEX has a lumped structure. The HBV contains a fixed model structure whereas the SUPERFLEX is based on a flexible model framework. These two approaches are intended to provide insights into the diversity of catchment properties and support the choice between a fixed or a flexible model structure.

2.2.1 HBV Model

The HBV is a well-known hydrological model which has been used in various studies and operational applications worldwide for a long time. Thus, it is a continuously enhanced modeling framework and well documented in many scientific publications and technical reports. The first version of the HBV as a semi-distributed conceptual model was presented by *Bergström* in 1972. It was developed for hydrological forecasting of Swedish drainage basins at the Swedish Meteorological and Hydrological Institute ($SMHI$). The basic skills included snow accumulation and melt, soil moisture accounting and runoff response (*Bergström*, 1992). A noteworthy development of the HBV was carried out by *Lindström et al.* (1997). A few drawbacks regarding areal representation and physical inconsistencies, such as neglecting of interception and elevation dependent evapotranspiration, were resolved. In this

study, a distributed version of the HBV is used (Figure 2.1). This model structure was stepwise developed at the IWS and is named *HBV-IWS* in the following. The basic architecture of all following versions contains conceptual routines for the calculation of snow accumulation and melt, soil moisture and runoff concentration. Snow processes are accounted for using a degree-day approach and a fast-melt function in the case of rainfall into the snow pack. A detailed description of the snow routine follows in Section 5.2.

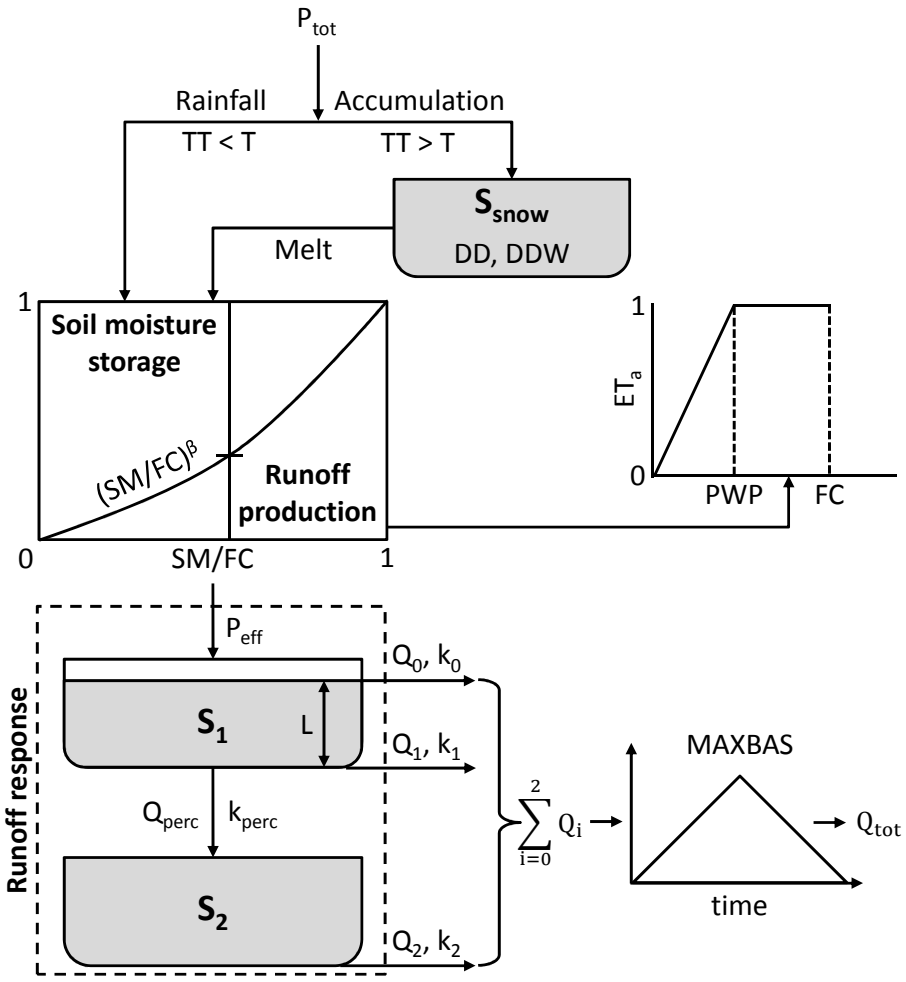


Figure 2.1: Scheme of the HBV-IWS model. S_1 and S_2 are the water levels of the corresponding upper respectively lower reservoir.

The first modifications of the original HBV was the partition of the river basin in different subcatchments. Inside the subcatchments a classification into zones regarding different properties like elevation, soil and land use is feasible (*Hundecha and Bárdossy*, 2004). *Götzinger and Bárdossy* (2007) modified the HBV-IWS model to a grid-based version. There, the input data (precipitation and temperature) and catchment properties (soil and land use) are considered for each single grid cell. Furthermore, the calculation of hydrological processes such as evapotranspiration, snow accumulation and melt, soil moisture and runoff concentration are performed for each grid cell. The runoff response still has a conceptual

structure with reservoirs for direct discharge and base flow. A further modification introduced by *Götzinger and Bárdossy* (2007) is the restructuring of the two reservoirs, where water can drain through percolation from the upper soil reservoir to the groundwater system. The production of direct runoff remains equal to the original version. The current version features another change in the calculation of the total discharge. The upper reservoir includes another outlet for representing the inter flow in the sub-surface layer (*Bárdossy and Das*, 2008; *Bárdossy and Singh*, 2008). A triangular weighting function with one free parameter *MAXBAS* transforms the direct discharge (near surface and inter flow) and base flow into the total discharge. Interconnected nodes represent the river network and the generated flows are calculated using the Muskingum flood routing method. An overview about different model structures of the HBV-IWS due to the spatial representation is given in *Das et al.* (2008).

The current version of the HBV-IWS requires numerous information a priori. In addition to the observation data discharge (Q), precipitation (P) and mean temperature (T_m), as well as information about topography (coordinates of grid cells and corresponding elevation), land use classes and soil properties are necessary. The modeling time step and temporal data resolution is daily in this study. The calculation of soil moisture (*SM*) requires input information of field capacity (*FC*) and permanent wilking point (*PWP*). Using these parameters the soil moisture is balanced by precipitation and evapotranspiration (*Bárdossy and Singh*, 2008) which is described by the following equations:

$$P_{eff} = (SM/FC)^{\beta} \cdot P + S_{melt}, \tag{2.1}$$

where P is the daily precipitation depth, P_{eff} the effective precipitation, S_{melt} the amount of snowmelt, SM the actual soil moisture, FC the maximum field capacity and β a model parameter considering the non-linear relationship between soil moisture and field capacity. The HBV-IWS provides three methods using different input information for calculating the potential evapotranspiration (PE):

- Method 1: Original linear approach (*Hundecha and Bárdossy*, 2004)

  $$PE = (1 + C_{ET} \cdot (T_m - T_M)) \cdot PE_M \tag{2.2}$$

  According to the daily mean temperature (T_m), the monthly long-term mean potential evapotranspiration (PE_M) is modified to the daily potential evapotranspiration (PE), where T_M is the monthly long-term mean temperature and C_{ET} a monthly factor accounting for temperature anomalies.

- Method 2 & 3: Hargreaves and Samani equation (*Hargreaves and Samani*, 1985)

  $$PE = 0.0023 \cdot R_A \cdot (T_{max} - T_{min})^{0.5} \cdot (T_m + 17.8), \tag{2.3}$$

  where R_A is the extraterrestrial radiation in $mm\,d^{-1}$, T_{max} the maximum and T_{min} the minimum daily temperature. The difference in method 2 and 3 is the calculation of R_A. Method 2 calculates R_A for a given latitude by considering the solar constant, solar declination and time of the year (*Allen et al.*, 1998). Method 3 uses the yearly variation of R_A for certain latitudes in the northern hemisphere given in *Iqbal* (1983).

According to these values and the latitude of the catchment area, method 3 assumes a monthly value of R_A as evapotranspiration equivalent in $mm\,d^{-1}$ in advance. These values are passed to the model as input information. Finally, the calculated PE of method 2 or 3 is corrected by different monthly crop coefficients (K_c).

In the case of reduced water availability the actual evapotranspiration ET_A is considered instead of the potential evapotranspiration PE. Here, the *PWP* is the limiting parameter if ET_A or PE take place.

$$ET_A = (SM/PWP) \cdot PE \tag{2.4}$$

Table 2.1: Model parameters and ranges of the HBV-IWS.

Parameter	Description	Unit	Min	Max
DD	Degree-day factor	$[mm\,d^{-1}\,°C^{-1}]$	0.5	5
DD_w	Rain-on-snow weight	$[d^{-1}\,°C^{-1}]$	0	2
k_0	Storage coefficient for direct runoff	$[d]$	0.1	2.5
k_1	Storage coefficient for inter flow	$[d]$	5	50
k_2	Storage coefficient for base flow	$[d]$	10	1000
k_{perc}	Storage coefficient for percolation	$[d]$	20	100
L	Threshold water level in upper reservoir	$[mm]$	1	50
T_0	Threshold temperature for snow/rain	$[°C]$	-2	2
β	Shape parameter for soil moisture routine	$[-]$	1	6

In total, 15 parameters describe the HBV-IWS model. The nine parameters shown in Table 2.1 are optimized through calibration, whereby the range can be assumed as catchment independent. Other parameter values and ranges used for calibration like field capacity (*FC*), permanent wilking point (*PWP*), model parameter C_{ET} and *MAXBAS* vary for different catchments. Two more parameters are necessary if a river network is considered. In this case, the flood routing is performed by the Muskingum method and includes a weighting factor x and a retention constant k. The optimization of these parameters is done with an automatic calibration using simulated annealing (*Aarts and Korst*, 1990). As objective function the Nash-Sutcliffe efficiency (NSE) is used (*Nash and Sutcliffe*, 1970). The maximization of the objective function is performed for different aggregation time steps (*Götzinger and Bárdossy*, 2007).

2.2.2 SUPERFLEX Framework

The SUPERFLEX framework is a conceptual hydrological model which deals with physically reasonable representations of catchment dynamics, moderate parameterization and computational efficiency (*Fenicia et al.*, 2011). The general model structure is built modular. It contains individual reservoirs representing different hydrological processes such as interception, groundwater storage, slow and fast runoff responses. The numerical framework

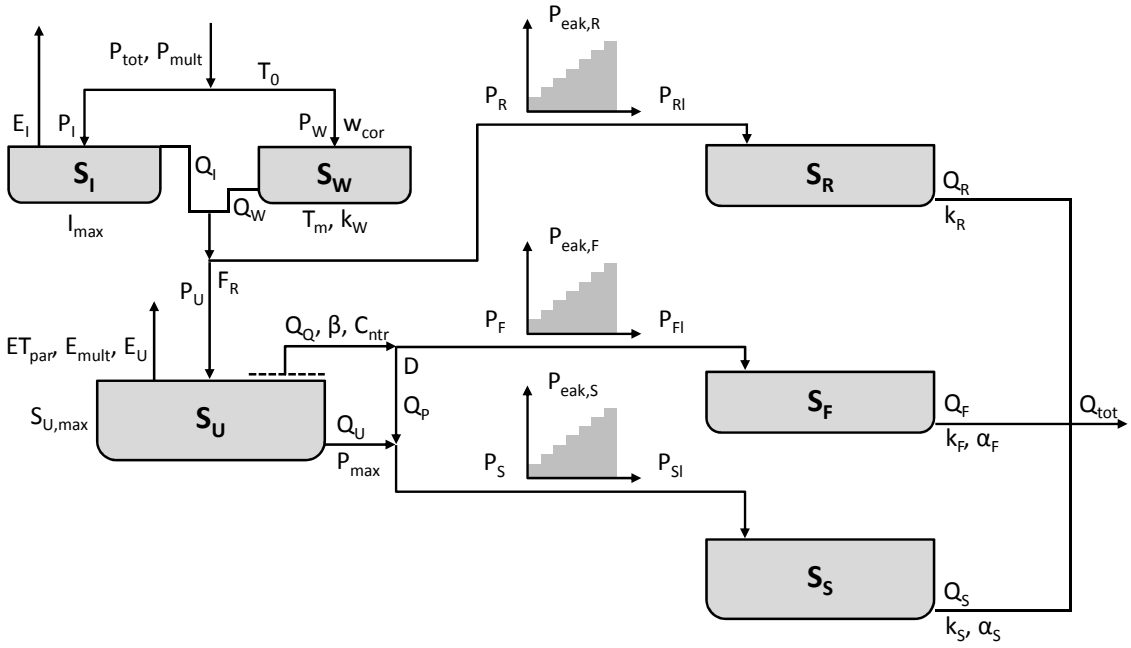


Figure 2.2: Scheme of the SUPERFLEX modeling framework. The reservoir index S_n refers to interception (I), snow (W), unsaturated soil (U), riparian zone (R), fast outflow (F) and slow outflow (S).

consists of robust techniques (*Fenicia et al.*, 2011). According to *Fenicia et al.* (2011) the basic components of the modeling framework are:

- reservoir elements representing water storage and release processes,
- lag functions for the transmission and delay of flows,
- junction elements for dividing, connecting and rescaling flows.

One advantage of this modeling framework is its flexible ability for arranging various reservoir configurations and applying variable lag functions. Thus, the model can be adapted to individual catchment properties and tested for different discharge characteristics. Another skill of this modular concept is the experimental learning of the driving hydrological processes within a catchment. Enabling and disabling of certain reservoirs present their influences and significance on the hydrological response of a catchment (*Kavetski and Fenicia*, 2011). The SUPERFLEX framework has a lumped model structure. In its full configuration, which includes a snow model routine, the model requires data on precipitation, discharge, mean temperature and potential evapotranspiration. All input and output variables are expressed in mm. The discharge rate is calculated by dividing the water volume by the catchment area and finally converting to the discharge depth. Precipitation input is represented by the mean areal precipitation for each time step. The mean values are gained from spatial precipitation fields interpolated by different methods, which are described in Section 4.3. The mean areal temperature values are processed in the same way. The areal values of the potential evapotranspiration (PE) are calculated either by the required mean areal temperature values or by prior interpolation.

The SUPERFLEX model itself can vary considerably in its complexity in terms of the reservoir structures, lag-functions and parameter sets. The model framework has the capability of easily modifying and simplifying its structure. The SUPERFLEX model incorporates six interconnected reservoir modules in its complete set-up (Figure 2.2), which are briefly explained in the following:

- **Interception reservoir (S_I)**
  The interception and snow reservoir are responsible for the generation of effective precipitation (P_{eff}). Depending on the reservoir configuration, the resulting amount of P_{eff} is splitted to the subsequent reservoirs S_U, S_R, S_F and/or S_S. S_I is conceptually an overfall reservoir with a maximum storage capacity. The interception (I) is generated by the potential evapotranspiration (PE) transferred from the input data and the available amount of water in the reservoir (*Gharari et al.*, 2014).

- **Snow reservoir (S_W)**
  The regulation of snow accumulation and melt is done by this reservoir. The snow processes are calculated by the degree-day approach (see Section 5.2). This reservoir allows some modifications due to the typically linear relationship between snowmelt and temperature. *Kavetski et al.* (2006a) implemented a smoother transition from rain to snow and vice versa. As well, they modified the melting process to a smoother behavior.

- **Unsaturated soil reservoir (S_U)**
  The reservoir of the unsaturated zone contains a ground (percolation) and an overfall (direct runoff) outlet. S_U describes the maximum soil moisture capacity in the root zone. Depending on the soil moisture conditions, the entering P_{eff} is completely stored in S_U and successively delivered to the slow reservoir (S_S) through percolation or parts of P_{eff} are directly routed to the fast reservoir (S_F). Another reduction of the soil moisture is considered by plant transpiration (*Gharari et al.*, 2014).

- **Fast reacting reservoir (S_F)**
  The direct runoff of the catchment is considered in the fast reacting reservoir (S_F). S_F is a linear reservoir controlled by the storage coefficient k_F (*Hrachowitz et al.*, 2014; *Gharari et al.*, 2014).

- **Slow reacting reservoir (S_S)**
  This reservoir calculates the slow portion of the total outflow and represents the groundwater dynamics of a catchment. S_S is also a linear reservoir controlled by the storage coefficient k_S (*Hrachowitz et al.*, 2014; *Gharari et al.*, 2014).

- **Riparian reservoir (S_R)**
  The riparian reservoir accounts for flow contributions from hillslopes and riparian zones which can also generate faster outflows (*Hrachowitz et al.*, 2014).

The model can be calibrated by two different regression schemes. These are the standard (SLS) and weighted (WLS) least square methods (*Kavetski and Fenicia*, 2011). Both methods differ in the assumption of the errors. The WLS error model assumes zero-mean Gaussian errors and accounts for heterogeneity of variance. In contrast, the error model based on

Table 2.2: Most relevant model parameters and descriptions of the SUPERFLEX modeling framework.

Parameter	Description	Unit
C_{ntr}	Preferential recharge coefficient	$[-]$
D	Partition coefficient from fast to slow reservoir	$[-]$
E_I	Interception	$[mm\,t^{-1}]$
E_{mult}	Evaporation multiplication factor	$[-]$
E_U	Transpiration from unsaturated reservoir	$[mm\,t^{-1}]$
ET_{par}	Transpiration threshold	$[-]$
F_R	Proportion of wetlands in the catchment	$[-]$
I_{max}	Maximum interception threshold	$[mm]$
k_n	Storage coefficient of reservoir n	$[t^{-1}]$
k_W	Degree-time factor for snowmelt	$[mm\,t^{-1}\,{}^\circ C^{-1}]$
$P_{eak,n}$	Time lag of transfer function for reservoir n	$[t]$
P_{max}	Maximum percolation capacity	$[mm\,t^{-1}]$
P_{mult}	Precipitation multiplication factor	$[-]$
P_n	Recharge of reservoir n	$[mm\,t^{-1}]$
P_{tot}	Total precipitation	$[mm\,t^{-1}]$
Q_n	Percolation/runoff from reservoir n	$[mm\,t^{-1}]$
Q_{tot}	Total runoff	$[mm\,t^{-1}]$
$S_{U,max}$	Maximum unsaturated storage capacity	$[mm]$
T_m	Snowmelt temperature	$[^\circ C]$
T_0	Threshold temperature for snow/rain	$[^\circ C]$
w_{cor}	Snow correction factor	$[-]$
α_n	Shape parameter for runoff generation of reservoir n	$[-]$
β	Shape parameter for C_{ntr}	$[-]$

the SLS regression scheme considers a constant variance of the residual errors (*Kavetski and Fenicia*, 2011). The individual parameters can be set to *active* or *fixed* in advance of the calibration, whereby only the *active* parameters are optimized. The most important parameters and their description are shown in Figure 2.2 and Table 2.2. The current version also includes some smoothing functions for the calculation of interception and snow processes which parameters are not explicitly illustrated. The model evaluation is performed by several objective functions simultaneously (*Fenicia et al.*, 2008). This multi-objective optimization method includes the Nash-Sutcliffe efficiency (NSE), coefficient of determination (R^2) and the root mean squared error ($RMSE$) in the applied model version. Such an approach allows a better constrain of parameter values (*Fenicia et al.*, 2008). The SUPERFLEX model uses a multistart quasi-Newton optimizer (*Nocedal and Wright*, 2006) and 100 random starting seeds (*Kavetski and Clark*, 2010). The posterior distributions of the predictions are investigated numerically

using Markov chain Monte Carlo ($MCMC$) methods. The uncertainty due to input and model structure is examined with the Bayesian total error analysis ($BATEA$). The theoretical background and numerical realization of the $BATEA$ parameter estimation equations are given in *Kavetski et al.* (2002) and *Kavetski et al.* (2006b). The model provides multipliers for precipitation and evaporation. This is a convenient option for the parameterization of input uncertainty (*Kavetski et al.*, 2006b). A more detailed description of the model numerics, optimization functions and uncertainty estimation is given in *Kavetski et al.* (2006b,c); *Fenicia et al.* (2008); *Kavetski and Clark* (2010); *Kavetski and Fenicia* (2011).

In this study, four different reservoir configurations are investigated. The different configurations and the labeling of the individual reservoirs are shown in Subsection 5.1.2. The complexity of the configurations is increased stepwise and the calibration is performed with both regression schemes. The model runs are executed with ten iteration steps.

2.3 Model Evaluation

All model results have to be comprehensively evaluated to check their reliability and ensure their comparability in terms of model performance, uncertainty and realism (*Wagener*, 2003). The confidence of reasonable model use and verifiable predictions is based on the agreement between the model and reality (*Gupta et al.*, 2008). An appropriate and well-selected variety of indicators specifies the model performance. Also, potential sources of uncertainty should be emphasized (*Schaefli and Gupta*, 2007). In this section, the applied methods of model evaluation are described. Firstly, it is discussed how the performance of a model can be estimated and which measures are available. Secondly, different approaches of model calibration are introduced. The calibration of a conceptual model is mandatory to obtain reliable and efficient parameterization (*Brath et al.*, 2004). The function of calibration is fitting the model output to the observations by adjusting the model parameters. The agreement between observations and model output is usually specified by some statistical criteria which allow an assessment of the model performance. Nevertheless, the indication of the model performance solely is insufficient. A comprehensive evaluation requires additional information because of its dependence on the quality of observed data, the area of application and modeling time period (*Bergström*, 1991). After the calibration, the model needs to be validated. The process of model validation examines if the model has a satisfactory range of accuracy within its scope of application (*Refsgaard and Henriksen*, 2004). In other words, the model calculations are performed by the optimized parameter set determined during calibration and the required input variables. Afterwards, the level of agreement between the observations which are the independent data and modeled output is investigated by performance measures. The evaluation of a model can be performed by statistical and graphical measures. A visual inspection of model results should always be included in an evaluation procedure. This kind of subjective model evaluation allows the control of systematic (e.g., under- and overestimation) and dynamic (timing, rise and decline of peak flows) model response (*Krause et al.*, 2005). *Willmott et al.* (1985) has already defined that a comprehensive procedure of evaluating the model performance has to include data plots, summary of univariate statistics, difference measures, sensitivity studies and assessment of reliability. A

sensitivity analysis provides a possibility to determine the influence of changed model inputs or parameters to the model output (*Oudin et al.*, 2006; *Moriasi et al.*, 2007; *Fenicia et al.*, 2008). A single statistical measure is not sufficient to evaluate the performance and particularly the reliability of a model. High values for the model performance can be achieved despite comparative poor model results. Also, the reverse case can occur where a poor index value suggest a rejection of reasonable model results (*Jain and Sudheer*, 2008).

In this context, the influence of different enhanced calibration periods and conditions is investigated. The interpretation of model *validation* is sometimes different and misleading. Within this study, *validation* defines the process of testing parameter sets identified by calibration against an independent data set (e.g., different time period). The model and its obtained results are accepted as valid if they fulfill certain performance criteria (*Refsgaard and Knudsen*, 1996). The independent data set for model validation should have similar properties like the data set used for model calibration (*Klemeš*, 1986). The reason is that it is assumed that the optimized parameter set is only able to capture *known* conditions within a certain range. The applied performance criteria are described in the subsequent section. The analysis and assessment of model validation is presented in Chapter 5 and 6.

2.3.1 Model Performance

The model performance specifies the level of acceptable match between reality (observations) and model output (*Wagener*, 2003). *Krause et al.* (2005) quoted three reasons for the necessity to evaluate model performance:

1. A quantitative assessment of model capability to simulate past, present and future catchment behavior
2. Evaluating the effect of modeling changes due to parameter adjustment, model structural modifications, involvement of additional information, and the spatial and temporal consideration of catchment characteristics
3. The comparison of present modeling achievements with former studies

Gupta et al. (2008, 2012) also described an urgent necessity for sophisticated and standardized schemes for model evaluation due to greater interaction of model components and thus more complex model structures. They argue that a model evaluation should be diagnostic in terms of illustrating if model modifications aim for more realistic representations of the real system and how a model should be updated. A performance evaluation implies the model calibration and validation (*Refsgaard and Henriksen*, 2004). The first step of evaluating the model performance is a comparative data description and analysis with data plots. Subsequently, the evaluation with different univariate measures such as mean and standard deviation of observed and modeled data takes place (*Willmott*, 1981). Then, the model performance is described by different, more sophisticated performance measures. The level of model performance can be affected by different factors. First of all, the quantity and quality of the required input data have an essential influence. Then, model type, structure and complexity can cause errors and reduce the power to explain the hydrological system. In

this context, a proper interpretation of the explanatory power of the performance measures is essential (*Schaefli and Gupta*, 2007). In general, the performance criteria can be classified in absolute (e.g., $RMSE$) and relative (e.g., NSE) criteria. An adequate assessment of model performance should always contain several absolute and relative measures (*Legates and McCabe*, 1999; *Fenicia et al.*, 2008; *Ritter and Muñoz-Carpena*, 2013). Several efficiency criteria used in this study imply a summation of the error term normalized by a measure of the variability in the observed input data. The summation considers the absolute or squared errors to exclude the compensation of errors with differing sign. Hence, larger errors receive a higher relevance whereas smaller errors tend to be neglected. The errors associated with high flow values are usually larger than the ones with low flow values. Consequently, using this kind of efficiency criteria as objective function for calibration procedures can cause a higher focus on fitting the higher portions of the hydrograph than on the lower portions (*Krause et al.*, 2005).

The following list describes graphical methods and performance measures used here for model evaluation. In this regard, the index "*obs*" denotes observed and "*mod*" modeled data. The index "*n*" is the total number of observations and "*i*" the time step.

Data visualization for examining model adequacy and performance is realized for instance with scatterplots, histograms, hydrographs and flow duration curves (*Refsgaard and Knudsen*, 1996). Such graphical analyses show to which degree observed and modeled data agree with each other. One can review if the model performs consistently over the total time period (calibration and validation range) or depends on certain discharge characteristics. Also, other features such as unexpected relationships between observed and modeled data sets, existence of outliers and considerable model bias are identifiable (*Ritter and Muñoz-Carpena*, 2013).

A scatterplot illustrates the relationship between two variables (e.g., observed versus modeled discharge). Such a relative simple representation of comparable data sets can already give a first hint for erroneous input data, model structure or parameterization. It is possible to detect consistent under- or overestimations of the dependent variables. Weak sections of the modeled output like certain discharge conditions can also be distinguished. A constant underestimation of high discharges refers to such an issue.

Hydrographs compare the time series of observed and modeled discharge. Thereby, model bias as well as differences in magnitudes can be detected. The difference to scatterplots is the ability to identify temporal discrepancies (timing) between observed and modeled data (*Moriasi et al.*, 2007).

Another application of plots is the comparison of observed and modeled flow duration curves. They display the general match of the discharge characteristics or rather frequencies over a selected time period. The difference to scatterplots or hydrographs is that the data pairs are not compared time-dependently, but according to their non-exceedance probability.

Coefficient of determination (R^2)

$$R^2 = \frac{\left(\sum_{i=1}^{n}\left(Q_{obs,i} - \overline{Q}_{obs}\right) \cdot \left(Q_{mod,i} - \overline{Q}_{mod}\right)\right)^2}{\sum_{i=1}^{n}\left(Q_{obs,i} - \overline{Q}_{obs}\right)^2 \cdot \sum_{i=1}^{n}\left(Q_{mod,i} - \overline{Q}_{mod}\right)^2} \tag{2.5}$$

R^2 specifies the proportion of the variance in observed data explained by the model. The range of R^2 is from 0 to 1 in which a value of zero signifies no correlation at all. Lower error variance exists by higher values of R^2 (*Moriasi et al.*, 2007). R^2 is a correlation-based measure. On the one hand, one drawback of such measures is their hypersensitivity to extreme values. On the other hand, they are insensitive to additive and proportional differences between the observed and modeled variable. One can be misled to interpret model results as good due to high values of R^2, even they are not (*Legates and McCabe*, 1999). Model results with almost no dispersion between observed and modeled data have values close to 1 and indicate very good model fits. But, even with high values of R^2 the model predictions can be wrong due to systematic under- and overestimations (*Krause et al.*, 2005). Thus, R^2 is not suitable as single performance measure for model evaluation.

Root mean squared error ($RMSE$)

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n}\left(Q_{obs,i} - Q_{mod,i}\right)^2} \tag{2.6}$$

The $RMSE$ describes the standard deviation of the model prediction error and thus the average error produced by the model. The range is from 0 to ∞. The unit is the same as the input variable. Lower values specify a better model performance (*Gupta et al.*, 1999). The missing clarification of error source or type is one disadvantage of the $RMSE$. Neither the proportion of systematic and unsystematic error nor the type of systematic error (additive or proportional) can be determined (*Willmott*, 1981).

Percent bias ($PBIAS$)

$$PBIAS = \frac{\sum_{i=1}^{n}\left(Q_{obs,i} - Q_{mod,i}\right)}{\sum_{i=1}^{n} Q_{obs,i}} \cdot 100 \tag{2.7}$$

$PBIAS$ measures the mean tendency of the modeled discharge to be larger or smaller than the observed correspondent in %. The range is between $-\infty$ and ∞. The optimal value is zero. Positive values indicate a model bias toward underestimation and negative values toward overestimation (*Gupta et al.*, 1999). *Moriasi et al.* (2007) reported satisfactory model performance if the $PBIAS$ is within a range of $\pm 25\,\%$. Since efficiency indices similar to the Pearson correlation, e.g., NSE can tend to low values due to model bias generated by the calibration, a measure of bias should be considered for a comprehensive model evaluation. Such a model bias can result from differences in magnitude or time offset for time-dependent models (*McCuen et al.*, 2006).

Nash-Sutcliffe efficiency (NSE)

$$NSE = 1 - \frac{\sum_{i=1}^{n} \left(Q_{obs,i} - Q_{mod,i}\right)^2}{\sum_{i=1}^{n} \left(Q_{obs,i} - \overline{Q}_{obs}\right)^2} \tag{2.8}$$

NSE displays the relative magnitude of the residual variance to the variance of the observations (*Nash and Sutcliffe*, 1970; *Gupta et al.*, 1999). The range is between $-\infty$ and 1. A value of 1 indicates a perfect fit between observed and modeled data. For a value smaller than 0, the mean observed discharge is a better predictor than the model. At a daily simulation time step, *Singh et al.* (2005) accepted a value of $NSE \geq 0.75$ obtained in the calibration period as satisfactory. In contrast, *Moriasi et al.* (2007) defines values of $NSE > 0.50$ as satisfactory and values > 0.75 as very good. The ability for its use to many different model types is an advantage (*McCuen et al.*, 2006). An important shortcoming of the NSE is the fact that sometimes high values can be achieved with a poor model and in spite of a high value not all information of the input data is extracted. If the variance of the considered variable is very large, even relatively poor model fits can generate high values of NSE (*Legates and McCabe*, 1999; *Schaefli and Gupta*, 2007; *Jain and Sudheer*, 2008). Its insensitivity to systematic model under- or overestimations, especially during low flow periods, is another disadvantage which should be considered (*Krause et al.*, 2005; *Pushpalatha et al.*, 2012).

Modified Kling-Gupta efficiency (KGE_m)

$$KGE_m = 1 - \beta_w \cdot \left[\frac{\sum_{i=1}^{n} \left(Q_{obs,i} - Q_{mod,i}\right)}{\sum_{i=1}^{n} Q_{obs,i}}\right]^2 - \left[1 - r\left(Q_{obs,i}, Q_{mod,i}\right)\right]^2 \tag{2.9}$$

Gupta et al. (2009) introduced an alternative model performance measure. The Kling-Gupta efficiency splits the three components of the NSE (linear correlation, bias and discharge variability) in its individual parts. The equation of the KGE weights all three components equally by computing their Euclidean distance from the ideal point. The optimum value of KGE is at unity. The applied version of the KGE is modified by *Bárdossy et al.* (2016). In this version, $r\left(Q_{obs,i}, Q_{mod,i}\right)$ describes the correlation coefficient between observed and modeled discharge. The first term controls if the modeled discharge fits the water balance of the observations and can be stronger weighted with the parameter β_w. In *Bárdossy et al.* (2016) a value of $\beta_w = 5$ is selected.

In addition to the model performance measures listed above, the following measures are sometimes applied to supplement a deeper investigation of flood events:

Relative volume error (EV)

$$EV = \frac{V_{obs} - V_{mod}}{V_{obs}}, \tag{2.10}$$

where V_{obs} and V_{mod} are the total volumes of the observed and modeled hydrographs of the particular evaluation period (*Brath et al.*, 2004). Values of EV close to zero indi-

cate similar observed and modeled volumes. A negative value of EV refers to higher modeled than observed volumes and an overestimation of the volume for the evaluation period. A positive value means that the volume of the modeled time series underestimates the observed volume.

The EV refers to relative values and, like the $PBIAS$, is another volume-based performance measure. The EV is listed here for completeness. It is sometimes used as an alternative to the $PBIAS$ in this study, particularly when evaluating peak flows. Some of the above presented efficiency criteria put a major effort on the reproduction of the flow dynamics. Thus, it is further recommended to quantify volume errors with, e.g., the EV or the $PBIAS$ for a complete model evaluation (*Krause et al.*, 2005).

Relative peak error (EP)

$$EP = 1 - \frac{\sum_{i=1}^{n} \left| Q_{obs,i}^{peak} - Q_{mod,i}^{peak} \right|}{\sum_{i=1}^{n} Q_{obs,i}^{peak}}, \tag{2.11}$$

where $Q_{obs,i}^{peak}$ and $Q_{mod,i}^{peak}$ are the observed and modeled peak flows. EP takes the absolute difference between observed and modeled peak flows for all n peaks (*Seibert*, 2003). In the case of multiple peaks within one event, the maximum observed and modeled peak are considered. The determination of peaks can be done in different ways. For instance, a peak is selected if it exceeds the long-term mean discharge by three times (*Seibert*, 2003). The range is between 0 and 1, while a value closer to 1 indicates a better match between modeled and observed peaks.

Time to peak error (ET)

$$ET = T(Q_{mod}^{peak}) - T(Q_{obs}^{peak}), \tag{2.12}$$

where $T(Q_{mod}^{peak})$ and $T(Q_{obs}^{peak})$ are the corresponding times to peak. The evaluation of the time to peak error takes the difference in time between the closest observed and modeled peak into account (*Brath et al.*, 2004). A value of zero indicates that observed and modeled peak occur at the same time. A positive value of ET represents a time-delayed modeled peak flow.

In summary, a combination of different statistical measures and a visual inspection of the observed against the modeled data should be performed for a comprehensive model evaluation. Many authors advise that the model evaluation by solely one performance measure should be avoided (*McCuen et al.*, 2006; *Fenicia et al.*, 2008; *Jain and Sudheer*, 2008; *Ritter and Muñoz-Carpena*, 2013). This prevents a user and decision maker drawing conclusions from mistakenly well performing models.

2.3.2 Split-sampling Methods

Hydrological models have to be evaluated to their reliability if they should be used for practical applications like flood predictions. Traditionally, models are tested with a simple split-sampling method, meaning that they are calibrated with observed discharge data for a certain time period and then validated on another independent time period (*Klemeš*, 1986; *Refsgaard et al.*, 1995; *Seibert*, 2003). A disadvantage of this test method is that, if conditions differ significantly from the calibration period, a robust check of model applicability for such unknown conditions is difficult because no reference data are available (*Xu*, 1999). An alternative to this procedure is the differential split-sample test. Here, a model is calibrated on certain conditions as very wet or dry periods as well as significant cold or hot periods. Then, the model is tested on the dissimilar counterpart (*Klemeš*, 1986; *Xu*, 1999). The evaluation of model performance on dissimilar or rather *unknown* hydrological conditions is a first approach to check the trustworthiness of predictions into the unknown (*Seibert*, 2003).

Different Time Periods
For sufficiently long time series, the data set is divided into sequences of equal length. Necessarily, the calibration time period has to be long enough for capturing wide information of the input data. A possible gain in information with longer data sets only results from data with additional input about the underlying system behavior (*Schoups et al.*, 2008). For the case of insufficiently long time series, the proportion of 70 % for calibration and 30 % for validation of the available data length is recommended (*Klemeš*, 1986). The reverse case is also considered which means the calibration and validation time sequences are swapped. Here, the model performance of both validation periods has to be similar and reasonable for an acceptable model output (*Klemeš*, 1986).

In this study, the different modeling periods for calibration and validation are defined by different starting points, such as calendar year or hydrological year.

Different Discharge Characteristics
Precipitation shows noticeable differences of seasonal variability (*Bárdossy and Das*, 2008). During the winter season, the extent of precipitation over a catchment is more uniform. In addition, snow accumulation and melting cause a temporal shift in the volume of water. In contrast, summer events often have convective characteristics with high intensities as well as high spatial variability. Thus, a model calibration and validation due to different seasonal flow conditions is advisable. In this study, to investigate different discharge characteristics, model calibration and validation is performed separately for hydrological summer and winter half-years. Furthermore, a model evaluation for different flow conditions such as snowmelt periods or peak flows is reasonable. This supports the identification of weak points representing the real system behavior.

In this study, all sampling methods consider a warm-up period of one year prior to the calibration period to avoid effects of inappropriate initial conditions in the various reservoirs. The warm-up period for the validation also considers the previous calibration period. This part of the model results is ignored for the graphical evaluation and calculation of the efficiency measures. The evaluation of the model performance is done for different time periods and discharge characteristics. Different time periods are the total time period, hydrological

summer and winter half-year, periods without major snow-driven processes and snowmelt periods. Different discharge characteristics are mainly indicated by significantly high flow conditions and so-called unusual events, as presented in the subsequent section.

2.3.3 Calibration on Unusual Events

Model calibration on certain sequences of the discharge time series which contain more information than others can be sufficient for the determination of good parameter sets in hydrological modeling. One advantage of this approach is that not the total time series is used and thus the amount of input data and the computation time of the model can be reduced significantly. An important issue in model calibration is the extraction of information contained in the observed values and their transfer into values for the model parameters. Due to measurement errors, the information content of the observations is imperfect (*Gupta and Sorooshian*, 1985). Thus, the selected time series of critical or unusual events requires enough information to activate all model functionalities and parameters.

Singh and Bárdossy (2012) presented a method of model calibration on hydrologically unusual events. The identification of critical sequences of hydrological data time series was done with the data depth function. In this study, the SUPERFLEX model is also calibrated on unusual events. The selection of data sequences with the half-space depth function is performed on different unusual hydrological conditions related to discharge and temperature. The next two subsections describe the concept of data depth and the identification of unusual events.

Definition of Data Depth
The concept of data depth is used for the determination of the center of a multivariate data set. This method was presented by *Tukey* (1974). The half-space depth $D_X(p)$ of a point p with respect to the finite set X in the d dimensional space $\Re^d$ is defined as the minimum number of points of the set X lying on one side of the hyperplane through the point p. The minimum is calculated over all possible hyperplanes (*Bárdossy and Singh*, 2008).

The formulation of the half-space depth $D_X(p)$ of the point p with respect to set X is:

$$D_X(p) = \min_{n_h} \left(min \left(|\{x \in X \, \langle n_h, x - p \rangle > 0\}| \right), \left(|\{x \in X \, \langle n_h, x - p \rangle < 0\}| \right) \right) \tag{2.13}$$

In this equation, $\langle x, y \rangle$ is the scalar product of the d dimensional vectors, and n_h is an arbitrary unit vector in the d dimensional space representing the normal vector of a selected hyperplane. If the point p is outside the convex hull of X then its depth is 0. Points on and near the boundary have a low depth and points further inside have higher depths (*Bárdossy and Singh*, 2008; *Singh and Bárdossy*, 2012). Amongst others, a detailed description of the depth function and algorithms for its computation is presented by *Bárdossy and Singh* (2008).

Identification of Unusual Events
Unusual events cannot necessarily be limited to extreme high and low flow conditions (*Singh and Bárdossy*, 2012). Hydrological catchment processes and dynamics are very complex and their responses often occur in very different time scales, depending on current

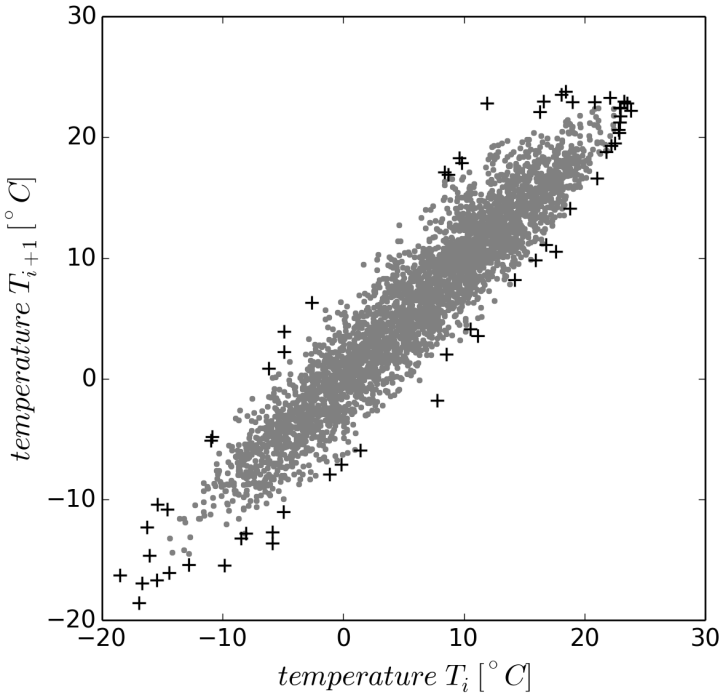


Figure 2.3: Two-dimensional data set X of daily mean temperature T_m. The plus markers indicate points with a depth $D_i \leq 5$.

discharge volume, precipitation intensity and duration, as well as soil moisture conditions due to antecedent water availability. Especially, in catchments where snow processes have a strong impact on discharge characteristics unusual temperature lapse rates over a certain time period can cause critical events. Accordingly, the detection of unusual events is done on the following criteria:

- Discharge of the total time series (Q_{tot})
- Discharge of the hydrological winter half-year (Q_w)
- Daily mean temperature (T_m)
- Antecedent temperature index (ATI)

The applied antecedent temperature index (ATI) is a parameter originating from a snow accumulation and ablation model (SNOW-17) developed by *Anderson* (1973). There, the ATI is used to describe the heat exchange between the surface layer of a snow pack and the air temperature. This index supports the identification of events with critical snowmelt potential:

$$ATI_i = ATI_{i-1} + TIPM \cdot (T_{m,i} - ATI_{i-1}) \tag{2.14}$$

The ATI at time step *i* considers the value of ATI at the previous time step *i-1*, the mean air temperature T_m at *i* and the antecedent temperature index parameter ($TIPM$) for its computation. The unit of the ATI is $^\circ C$. $TIPM$ is a weighting factor on temperatures of

previous time intervals and ranges between 0 and 1. Smaller values of $TIPM$ put more weight on previous time intervals. In this study, a value of 0.5 is used for $TIPM$ because *Anderson* (1973) has already found reasonable results for mountainous watersheds using this value.

The data set X passed to the depth function is built up of two consecutive time steps i (e.g., daily) defining the point p. It means that a data pair generated by discharge or temperature data include the value of the actual (X_i) and the subsequent (X_{i+1}) time step. The ATI considers the previous time step (X_{i-1}). Figure 2.3 shows such a two-dimensional data set X using the daily mean temperature T_m. For each time step, the data depth D_i of point p with respect to the total two-dimensional data set X is calculated. In *Singh and Bárdossy* (2012) a range of 1 to 5 for D_0 is proposed. In this study, a threshold depth D_0 of 5 is chosen. Thus, events with $D_i \leq D_0$ are assumed to be unusual.

3 Study Area and Database

The first part of this chapter provides a brief introduction to the selected hydrological catchment and its properties. It is shown why this study area is appropriate for the intended study objectives. In the second part of the chapter, the required data availability is presented.

3.1 Saalach Catchment

The *Saalach* catchment is located in the south-east of Bavaria in Germany and the province of Salzburg in Austria. The total catchment area is $1034\,km^2$. The river *Saalach* has a length of $103\,km$ and it is a tributary of the river *Salzach*. The upper part flows $70\,km$ through Austrian territory, then $19\,km$ through Germany until the river forms the border between Austria and Germany on a length of $14\,km$. The *Saalach* originates in the Kitzbühel Alps at an elevation of about $2000\,m.a.s.l.$ and flows into the *Salzach* at an elevation of $400\,m.a.s.l.$ northern of the city of Salzburg.

The upper river course of the *Saalach* on the Austrian side is mostly unaffected by structures in the watercourse. In the German part, shortly downstream of the gauge *Unterjettenberg* (Figure 3.1), a dam impounds the river to the lake *Saalachsee* and provides water to an hydroelectric power plant. From there on, the river is considerably anthropogenically influenced by straightening of the watercourse, embedded by dams and other hydraulic structures such as ramps and weirs. The basic information of the catchment and river characteristics is taken from the technical reports *Schaipp* (2002) and *Savora and Sackl* (2011).

3.1.1 General

The *Saalach* catchment is considered as one unit with an area of $929\,km^2$ until the gauge *Unterjettenberg* in this study. Figure 3.1 shows the topography of the region, the geographical extent of the investigated catchment and the course of the *Saalach*. It can be seen that the catchment is surrounded by high mountain ranges and the river flows to a large extent through steep valleys. It only flows about $15\,km$ through a relatively broad flat area where the flow direction turns from eastward to northward. The investigated upper river section has a length of $77\,km$ and the main part of the catchment area is located in Austria. Altitudes ranging from $420\,m.a.s.l.$ to $2600\,m.a.s.l.$ indicate a high topographic variability. The average elevation of $1260\,m.a.s.l.$ suggests a noteworthy influence of snow. The vegetation is mainly characterized by forest which covers more than a third of the area. The valleys are used agriculturally in which meadows and pastures predominate (*Savora and Sackl*, 2011).

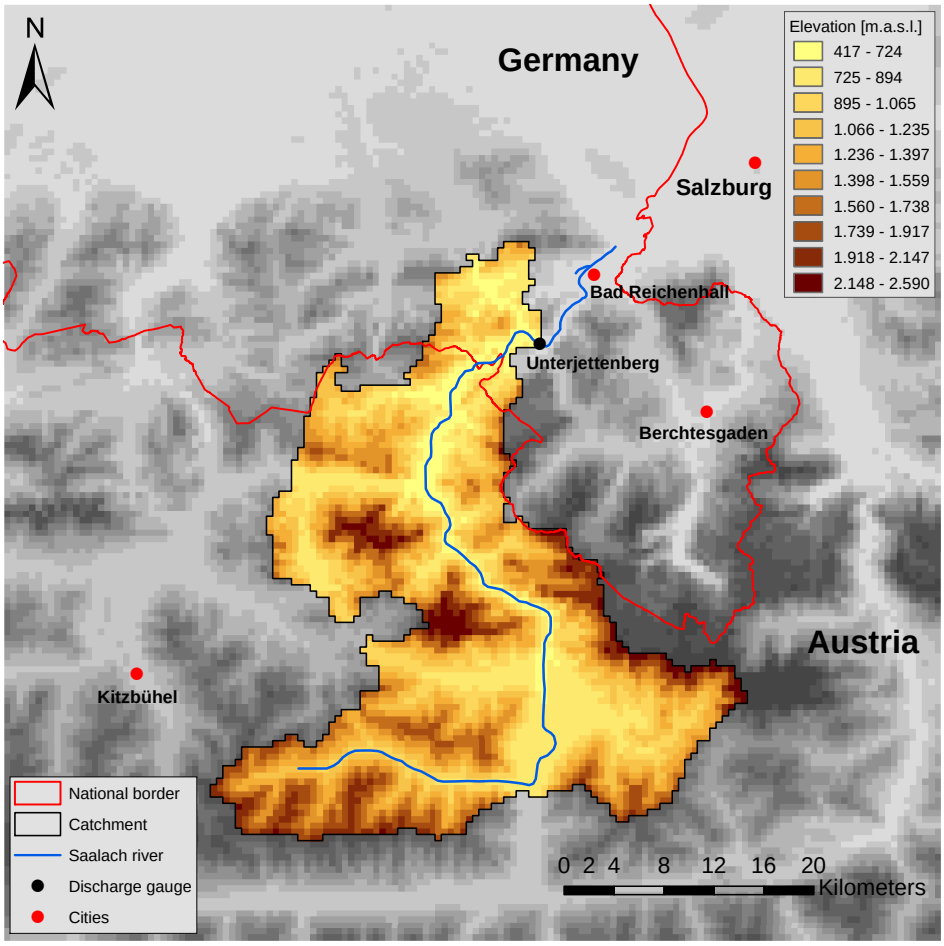


Figure 3.1: Topography of the study area and the upper *Saalach* catchment until the gauge *Unterjettenberg*.

3.1.2 Meteorology

The catchment area is located on the north side of the Alps and is therefore affected by orographic precipitation. Beside the Black Forest, the Alps and their foreland belong to the regions with the highest sums of mean annual precipitation in Germany as well as in Austria. Thus, this region is at risk of extreme precipitation events, especially if humid air masses from the north or north-east are blocked at the mountain ranges and are forced to release the moisture as rain.

The mean annual precipitation ranges from $1100\,mm$ in the lower regions and more than $2000\,mm$ in the highest parts of the catchment. Figure 3.2 shows the long-term monthly mean precipitation sums for four different rain gauge stations close to the catchment area. Table 3.1 gives some general station information. The annual distribution of precipitation shows a slight seasonality. Approximately $40\,\%$ ($650\,mm$) of the precipitation occurs within the hydrological winter half-year and $60\,\%$ ($950\,mm$) in the hydrological summer half-year. The monthly maximum precipitation amounts are observed from June to August ($170 - 230\,mm$). Months with minimum precipitation amounts are October ($85 - 110\,mm$)

and January/February ($85 - 115\,mm$). The mean annual temperatures vary between $4\,°C$ and $7\,°C$ in the low and middle zones and $0\,°C$ up to $3\,°C$ for the highest zones. In general, the minimum temperatures are measured in January and the maximum in July. The mean annual potential evapotranspiration ranges between $350\,mm$ for the upper parts of the catchment, and up to $550\,mm$ in the valleys.

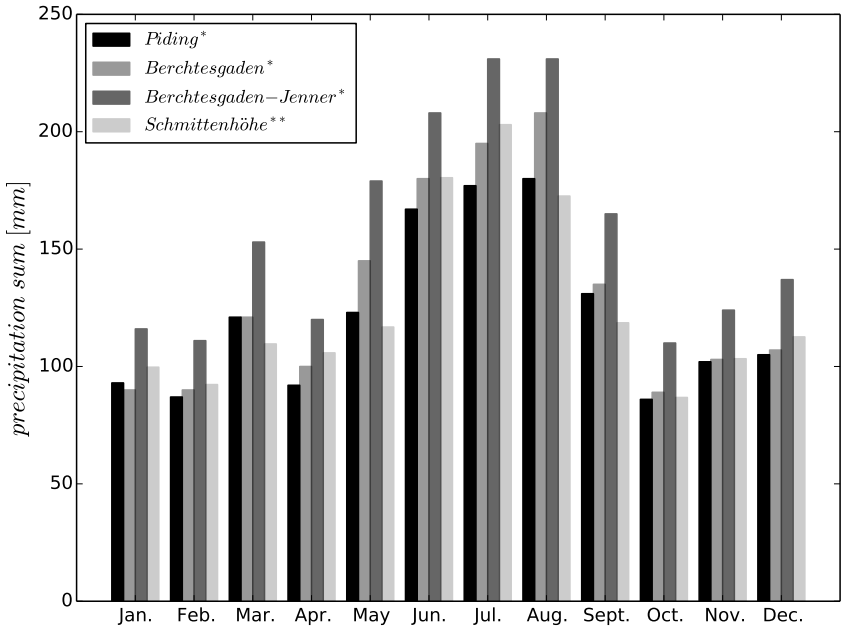


Figure 3.2: Long-term monthly mean precipitation (*1981 - 2010, **1971 - 2000) for different rain gauges in the study area.

Table 3.1: General information of some rain gauge stations in the study area.

station name	coordinates [m]		elevation	precipitation [mm]		
	E	N	[m.a.s.l.]	annual	winter	summer
Piding	4568115	5293082	458	1464	600	864
Berchtesgaden	4575742	5278355	600	1563	611	952
Berchtesgaden-Jenner	4576453	5274562	980	1885	761	1124
Schmittenhöhe	4555675	5243633	1973	1502	623	878

3.1.3 Hydrology

The water body type can be classified as alpine with cold water temperatures in summer and significant sediment transport. The *Saalach* has an high-alpine discharge character due to the altitude of the headwaters. Beside rainfall events, the discharge is regulated by snowmelt. With the beginning of the snowmelt period in March/April, the discharge increases until its maximum is reached during the summer season with high precipitation amounts. In au-

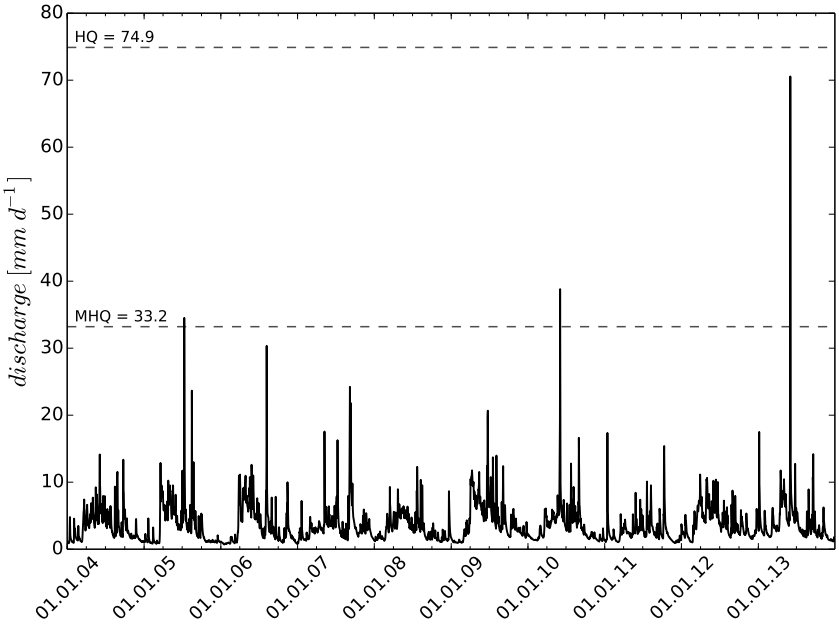


Figure 3.3: Daily discharge time series at gauge *Unterjettenberg* for the investigation period 2004 - 2013.

tumn, the flow rate starts decreasing and reaches its minimum during the subsequent winter months (Figure 3.3). The long-term mean discharge (MQ) at the gauge *Unterjettenberg* is $3.6\,mm\,d^{-1}$. For the reference period from 1901 until 2012, the mean flood discharge (MHQ) is $33.2\,mm\,d^{-1}$ and the highest observed flood discharge (HQ) is specified with $74.9\,mm\,d^{-1}$ on August 12, 2002.

3.2 Data Availability

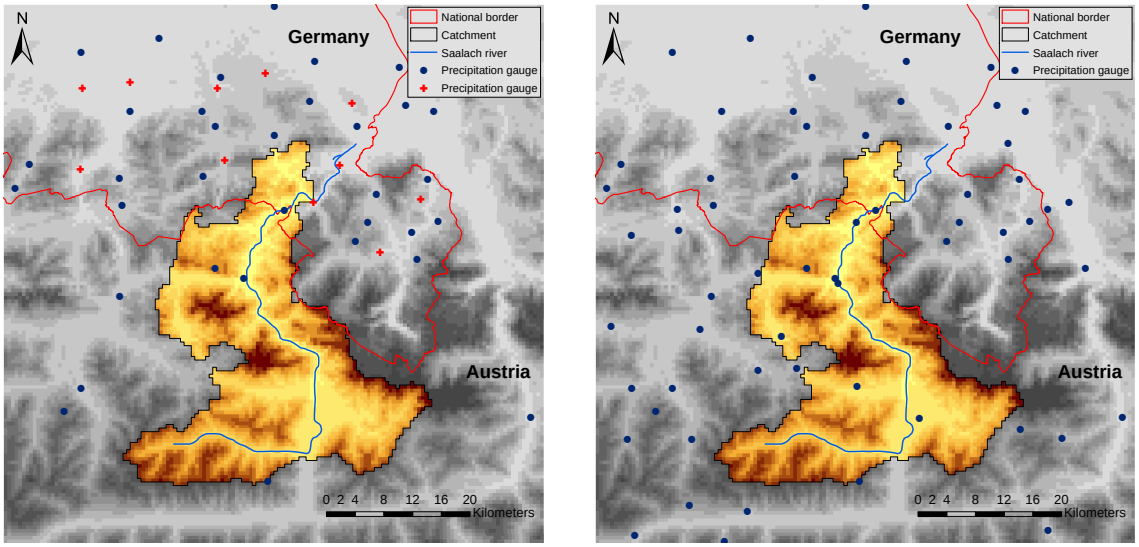
An adequate availability of data is essential and required for all hydrological modeling purposes. This section itemizes all information about the used data like source, time period, amount of gauging stations, spatial and temporal resolution. The total investigation time period is ten years and is set from the year 2004 until 2013. Although the available time series for daily precipitation and discharge data are significantly longer, only a time period of ten years is chosen. This provides a flexible data availability for the application at higher temporal resolutions, because hourly data prior to 2004 are very sparse.

Catchment Properties
The catchment is subdivided into regular grid cells with a spatial resolution of $500\,m$ by $500\,m$. All required input variables like precipitation (P), temperature (T) and potential evapotranspiration (PE) are processed on this grid. The underlying digital elevation model (DEM) is provided by the LfU, Bavaria. Spatial information on land use, soil, and field capacity is required for the HBV model. These data are not available in a digital grid-based

format. Thus, an underlying land use map is digitized manually with the information of BIS (LfU). The required soil classification and field capacities are derived from the HAD (BMU).

Meteorological Data

The precipitation and temperature data are mainly provided by the DWD and $ZAMG$. Supplementary observation data are obtained by publicly accessible online platforms of German (GKD) and Austrian ($eHYD$) state agencies. The data are not available to the same extent in daily and hourly resolution. For the spatial representation and interpolation of the meteorological data, a rectangular area with a spatial resolution of $500\,m$ by $500\,m$, and an extent of $150\,km$ by $150\,km$ with a total area of $22500\,km^2$ is used.



(a) Configuration *I* & *II* (b) Configuration *III*

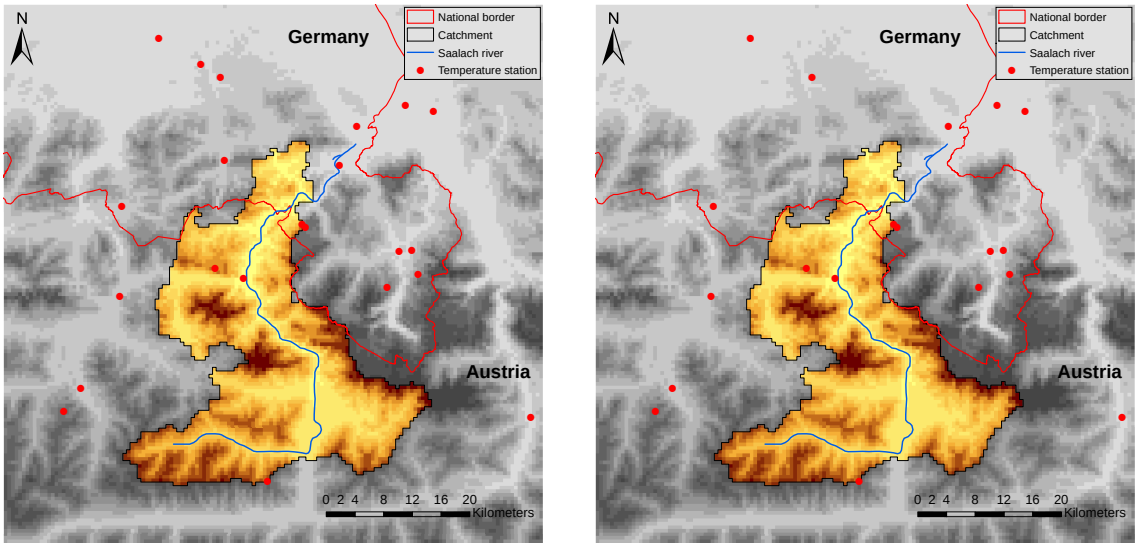
Figure 3.4: Distribution of different daily rain gauge configurations (The plus signs indicate the additional stations of configuration I.).

In this study, three configurations (*I* - *III*) of differently spatially distributed daily rain gauge stations are considered. These configurations vary in the number of stations and thus in the density capturing the study area. The spatial arrangement of the daily observation stations is shown in Figure 3.4 and further information is given in Table 3.2. It can be seen that configuration *II* has the least number of stations and configuration *III* the most. Originally, configuration *I* has more stations than *II* but configuration *I* contains several stations with incomplete time series which often just have data until the end of 2005. Therefore, the configuration *I* and *II* are almost identical after 2005. The density of stations within the catchment is 1.1 stations per $100\,km^2$ for configuration *III* which is more than twice as the other configurations. Although, the average minimum distance of two stations for configuration *I* and *III* is the same, the ranges of the absolute minimum distances vary considerably. For the configurations *I* and *II* the minimum distances of two stations ranges between $1.5\,km$ up to $28.4\,km$ whereas configuration *III* has minimum distances of $0.7\,km$ to $13.0\,km$. Especially, the high ranges of the minimum distance between two stations of configuration I and

Table 3.2: Basic information of the investigated daily rain gauge configurations (The order of operators is: DWD/ZAMG/eHYD).

description	daily rain gauge configuration		
	I	II	III
total number of stations	45	34	61
number per operator	36/9/-	25/9/-	25/9/27
station density of catchment [$1/100\,km^2$]	0.5	0.4	1.1
minimum distance of two stations [km]	1.5 - 28.4	2.9 - 28.4	0.7 - 13.0
average minimum distance [km]	5.5	6.9	5.6
average station altitude [$m.a.s.l.$]	716	748	767

II indicate a relative irregularly distributed and in some areas sparse gauging network. The minimum station altitude is $412\,m.a.s.l.$ and the maximum is $1973\,m.a.s.l.$ for all three configurations. There are 24 hourly rain gauge stations available (see Figure 5.9). The minimum distance of two hourly stations varies between $3.9\,km$ and $28.4\,km$. The average minimum distance between two stations is $8.7\,km$ and the density of rain gauge stations is 0.3 stations per $100\,km^2$ for the catchment area.



(a) Daily stations (b) Hourly stations

Figure 3.5: Daily and hourly configuration of available temperature stations.

The availability and distribution of the daily and hourly temperature stations is shown in Figure 3.5. The number of temperature stations is 22 for daily and 19 for hourly resolution. The minimal distance between two stations ranges from $0.6\,km$ to $28.4\,km$ and the average minimum distance between two stations is about $7.0\,km$. The station density for the catchment is $0.5/100\,km^2$. The lowest station is at $420\,m.a.s.l.$ and the highest at $1973\,m.a.s.l.$

Hydrological Data

The discharge data in daily and fifteen minute resolution for the gauge *Unterjettenberg* are provided by the LfU. The daily discharge time series for the investigation period from 2004 until 2013 is shown in Figure 3.3. In total, three flood events occurred exceeding the mean flood discharge (MHQ) of $33.2\,mm\,d^{-1}$ within this time period. One of the highest ever observed flood events, with a daily peak flow of $70.5\,mm\,d^{-1}$, happened on July 2, 2013. No noteworthy peak flow conditions were observed at all in 2004, 2008, and 2012. A sharp increase of the discharge caused by snowmelt in spring can be noticed in 2005, 2006 and 2009. An extensive low flow period for a couple of months occurred between the last quarter of 2005 and the first quarter of 2006.

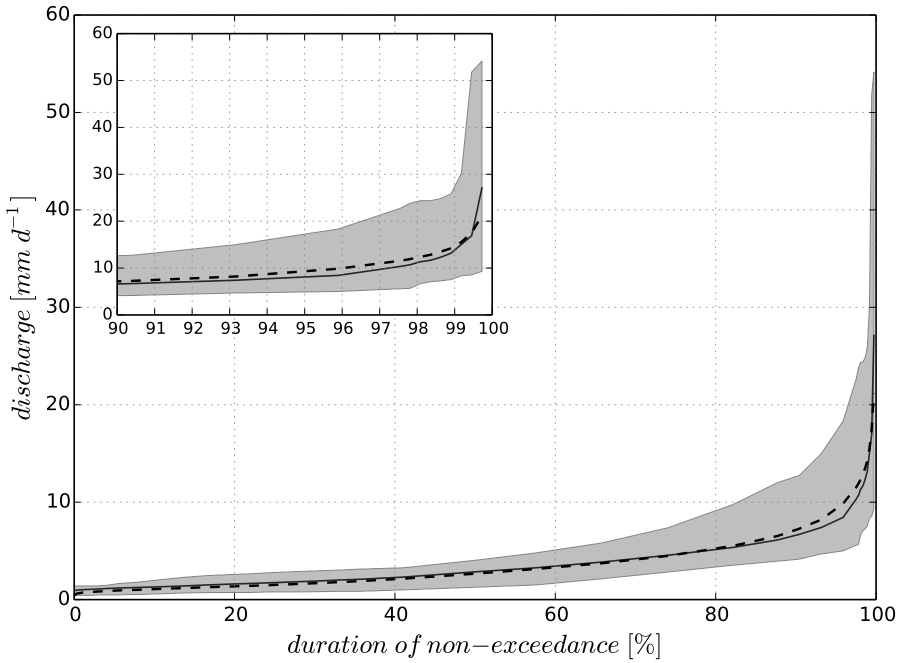


Figure 3.6: Flow duration curve of gauge *Unterjettenberg*. The solid line defines the mean duration of non-exceedance for the period 2004 - 2013, the dashed line indicates the long-term mean (1901 - 2006) and the gray shaded area represents the range of lower and upper hull curve.

The flow duration curve of the investigation time period and the long-term observations in Figure 3.6 match very well for low and middle flow conditions until a duration of non-exceedance of about $82\,\%$ (300 days). In the range of $82\,\%$ up to $99.5\,\%$ (363 days), the mean discharge of the investigation period is slightly below the long-term mean observations. For a duration of non-exceedance of $99.7\,\%$ (364 days), the mean discharge of the investigation time period is clearly above the long-term mean. This can be explained by the fact that within 2004 and 2013 several flood discharges were observed with higher flows than $21.2\,mm\,d^{-1}(99.7\,\%)$. In particular, the flood events in 2005, 2010, and 2013 significantly exceeded the long-term mean of $99.7\,\%$ duration of non-exceedance.

4 Data Validation and Processing

Every application of conceptual hydrological models requires a variety of input data for model calibration and evaluation. The data quality has a significant influence on the estimation of model parameters and thus on model performance. In the first section of this chapter, some aspects regarding the potential errors and resulting uncertainty of the input data are discussed. The second part of the chapter describes how data quality is analyzed and validated. Finally, the methods of processing the input data are introduced. Here, mainly meteorological point observations such as precipitation and temperature are converted to a spatially distributed representation. In a first step, conventional interpolation methods such as kriging are presented. Afterwards, a new conditional simulation method, called *random mixing* (RM) developed by *Bárdossy and Hörning* (2016a,b), is introduced. This method is versatile for many hydrological applications. Among other things, different constraints can be implemented to account for uncertainties due to measurement errors of input data.

4.1 Uncertainty of Input Data

It is widely accepted that measured data have inherently a degree of uncertainty, but this has so far often been disregarded in model evaluation procedures (*Moriasi et al.*, 2007). In this section, potential error sources of the input data are shortly pointed out. Primarily, the errors due to the measurement techniques and conditions are discussed. The amount of available data also has a strong influence on input uncertainty. The size of a data set is dependent on its spatial (lumped or distributed modeling approach) and temporal resolution, as well as on the length of the used time series. An increase from daily to hourly resolution has already a significant influence on data preprocessing and model computing time. For this reason, a manual control of the data quality is unfeasible and just simple automatic validation techniques can be executed to remove the most clear erroneous data values. The data maintained by public institutions should have a certain level of quality (*Andréassian et al.*, 2009). Nevertheless, a certain degree of uncertainty remains.

4.1.1 Meteorological Data

The available meteorological data used in this study are basically point measurements from weather stations. These data are subject to various error sources and thus to uncertainty. All input data are uncertain because of limited measurement procedures. Then, observations on a few points within a domain produce uncertainty due to the spatial variability of the meteorological variables. Consequently, a well established observation network with a

sufficient density and a preferably uniform distribution of monitoring stations is required for the estimation of areal precipitation (*McMillan et al.*, 2011). In regions with strong topographical variability, the locations of stations also have to cover the differences in altitude. This accounts for the vertical variation of precipitation due to orographic effects. In literature, a density of one rain gauge station per $40 - 50\,km^2$ is indicated as sufficient. Whereas, in mountainous regions, a much denser rain gauge network is required depending on the prevailing wind conditions (*Lecher et al.*, 2015). The required number of rain gauges can also depend on the temporal resolution of the measurements. In *Villarini et al.* (2008), it is reported that for the equal estimation of areal precipitation within a certain range of accuracy less stations for larger accumulation time steps, e.g., daily are necessary than for hourly or even higher temporal resolutions. Nevertheless, a network of observation stations can hardly specify the exact extension of precipitation fields or the amount of precipitation at ungauged locations between the stations. Thus, different geostatistical interpolation methods to convert point information to a spatial representation are used. Such a conversion of meteorological variables adds additional uncertainty to the input data (*Götzinger and Bárdossy*, 2008).

To shift back the focus on measurement uncertainty of precipitation, the commonly used measuring devices are briefly introduced. In Germany as well as in Austria, the Hellmann gauge types are generally used. The recording of precipitation is based on the principle of weighing or tipping buckets (*McMillan et al.*, 2011; *Lecher et al.*, 2015). Precipitation measurements are exposed to systematic and random errors due to limitations of the devices (*Sevruk*, 1996; *Yu et al.*, 1997; *Molini et al.*, 2001, 2005; *Sevruk and Chvíla*, 2005) and external influences (*Sevruk*, 1983; *Michelson*, 2004; *Chvíla et al.*, 2005; *Sevruk et al.*, 2009). The main reason for systematic errors are losses through wind. Depending on the location and altitude of the station, and the physical state of precipitation, e.g., rain or snow as well as the size and distribution of raindrops, the wind-induced losses vary considerably. In literature, average values for wind-induced errors are reported between $4 - 6\,\%$ (*Sevruk et al.*, 2009) up to $15\,\%$ for rainfall (*Sevruk*, 1996; *Chvíla et al.*, 2005). The error due to wind for snow and mixed precipitation is significantly higher and can reach up to $50\,\%$ for mixed precipitation and up to $80\,\%$ for snow (*Sevruk et al.*, 2009). There are more losses with increasing wind speeds and decreasing rain intensities (*Chvíla et al.*, 2005). In other words, larger wind-induced measurement errors occur in high mountain regions due to higher wind speeds and proportion of snow. Whereas, the influence of wind for convective rainfall events with larger raindrop sizes in summer at shielded lower altitudes is less.

Other sources of systematic errors are evaporation, wetting, splashing and drifting. In *Michelson* (2004), some loss constants for evaporation and wetting are listed. For a Hellmann rain gauge, loss constants for daily evaporation between $0.01 - 0.04\,mm\,d^{-1}$ for the hydrological winter half-year and $0.09 - 0.16\,mm\,d^{-1}$ for the hydrological summer half-year are specified. According to this, wetting constants vary between $0.14\,mm$ for rain, $0.10\,mm$ for solid precipitation and $0.18\,mm$ for mixed precipitation per event in a period of twelve hours. Furthermore, random errors can affect precipitation measurements. These are, e.g., clogging of the collecting vessel and funnel, reading errors, digitizing and transmission errors (*Sevruk*, 1996). Another issue of precipitation data are volume errors which can result from the measurements itself or the subsequent processing to areal precipitation. The conse-

quence for rainfall-runoff modeling may be a biased parameter estimation due to inadequate properties of the water balance and thus poor predictions (*Kavetski et al.*, 2006b). The determination of other sources of errors and uncertainty due to the model structure as well as parameterization inadequacies are also complicated by erroneous input variables (*McMillan et al.*, 2011).

4.1.2 Hydrological Data

Discharge observations are required in every application of conceptual hydrological models. Discharge values are necessary for the parameter estimation in the calibration phase and for the comparison and evaluation of the model output in the validation phase. Commonly, discharge data are not directly measured but derived from rating curves which have to be determined for every discharge gauge individually. Such a rating curve puts a discharge value explicitly in relation to a value of the water level. Water levels can be directly measured by different methods whereas the direct measurement of discharge is more complex. Therefore, water levels are continuously measured and subsequently converted to discharge data using the fitted rating curve. The rating curve is developed by several single discharge measurements at a preferably wide range of water levels. For this, the flow velocity is measured with a current meter at a defined cross section of the gauge. To obtain a good approximation of the average flow velocity, the total cross section is divided into several vertical sections. At these vertical sections, point measurements of the flow velocity are performed at defined water depths with the current meter. The width between two cross sections and the associated water levels are also measured to calculate the cross-sectional areas. These information allow the computation of the total discharge at a known water level (*Linsley et al.*, 1988). The individual data pairs of discharge and water level have to be transferred to a continuous curve. The relationship between the water level and the discharge can usually be described by a power or polynomial function. Probably, the most common function to fit the data is: $Q(W) = a \cdot W^b$. Q is the discharge and W the corresponding water level. The parameters a and b describe the flow dynamics and account for the geometry of the cross section (*Jónsson et al.*, 2002). The determination of these parameters are often done with a linear regression if a logarithmic transformation $Log(Q) = Log(a) + b \cdot Log(W)$ is applied to the function above. Then, the variables $Log(Q)$ and $Log(W)$ have a linear relationship. To obtain the best fitting values of the parameters a and b, the least square method is used where the square sum of the residuals is minimized (*Jónsson et al.*, 2002).

The uncertainties of discharge data are often disregarded for the use of hydrological modeling. This is justified with the fact that measurements of the discharge are the supposedly only and best available data for model calibration. Nevertheless, erroneous discharge data can lead to inadequate parameter estimations and consequently to model predictions affected by uncertainty (*Beven*, 2012). The possible range of uncertainty due to imperfect discharge data is discussed in the following sentences. The two error components of discharge data are the measurement uncertainty and the uncertainty coming from the methods of determining the rating curve (*Tomkins*, 2014). Measurement errors originate from the used devices as well as the observer. Some sources of measurement uncertainty using a current meter are the accurate determination of the water depth and the measurement of the flow

velocity itself. Another problem is the change of cross section properties due to river bed erosion or deposition, weedage and bank erosion (*Harmel et al.*, 2006). This can have a crucial influence on the relationship between water level and discharge. For this reason, a frequent control through discharge measurements and if necessary adjustment of the rating curve is compulsory. The most uncertain ranges of a rating curve are the very low and very high water levels. For these water levels, measurements are usually not available and this part is determined by extrapolation (*Tomkins*, 2014). In particular, the high discharge values are of major importance for flood predictions. *Jónsson et al.* (2002) investigated the methodological and personal uncertainty of determining rating curves. They applied different methods of establishing rating curves with the same discharge measurements. They found that particularly for extreme high water levels, discharge differences up to $20\,\%$ can occur if different rating curves are used. *Harmel et al.* (2006) specifies a survey of several values for the uncertainty of discharge measurements and rating curves summarized from various studies. Accordingly, the uncertainty of determining the discharge with flow velocity measurements ranges between $\pm 2\,\%$ for ideal conditions and up to $\pm 20\,\%$ for poor conditions. On average conditions, an uncertainty of $\pm 6\,\%$ remains. Uncertainties in the range of $\pm 5\,\%$ and $\pm 10\,\%$ are named for the relationship of water level and discharge (*Harmel et al.*, 2006).

4.1.3 Estimation of Evapotranspiration

Evapotranspiration, besides precipitation and discharge, is a part of the water balance in a catchment. For longer, continuous hydrological modeling purposes, the actual evapotranspiration (ET_A) has to be estimated for a sufficient consideration of antecedent catchment conditions before rainfall events (*Beven*, 2012). The potential evapotranspiration (PE) has to be estimated for a subsequent calculation of ET_A. The direct measurement of PE is difficult whereas values of PE are usually estimated with empirical equations. PE is the amount of water loss over an area which would occur with unlimited water availability (*Beven*, 2012). Especially, during longer dry periods the actual evaporated amount of water (ET_A) is not equal to PE due to limited water availability. The calculation of ET_A usually takes place in a subroutine of the hydrological model. For instance, this is shown for the HBV model in Subsection 2.2.1. There, the current water availability is represented by soil moisture storage which is related to a proportion of ET_A to PE. In this study, due to the lack of available meteorological data, the approach of *Hargreaves and Samani* (1985) was used to estimate PE. This method only uses data for the minimum (T_{min}), maximum (T_{max}) and mean (T_m) temperature as presented in Subsection 2.2.1. Thus, a minimum of data is required compared to more physically-based approaches such as Penman-Monteith. In addition, the devices measuring weather parameters such as solar radiation and humidity are often prone to stability errors (*Samani*, 2000). The estimation of PE can vary depending on the calculation method (*Singh et al.*, 2005).

An adequate assumption of evapotranspiration often gets insufficient attention in hydrological modeling purposes. Consequently, data of evapotranspiration are also subject to uncertainty because of their indirect determination through more or less complex empirical equations and the required input variables. In literature, errors are reported up to $\pm 10\,\%$ and for small evaporation rates even up to $\pm 20\,\%$ (*Lecher et al.*, 2015). The influence of inadequate

evapotranspiration data to the results of hydrological models are also investigated in several studies. Different data sets of improved PE estimations are tested on two different rainfall-runoff models in *Andréassian et al.* (2004). The evaluation of model performance shows that both models react remarkably insensitive to different data inputs of PE. Nevertheless, the calibrated model parameter sets differ from each other which allows the assumption that the models are able to compensate different inputs of PE. Another study of *Oudin et al.* (2005a) investigates the influence of considering PE with a long-term mean time series identically for each year or an individual estimation of PE on a daily resolution. The values of PE are estimated with the Penman approach and applied to four different rainfall-runoff models on 308 catchments. The study shows that no significant improvement of the model performance can be obtained with daily estimated inputs of PE. The differences in model performance regarding the Nash-Sutcliffe efficiency (NSE) ranges between $0.1\,\%$ and $0.8\,\%$. The companion study analyzes 27 different approaches for estimating PE on 308 catchments. The improvement of model performance (NSE) by using different PE time series is around $1\,\%$. They conclude that even simple temperature-based equations for estimating PE are capable for improving the model performance (*Oudin et al.*, 2005b).

4.2 Quality Control

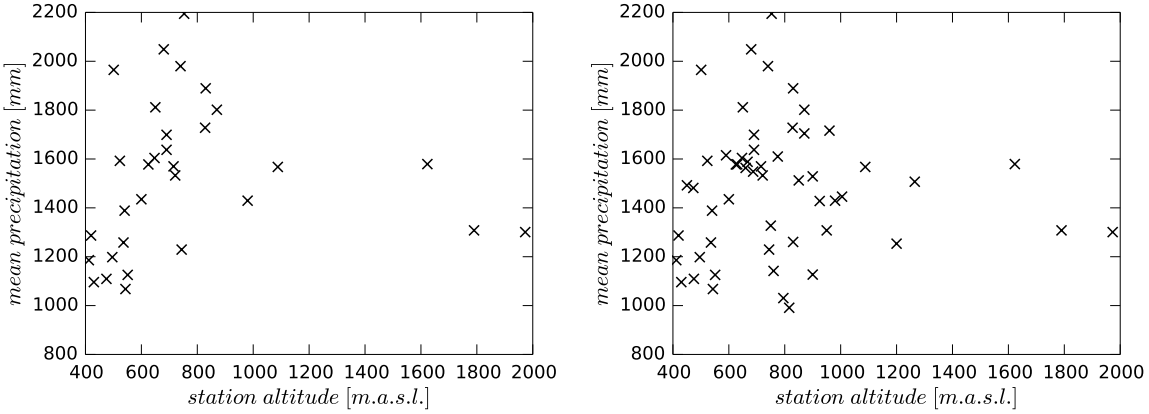
The data availability and quality influence the application of hydrological models and especially their proper calibration in different ways. Difficulties for model calibration can occur if the data have limited information content, considerable measurement errors or poor representation of the spatial variability of meteorological input variables due to sparsely distributed point measurements (*Gupta and Sorooshian*, 1985; *Gan et al.*, 1997). The quality of observation data is generally not determinable as well as unchangeable. Only the inadequacy of data can be identified and must be carefully verified. Thus, a quality control (QC) of the meteorological input variables as well as discharge data are an important requirement before using them in hydrological models. At least, a QC of model input variables should avoid the use of apparently inadequate data which result in insufficient model calibration and unreliable model predictions. For this purpose, a comprehensive QC starts with the operation and the frequent maintenance of the measurement devices. This includes the calibration of the measuring sensors, a regular revision of its validity and a manual inspection. In a second step, observation data have to be checked with different statistical procedures (*Estévez et al.*, 2011).

The subsequent sections show the results of QC regarding the meteorological input variables and discharge data. The subsection on rain gauge data describes the verification of meteorological data in more detail. Potential inadequacies of the required data can already be a hint for later occurring modeling limitations.

4.2.1 Rain Gauge Data

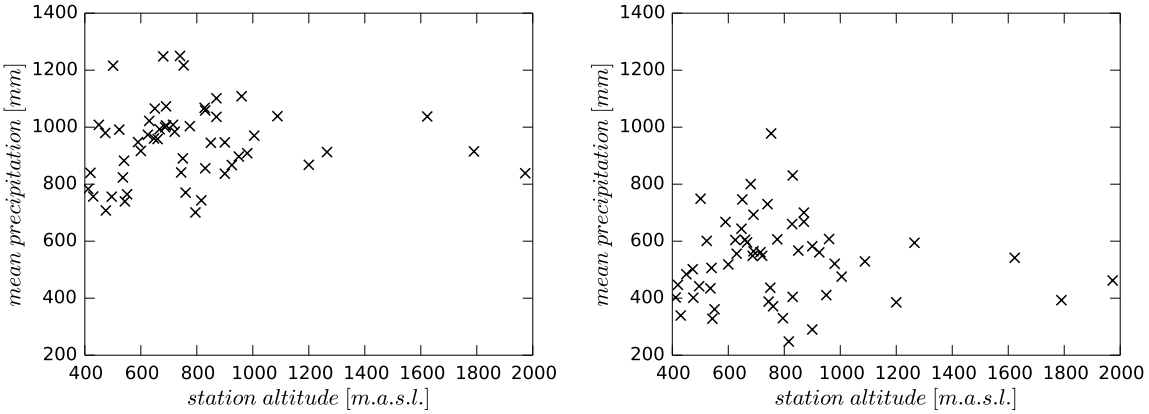
Subsection 4.1.1 specifies a variety of potential error sources for precipitation data. The quality of rain gauge data is a highly influencing aspect for hydrological modeling. Among

experts exists the conclusion that precipitation point measurements can only be checked by human eye to obtain a reliable and good data quality (*Einfalt and Michaelides*, 2008). Nevertheless, the manual plausibility check of data is difficult to apply if, e.g., real-time data has to be further processed for flood warning or the amount of investigated data is too large (*Einfalt and Michaelides*, 2008).



(a) Configuration *II* (b) Configuration *III*

Figure 4.1: Relationship between station altitude and mean annual precipitation sum.



(a) Summer (b) Winter

Figure 4.2: Relationship between station altitude and mean precipitation sum of rain gauge configuration *III* for the hydrological half-years.

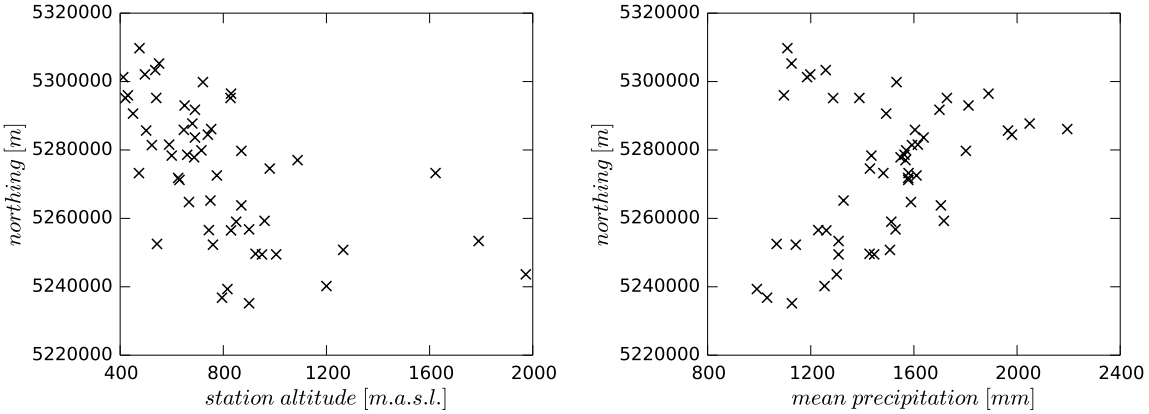
The precipitation data sets used are checked and prepared for further processing. Firstly, for every single rain gauge station data gaps within the time series are detected and filled with missing values. Thus, every time series is complete and comparable according to the time stamp over the total investigation period. Then, the time series are checked for extreme high values and their physical possibility. Further procedures of detecting inadequate data such as constant values for a certain time or unusually low as well as high values are suggested in *Einfalt et al.* (2008), but not applied in this study. Additionally, the mean annual

precipitation sums of each individual rain gauge is compared to geographical properties. Such an analysis provides information about expected relationships between precipitation patterns and elevation or orientation. Mean annual sums of precipitation typically increase with elevation. The geographic direction of the gauging stations can also influence the mean annual precipitation sums. The windward side of a mountain range usually records higher mean precipitation sums than the lee side. Due to accumulation effects on the windward side, even lower located gauging stations can measure above-average sums of precipitation. Two gauging stations in the investigation area that exhibit just such an accumulation effect are listed in *BMU* (2003). Both stations, *Aschau-Stein (680 m.a.s.l.)* and *Ruhpolding-Seehaus (753 m.a.s.l.)* observe more than $2000\,mm$ of mean annual precipitation. Thus, higher mean annual precipitation is not just related to increasing station altitudes (*BMU*, 2003) but also to the geographical location. Due to the geographical orientation of the Alps in the study area (Figure 3.1), such accumulation effects in the north-south direction must be taken into account when evaluating precipitation data.

Table 4.1: Correlation (R) between station altitude [m.a.s.l.] and mean precipitation sum [mm] for different station configurations and time periods.

time period	configuration II				configuration III			
	all	< 1000	< 900	< 700	all	< 1000	< 900	< 700
annual	0.06	0.54	0.65	0.65	-0.01	0.18	0.30	0.60
summer	0.13	0.53	0.64	0.62	0.05	0.20	0.29	0.54
winter	-0.01	0.53	0.63	0.66	-0.06	0.15	0.29	0.63

Stations with more than $12.5\,\%$ missing values are excluded for further data evaluation. The threshold value corresponds to a maximum of one year of missing data within the investigated period of eight years. This adjustment is intended to avoid too much corruption of the mean annual values due to too many missing values. The remaining missing values in the time series are considered as zero for calculating the precipitation sums. The mean annual precipitation sums tend to be underestimated depending on the number of missing values per gauging station. Figure 4.1 shows the relationship between station altitude and mean annual precipitation sum of the rain gauge configurations *II* and *III*. The majority of the stations are located below an altitude of $900\,m.a.s.l.$ and there are only a few stations above $1000\,m.a.s.l.$. For configuration *II*, a trend toward higher mean annual precipitation with increasing station altitude up to $900\,m.a.s.l.$ is visible. A wide range of mean precipitation for stations at the same elevations is also apparent. The total annual amounts of precipitation do not exceed more than $1500\,mm$ for stations above $1000\,m.a.s.l.$ for both configurations. Configuration *III* has more stations available between $800\,m.a.s.l.$ and $1000\,m.a.s.l.$. The mean precipitation sums show a high variability in this range and a trend of increasing precipitation sums for higher altitudes is not obvious. The relatively low mean annual precipitation sums at the higher stations may indicate underestimation due to systematic measurement errors. Whereas, the partly high mean annual sums of lower stations are plausibel due to the above described accumulation effects at the foothills of the Alps. Figure 4.2 shows the mean precipitation sums of configuration *III* separately for the hydrological half-years. The



(a) Station altitude - station northing

(b) Precipitation sum - station northing

Figure 4.3: Relationship between altitude and northing of the gauging stations (a) and mean annual precipitation and northing (b) of configuration *III*.

patterns of precipitation sum and station altitude look quite similar for the summer and winter half-years. Solely, the amount of precipitation is higher for the hydrological summer half-years. A systematic measurement error due to snow, especially for higher gauging stations, is not clearly recognizable for the winter half-years. In the left plot of Figure 4.3, the locations of the gauging station according to the north-south orientation are given. Station altitudes generally increase in a southward direction. Therefore, shadowing effects can lead to lower mean annual precipitation sums for higher, but also more in the hinterland located stations.

Table 4.1 shows the correlation (R) between station altitude and mean precipitation sum for the configuration *II* and *III*. The analyzed time period refers to the hydrological years from 2005 until 2012. Both configurations are further classified into groups of different station compositions. The denomination *all* involves all available stations, while the numbers, e.g., *1000*, specify the altitude level above which all stations are disregarded in the calculation of R. If all stations are considered, the results indicate that there is almost no correlation between station altitude and precipitation sum. For the hydrological winter half-years, even a light anticorrelation is observable. This means that there are decreasing mean precipitation sums with increasing station altitude. Regarding the different half-years, R is higher for the summer half-years than for the winter half-years. This may be an indicator for measurement errors due to snowfall. The more stations of higher altitudes are disregarded, the higher is R. For configuration *II*, a significant increase of R already occurs if the stations above $1000\,m.a.s.l.$ are ignored. Similar values of R are obtained by configuration *III* if just the stations below $700\,m.a.s.l.$ are considered. This can be explained by the fact that between $700\,m.a.s.l.$ and $1000\,m.a.s.l.$ more stations are available which may influence the values of R. These stations in the middle altitudes can be more affected by shadowing effects because of the leeward locations in the valleys of the hinterland. If stations above $1000\,m.a.s.l.$ are disregarded, another finding is that the differences in R between summer and winter half-

years almost disappear. In the case where just stations below $700\,m.a.s.l.$ are used, R is even higher during the winter half-years. The reason may be that stations at lower elevations are less affected by measurement errors due to snow. On the other hand, during the summer half-years, other circumstances such as high rainfall intensities can cause measurement uncertainties. The higher spatial variability of rainfall patterns due to convective events may also influence the distribution of precipitation within the study area independently of station altitudes.

According to the location of the study area on the north side of the Alps, a relationship between mean precipitation sum and the north-south orientation of the stations is checked (Figure 4.3). Despite mainly increasing station altitudes southwards, the mean annual precipitation sums do not show the same trend. The distribution of precipitation due to their latitudinal location follows a rather horizontal parable. Above a northing of $5280000\,m$ an increasing trend of the precipitation sums to the south exists. But, below the northing of $5280000\,m$ a decreasing trend of precipitation sums together with increasing station altitudes is apparent southwards. This northing crosses the study area approximately between the gauge *Unterjettenberg* and the city of *Berchtesgaden* (Figure 3.1). Looking at the topography, it can be seen that south of this northing the first high mountain ranges appear. Thus, the increasing trend of precipitation can again be caused by accumulation effects on the north side of the mountain range as well as the lower amounts of precipitation to the south can be the result of shadowing effects.

4.2.2 Temperature Data

Especially in mountainous catchment areas, such as the investigated *Saalach* catchment, adequate temperature data are an important prerequisite. Temperature data are required for the estimation of evapotranspiration as well as a proper representation of snow processes. As already mentioned in Section 3.2 and illustrated in Figure 4.4, the temperature stations and consequently the available point measurements are even more sparse than the rain gauge stations. After revising the data time series with the requirement of less than one year of missing data ($12.5\,\%$), only ten stations remain for the analysis.

After fitting a regression line to the data pairs of station altitude and mean annual temperature, the slope of the regression line provides an average temperature lapse rate of $4\,°C\,km^{-1}$. In literature, mean lapse rates of about $6.5\,°C$ to $7\,°C$ decrease per kilometer height are specified for the lower troposphere (*Linsley et al.*, 1988; *Häckel*, 2008). The estimated temperature lapse rate is slightly below the literature values. This may be due to regional variations or the relative short time series of eight years. Another reason is that the slope refers to mean annual temperatures which also involves inversions. Usually, the temperature decreases with increasing altitude but during an inversion the temperature profile reverses and the temperature increases with height. This fact can reduce the estimated temperature lapse rate somewhat. Nevertheless, the relation between station altitude and temperature observations has a coefficient of determination (R^2) of 0.99 and can be considered as linear. The assumption that temperatures decrease linearly with increasing elevation is reasonable. Therefore, estimating the temperatures with linear lapse rates in grid cells

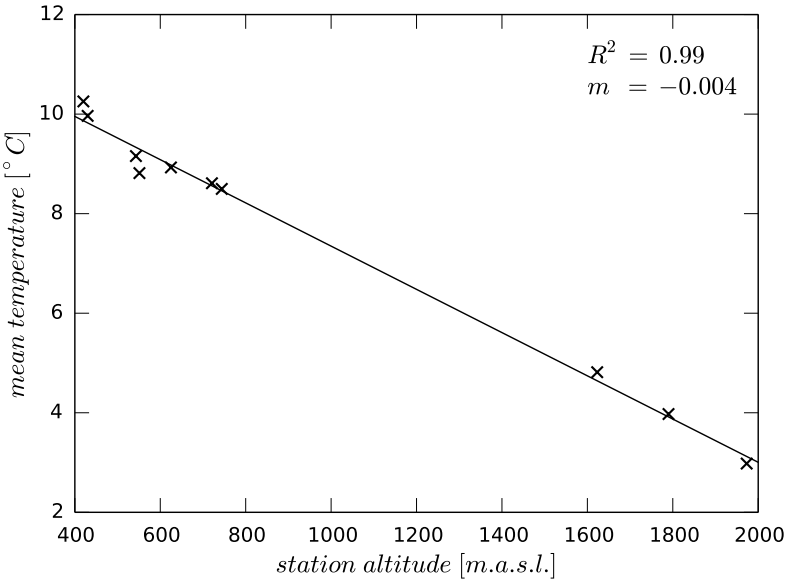


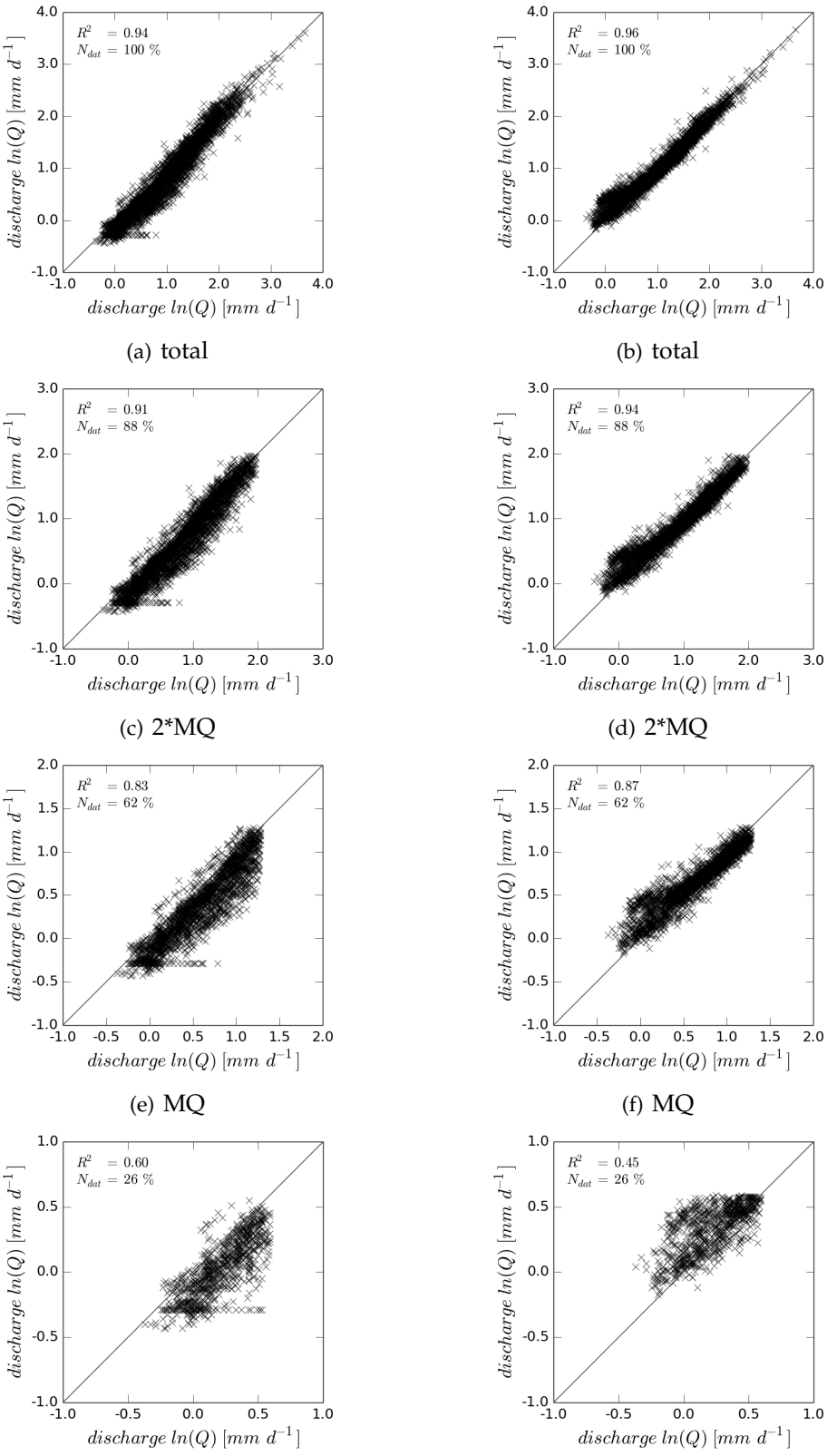
Figure 4.4: Relationship between station altitude and mean annual temperature.

without observations on different altitudes might be practicable. The finding of an explicit relation between elevation and mean temperature is essential for subsequent spatial interpolations. Especially, if wide vertical ranges are not covered by monitoring stations. Figure 4.4 illustrates that between $800\,m.a.s.l.$ and $1600\,m.a.s.l.$ no temperature stations are available. Thus, the almost linear relationship between station altitude and mean temperature allows for reliable temperature estimations in this ungauged range.

4.2.3 Discharge Data

The QC of the discharge time series provides a rough survey of data inconsistencies. In a first step, daily discharge time series of one upstream and one downstream gauge are compared to the data of gauge *Unterjettenberg*. The gauge *Weissbach* is located $27\,km$ upstream of gauge *Unterjettenberg* and has a catchment area of $567.5\,km^2$. The gauge *Staufeneck* is $11\,km$ downstream and its catchment covers about $1022\,km^2$. The time series have no data gaps. Due to the relative short distances of gauge *Unterjettenberg* to the other two gauges and average slopes between $0.4\,\%$ and $0.6\,\%$, flow times of some hours are assumed. Thus, extreme discharge changes between the three gauges for a daily time step should be very rare. Consequently, a certain correlation between discharge data of two gauges at the same time step is expectable. The discharge values of gauge *Unterjettenberg* are compared with the two other gauges.

Figure 4.5 shows the different data sets, the coefficient of determination (R^2) and the used data quantity related to the total data set (N_{dat}). The left column refers to the data pairs of *Unterjettenberg/Weissbach* and the right column to *Unterjettenberg/Staufeneck*. For the analysis, the total data set and some modifications related to the long-term mean discharge (MQ)



(a) total (b) total

(c) 2*MQ (d) 2*MQ

(e) MQ (f) MQ

(g) 0.5*MQ (h) 0.5*MQ

Figure 4.5: Comparison of different discharge data sets between two gauges. The left column presents the relationship between the gauges *Unterjettenberg* and *Weissbach* and on the right side gauge *Unterjettenberg* is compared to *Staufeneck*.

are considered. In the modified data sets, values above certain thresholds are disregarded for all gauges. For example, the modification *2*MQ* considers only values which are not greater than twice the MQ. A quite high relationship between the respective pairs of gauges is identified for the *total* data set and *2*MQ*. The smaller the discharge values compared, the lower the correlation between the data pairs. The gauges *Unterjettenberg* and *Staufeneck* show a stronger relationship with the exception of the configuration *0.5*MQ*. Reviewing all scatterplots in Figure 4.5, two facts can be highlighted. Firstly, the discharge data of gauge *Weissbach* show some constant values for varying values of gauge *Unterjettenberg*. These values appear successively from the end of January until mid of February 2006 in the time series. The mean daily temperatures in this period show values of almost permanently below $0°C$ and often with severe frost below $-4°C$ in daily mean. This data inadequacy is probably related to the gauge *Weissbach* and could be caused by frozen measurement devices. The second finding is a clear decline of the linear relationship for the very low discharge values for both pairs of gauges. Several reasons could cause these effects. Changes of low flows can occur due to local effects such as melting of remaining snowfields or intensive convective rainfall events with limited spatial extent in a subcatchment. Consequently, such small discharge changes do not necessarily result from the total catchment and are just recorded by single gauges. Another reason may be the influence of regulating the flow rate with dams and weirs for hydro-electric power plants. Such an anthropogenic regulation of the flow rate certainly occurs between *Unterjettenberg* and *Staufeneck*. Finally, differences in discharge values can also be caused by peak flows passing the gauges on different time steps. In summary, no apparent data inconsistencies can be detected for the gauge *Unterjettenberg* for this data control.

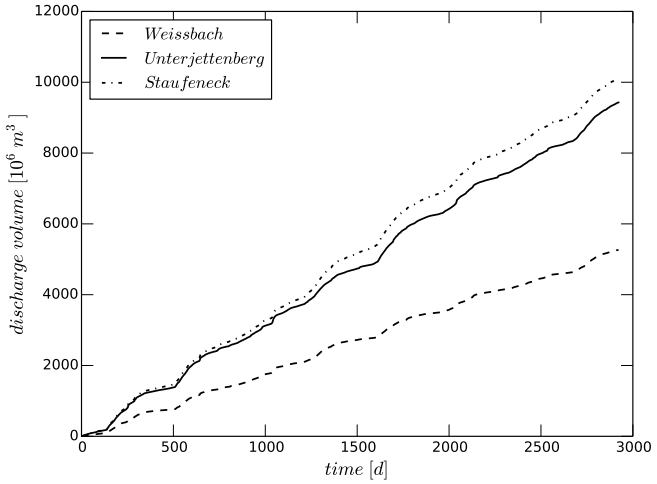


Figure 4.6: Cumulative curve of discharge volume for the hydrological years 2005 - 2012.

Another approach of data control is the comparison of discharge volumes between the different gauges. Here, the cumulative curve of discharge volume (Figure 4.6) and the annual discharge volume (Figure 4.7) of all three gauges is presented for the hydrological years from 2005 until 2012. The cumulative discharge volumes represent a reasonable increase over the time period in which the most upstream gauge has the lowest and the most downstream

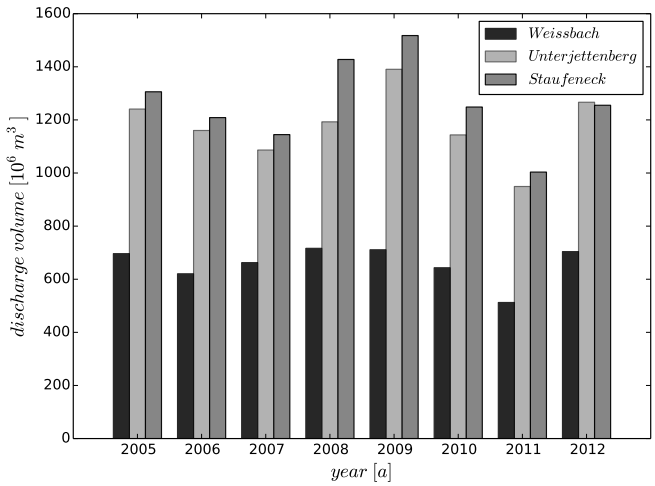


Figure 4.7: Annual discharge volume for the hydrological years 2005 - 2012.

gauge the highest total volume. As rough estimation, the ratio of total volume is compared to the ratio of additional catchment area between the gauges. The average extra discharge volume between the gauges *Weissbach* and *Unterjettenberg* is $44\,\%$ and the corresponding increase of catchment area is $39\,\%$. Accordingly, the additional discharge volume between the gauges *Unterjettenberg* and *Staufeneck* is $7\,\%$ regarding $9\,\%$ larger catchment area. Thus, the ratio of total volume between the gauges and additional catchment area also seems to be in a realistic range.

Solely, the gauges *Unterjettenberg* and *Staufeneck* show an almost equal gradient for the last part of the cumulative curve. Figure 4.7 shows that the differences in discharge volume are already small for the hydrological year 2011 and for 2012 even slightly less discharge volume is calculated for the downstream gauge *Staufeneck*. The ratios of discharge volume between the gauges *Weissbach* and *Unterjettenberg* are in a reasonable range for the two years. Therefore, it can be an indicator for errors in the observations of gauge *Staufeneck* within this period. The year 2008 also shows an disproportionate increase of discharge volume ($18\,\%$) between the gauge *Unterjettenberg* and *Staufeneck* whereas the ratio between the gauges *Weissbach* and *Unterjettenberg* is below the mean ($40\,\%$). This could be an indicator for inconsistent data of the gauge *Unterjettenberg*.

4.3 Conventional Spatial Statistical Methods

Meteorological observation data from point measurements at different locations are often available. But, for hydrological modeling purposes estimates of meteorological input variables are usually required at any location in the catchment area. Such a spatial representation of the input variables should consider the variability of, e.g., temperature and precipitation patterns due to topographic effects and their non-uniform distribution. Thus, the input variables need to be transferred from a point to a spatial representation. Traditional

interpolation methods of precipitation point measurements are *inverse distance weighting* or the *Thiessen polygon method* (*Beven*, 2012). Other interpolation approaches are geostatistical methods like kriging. The kriging methods and some specific results for this study are briefly presented in the first part of this section. Further reading about the different kriging methods can be found in, e.g., *Kitanidis* (1997), *Wackernagel* (1998) and *Sorooshian* (2008).

4.3.1 Ordinary Kriging

One of the most widely used geostatistical methods of interpolating meteorological variables is ordinary kriging (ODK). Kriging is based on a linear interpolation approach and provides the possibility to estimate a value $z^*(x_0)$ at an unobserved location x_0. A prerequisite for kriging are values $z(x_i)$ from observation stations at locations x_i in the neighborhood which allow the determination of a variogram. A variogram represents the spatial dependence of a variable $z(x_i)$ for increasing distances. The experimental variogram $\gamma(h_{ij})$ is defined as follows:

$$\gamma(h_{ij}) = \frac{1}{2n(h_{ij})} \cdot \sum_{i=1}^{n(h_{ij})} [z(x_i) - z(x_i + h_{ij})]^2 \tag{4.1}$$

where $n(h_{ij})$ is the number of station pairs separated by the vector h_{ij}. Before applying the kriging algorithm, a continuous function, called the theoretical variogram, must be fitted to the discrete values of the experimental variogram. Different standard functions of theoretical variograms can be found in *Kitanidis* (1997) or *Sorooshian* (2008).

The weighted linear combination of observations for estimating a value $z^*(x_0)$ can be described as follows:

$$z^*(x_0) = \sum_{i=1}^{n} \lambda_i \cdot z(x_i) \tag{4.2}$$

where n is the amount of neighboring observations and λ_i represents the weight of the particular observation $z(x_i)$. Then, the linear equation system of ODK introducing the Lagrange multiplier ν:

$$\sum_{j=1}^{n} \lambda_j \gamma(x_i - x_j) + \nu = \gamma(x_i - x_0) \qquad i = 1, ..., n \tag{4.3}$$

$$\sum_{j=1}^{n} \lambda_j = 1 \tag{4.4}$$

is solved to obtain the weights λ_j that are required in Equation 4.2.

4.3.2 External Drift Kriging

External drift kriging (EDK) is used for the incorporation of secondary information. The assumption is a linear relationship of the secondary variable $y(x_i)$ to the variable $z(x_i)$. For example, elevation data from a digital elevation model (DEM) can be used as external drift for interpolating meteorological variables such as precipitation or temperature. The equation system for EDK with two Lagrange multipliers ν_1 and ν_2 is:

$$\sum_{j=1}^{n} \lambda_j \gamma(x_i - x_j) + \nu_1 + \nu_2 y(x_i) = \gamma(x_i - x_0) \qquad i = 1, ..., n \tag{4.5}$$

$$\sum_{j=1}^{n} \lambda_j = 1 \tag{4.6}$$

$$\sum_{j=1}^{n} \lambda_j y(x_j) = y(x_0) \tag{4.7}$$

The advantage of EDK over ODK is the consideration of secondary information such as elevation. In this study, many observation stations for precipitation as well as temperature are located in lower regions whereas the topography of the investigation area is spatially highly variable. The availability of stations in upper altitudes is very sparse and no observations at all exist for the highest ranges. Especially, the interpolation of temperature point measurements with ODK leads to erroneous estimations at unobserved locations in high altitudes. The characteristic of decreasing temperature with increasing elevation is not correctly represented by ODK. EDK also accounts for the assumption of increasing precipitation sums with increasing elevations.

4.3.3 EDK with DEM smoothing

The modified EDK_s uses a smoothed DEM as secondary information. Accumulated precipitation measured by rain gauges is generally not affected by small scale variability of the topography. Thus, interpolated precipitation patterns using the conventional EDK reproduce the structures of the underlying DEM in an unreasonable way. In *Bárdossy and Pegram* (2013), a directional smoothing approach for the consideration of orographic and advective effects on precipitation patterns is introduced. The original elevation $h(x)$ at a location x can be smoothed and shifted to the elevation $h_s(x)$:

$$h_s(x) = \sum_{g=1}^{G} w(x - x_g) h(x_g) \tag{4.8}$$

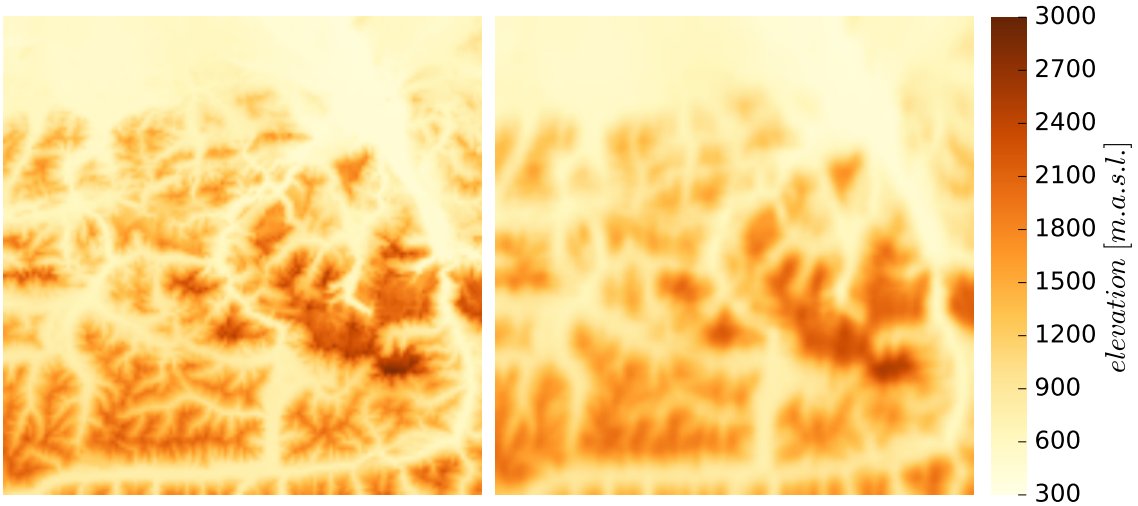
where

$$w(x - x_g) = \begin{cases} 0 & if\ |x - x_g| > d \quad or \quad \langle v, (x - x_g) \rangle < 0 \\ c_w \cdot \left(1 - \frac{|x - x_g|}{d}\right) & otherwise \end{cases} \tag{4.9}$$

and

$$\sum_{g=1}^{G} w(x - x_g) = 1 \tag{4.10}$$

G indicates the number of elevation grid points and x_g is a grid point of the DEM. Vector v accounts for the predominant wind direction with a north and east component. The smoothing distance is represented by d and w are weights. The selection of the constant c_w has to fulfill the constraint of Equation 4.10 (*Bárdossy and Pegram*, 2013).



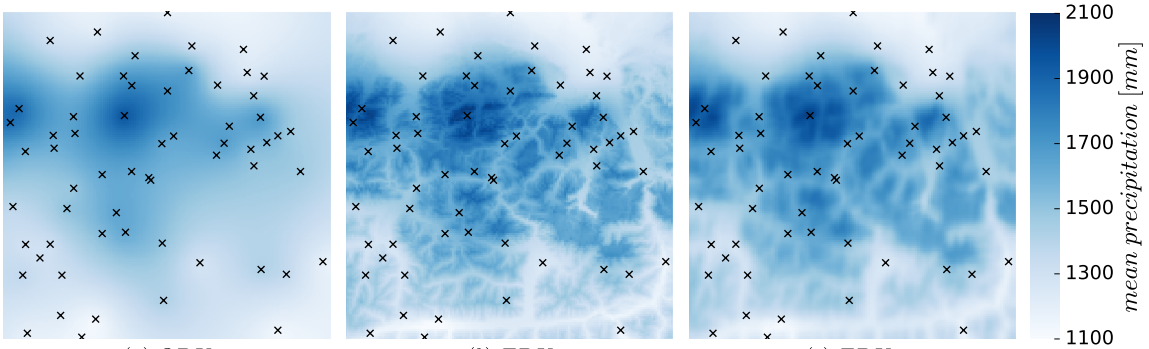
(a) *original DEM* (b) *smoothed DEM*

Figure 4.8: Difference between original and smoothed DEM.

In this study, the original DEM is smoothed with a distance of $d = 5\,km$ without shifting into the prevailing wind direction. Figure 4.8 shows the difference between the original and smoothed DEM. In the version of smoothed DEM, the microtopography such as small steep valleys disappears.

4.3.4 Comparison of Kriging Methods and Water Balances

The different interpolation results for mean annual precipitation, obtained using the three kriging methods for rain gauge configuration *III*, are shown in Figure 4.9. The map generated by ODK indicates a decrease of the mean annual precipitation in southern direction. Especially, in the south-eastern part of the investigation area the high mountain ranges and



(a) ODK (b) EDK (c) EDK_s

Figure 4.9: Mean annual precipitation interpolated by three different kriging methods. The black crosses indicate the gauging stations of configuration *III*.

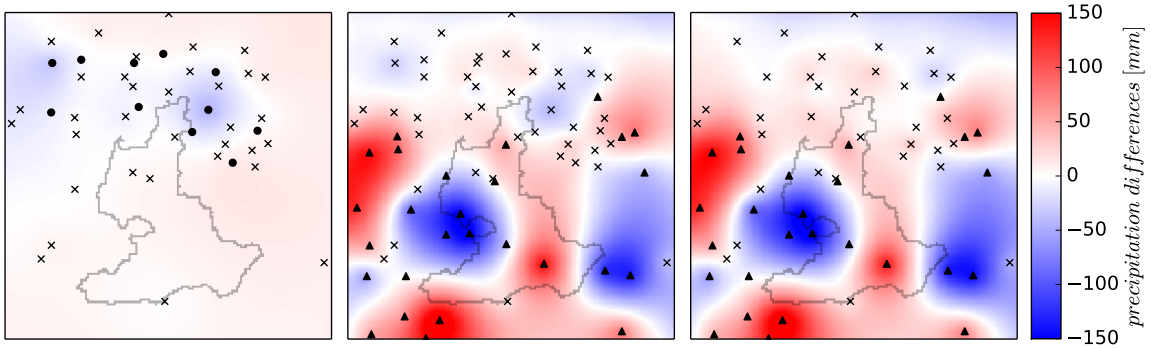
the expected higher precipitation amounts are not reproduced. The precipitation maps using a DEM as additional information account better for the orographic variability of the precipitation. In both cases, the topography of the investigation area is visible whereas the map of EDK_s has smoother contours. There, only the broad valleys in the hinterland shielded by the high mountain ranges receive clearly less precipitation in annual mean.

The mean annual differences of areal precipitation due to the rain gauge configuration and kriging method are analyzed in Table 4.2. The spatial distribution of the precipitation differences is shown for ODK in Figure 4.10 and EDK_s in Figure 4.11.

Table 4.2: Mean annual differences of areal precipitation [mm] between different rain gauge configurations and kriging methods for the catchment *Unterjettenberg*.

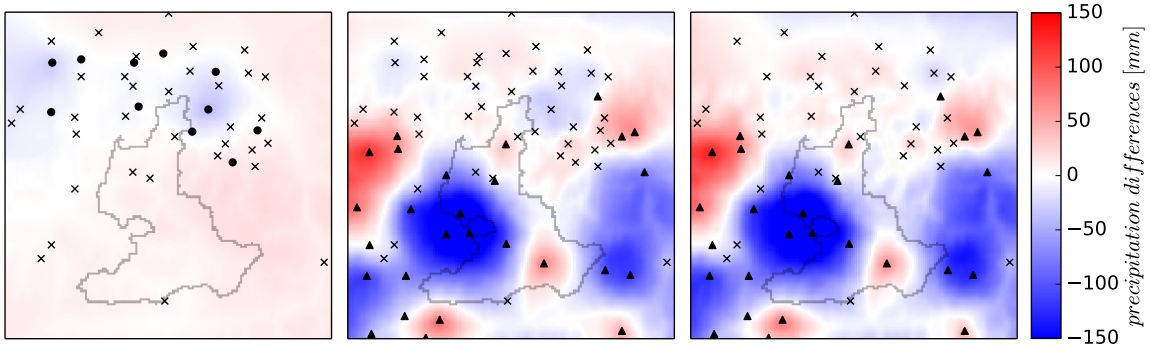
	ΔP_{I-II}	ΔP_{I-III}	ΔP_{II-III}
ODK	2.1	4.2	2.1
EDK	4.6	-35.1	-39.7
EDK_s	4.6	-35.2	-39.8

The mean annual differences between all rain gauge configurations interpolated by ODK are marginal. The differences of processing the data with EDK or EDK_s are negligible. Taking topography into account as secondary information shows a clear increase of mean annual precipitation for configuration *III*. On average, configuration *III* estimates between $35\,mm$ and $40\,mm$ more areal precipitation per year using EDK or EDK_s. Thus, the consideration of topographic information is essential in such a catchment. The strong influence of a single gauging station located in a shielded valley in the south-east of the catchment and recording low mean annual sums is given for ODK in Figure 4.10.



(a) Configuration $I - II$ (b) Configuration $I - III$ (c) Configuration $II - III$

Figure 4.10: Differences of mean annual precipitation between the configurations *I* - *III* calculated by ODK.



(a) Configuration $I - II$ (b) Configuration $I - III$ (c) Configuration $II - III$

Figure 4.11: Differences of mean annual precipitation between the configurations *I* - *III* calculated by EDK_s. The black dots are additional stations of configuration *I* and the black triangles are stations of configuration *III*. The bright gray line defines the catchment.

Finally, the results for precipitation from the different interpolation methods are used to look more closely at the water budget. All calculations refer to the hydrological year. The hydrological year starts at the 1st of November and the denotation corresponds to the subsequent year (*Maniak*, 1997). The reason for the shift to November is that a catchment usually has the lowest water resources during this season. Thus, snow accumulation in November and December which melts in spring of the next year is considered correctly. The hydrological half-years are defined from 1st of November until 30th of April (winter) and from 1st of May until 31th of October (summer).

In addition to the general water balance, the climatic water balance (CWB) is first considered as an important quantity for estimating water availability. The calculation of the CWB takes the difference between precipitation depth (P) and the depth of the potential evapotranspiration (PE) for a certain location and time period into account (*BMU*, 2003):

$$CWB = P - PE \tag{4.11}$$

Table 4.3: Annual and semi-annual climatic water balance (CWB) for different interpolation methods of rain gauge configuration *III*. The analysis refers to the hydrological year and half-years in the period of 2005 - 2012.

	annual [mm]			summer [mm]			winter [mm]		
	ODK	EDK	EDK_s	ODK	EDK	EDK_s	ODK	EDK	EDK_s
mean	836	908	911	447	486	487	388	422	423
min	637	735	738	202	250	252	222	261	262
max	956	1025	1027	565	599	600	540	579	580

The unit of all variables is mm. Initially, it is checked whether situations with water deficits occur within the catchment during the investigation period. It is also possible to identify seasonal differences in water availability. The quantities of available water for different time periods are shown in Table 4.3. The results are mean areal values of the catchment *Unterjettenberg* and refer to the rain gauge configuration *III*. They have a positive balance for all time periods. The mean annual values vary between $637\,mm$ and $1027\,mm$ depending on the interpolation method. The mean amount of available water is almost the same for the hydrological summers and winters. On average, slightly less water is available ($46.5\,\%$) in winter. An almost equal ratio of the mean semi-annual water availability between both half-years is plausibel for the Alps (*BMU*, 2003). The CWB in Table 4.3 show that the interpolation methods using elevation as external drift outperform ODK. Higher amounts of precipitation are assumed when considering elevation, and thus more water is available for runoff generation. The average differences of available water are about $8\,\%$ for the annual as well as semi-annual time periods. This indicates that the larger estimated quantities of available water with EDK methods are independent of the season. The differences between EDK and EDK_s are negligible. The fact that the CWB are always positive within the investigated time period is confirmed by the evaluations in *BMU* (2003). However, mean annual values up to $2000\,mm$ are reported for the Alps which are more than twice of the estimated values here.

The general water balance determines the amount of water and its changes within a catchment. The applied equation contains the depth of precipitation (P) as input, discharge (Q) and the potential evapotranspiration (PE) as output of the system:

$$P = Q + PE + \Delta S \tag{4.12}$$

The term ΔS accounts for the storage and release change in the catchment through, e.g., soil moisture or snow cover (*Lecher et al.*, 2015). ΔS can be neglected for time periods longer than one year. The unit of all variables is mm. In this study, the storage change and lateral in- and outflows of the catchment are neglected. Nevertheless, this simplified estimation of the water balance can be used as an indicator for inconsistent quantities of the system variables. A distinct deficit or excess of the water balance for the total investigation period or individual years are potential sources of uncertainty for the subsequent hydrological modeling.

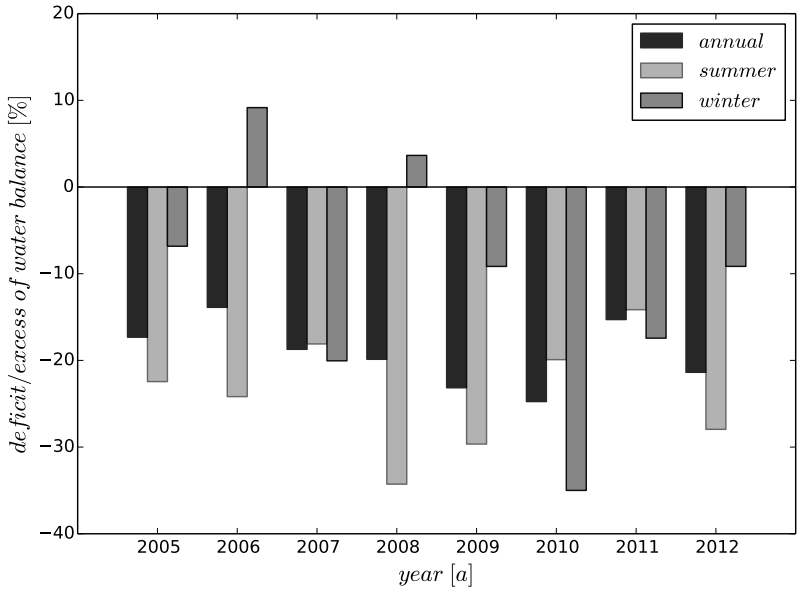


Figure 4.12: Annual and semi-annual water deficit or excess of rain gauge configuration *III* interpolated with EDK_s.

Figure 4.12 presents the water deficit or excess of the total water balance for each hydrological year or half-year of the investigation time period. It can clearly be seen that the annual and the semi-annual water balance for the hydrological summer are always negative. Only, two half-years in winter have a positive water balance with a maximum excess of $9\,\%$. The summer half-years tend to higher deficits than the winter half-years. There, the semi-annual deficit ranges between $-14\,\%$ and $-34\,\%$. The maximum semi-annual water deficit reaches values of $-35\,\%$. Looking at the mean annual and semi-annual water balance calculated for the investigated time period from 2005 until 2012, it can be seen that the system has an overall water deficit. As presented in Table 4.4, the mean annual water deficit ranges between $19.3\,\%$ and $23.6\,\%$, depending on the used interpolation method and rain gauge configuration. The mean differences of water deficiency between the summer and winter half-years range about $9.5\,\%$ and $13.0\,\%$.

Table 4.4: Mean annual and semi-annual deficit of the water balance [%] for different configurations of rain gauge stations (*I* - *III*) and interpolation methods.

	annual (2005 - 2012)			summer			winter		
	ODK	EDK	EDK_s	ODK	EDK	EDK_s	ODK	EDK	EDK_s
I	-23.6	-21.6	-21.5	-26.8	-25.8	-25.7	-17.4	-13.5	-13.4
II	-23.6	-21.8	-21.7	-27.0	-25.9	-25.8	-17.2	-13.8	-13.7
III	-23.2	-19.4	-19.3	-26.9	-23.9	-23.8	-16.2	-10.8	-10.6

Distinguishing the results between the different interpolation methods, it can be concluded that the highest deficits are estimated with ODK. The used EDK and EDK_s provide almost equal results. According to the rain gauge configurations, it is shown that configuration *I* and *II* have very similar results. Configuration *III* has about $2\,\%$ less deficit compared to the other configurations. Whereby, the differences are slighty higher during the winter months. Only, the water deficits interpolated by ODK do not show a noteworthy difference for all rain gauge configurations. Consequently, the rain gauge configuration *III* interpolated by EDK_s is supposed to provide the most promising input data set for the hydrological models due to the lowest water deficit in the water balance.

4.4 Application of a New Conditional Simulation Method - Random Mixing

Precipitation uncertainty due to errors of point measurements and the derivation of interpolated precipitation fields at ungauged locations influences the calibration of hydrological models. The introduced kriging methods only provide estimations at ungauged locations with a minimum error variance. Consequently, the variability of real precipitation fields is smoothed out (*Götzinger and Bárdossy*, 2008). Thus, the use of conditional simulation techniques is an option to consider spatial variability and to quantify the uncertainty of precipitation data more reasonable. Conditional simulation methods embedded in a Monte Carlo framework generate a set of realizations of a variable such as precipitation following certain constraints. Each realization keeps the observed values at the conditioning points as well as the spatial variability of the observations.

4.4.1 Basic Theory

The methodology of *random mixing* (RM) is primarily introduced in the paper of *Bárdossy and Hörning* (2016a) where this approach is applied to inverse modeling problems of groundwater flow and transport. Here, a short introduction of the basic theory of RM is given and for a comprehensive description it is referred to *Bárdossy and Hörning* (2016a,b) and *Hörning et al.* (2019). RM takes place in the standard normal space for which the data need to be transformed if they have a non-Gaussian marginal distribution. A spatial random field $Z(x)$ is generated by a linear combination of n unconditional random fields $Y_i(x)$, all having the same marginal distribution and covariance structure:

$$Z(x) = \sum_{i=1}^{n} \alpha_i Y_i(x) \tag{4.13}$$

Different methods of generating such independent random fields $Y_i(x)$ and their sources are specified in *Bárdossy and Hörning* (2016a). The weights α_i of the linear equation system have to fulfill:

$$\sum_{i=1}^{n} \alpha_i^2 = 1 \tag{4.14}$$

in order to preserve the covariance structure. Furthermore, the value of the conditional field $Z(x_k)$ at the observation location x_k has to be equal to the observed value z_k:

$$Z(x_k) = z_k \qquad k = 1, ..., K \tag{4.15}$$

Finally, the weights α_i for n independent realizations of unconditional fields $Y_i(x)$ are identified so that:

$$\sum_{i=1}^{n} \alpha_i Y_i(x_k) = z_k \qquad k = 1, ..., K \tag{4.16}$$

The method of RM also allows for an extension of the simulation procedure by additional information such as the topography represented by a DEM or other constraints.

4.4.2 Linear Inequality Constraints

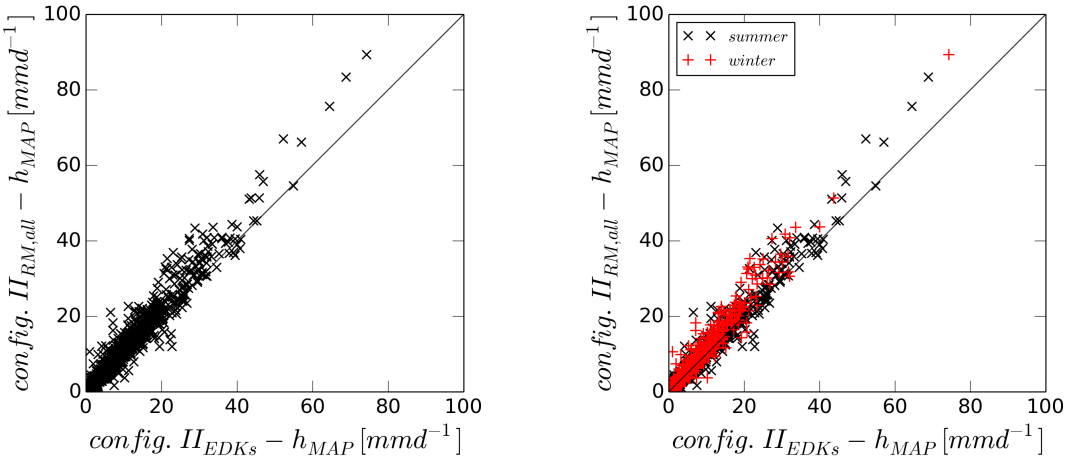
The implementation of linear inequality constraints provide a stochastic approach for a stronger consideration of uncertainties due to measurement errors. In this study, *greater-equal* constraints are applied to precipitation data:

$$\sum_{i=1}^{n} \alpha_i Y_i(x_s) \geq z_s \qquad s = 1, ..., S \tag{4.17}$$

which allow for values greater than the observed values z_s at the conditioning points x_s. The implementation of such constraints can be limited to a number of conditioning points. The selection of conditioning points needs to fulfill certain requirements such as exceeding a predefined station altitude. This extension enables the analysis of potentially underestimated precipitation amounts due to external influences such as wind or snow.

4.4.3 Potential for Hydrological Modeling

Conditional simulations such as RM enable to investigate the impact of input uncertainty on the model calibration and its output. Primarily, the capability of RM is investigated based on precipitation data in this study. Other meteorological input data such as temperature can also be used. Hydrological models usually require time series for several years in a high temporal resolution, e.g., daily or hourly, for calibration and evaluation procedures. Thus, the simulation method needs to be fast for generating many realizations per time step for long time series. RM satisfies this prerequisite and time series with a set of precipitation



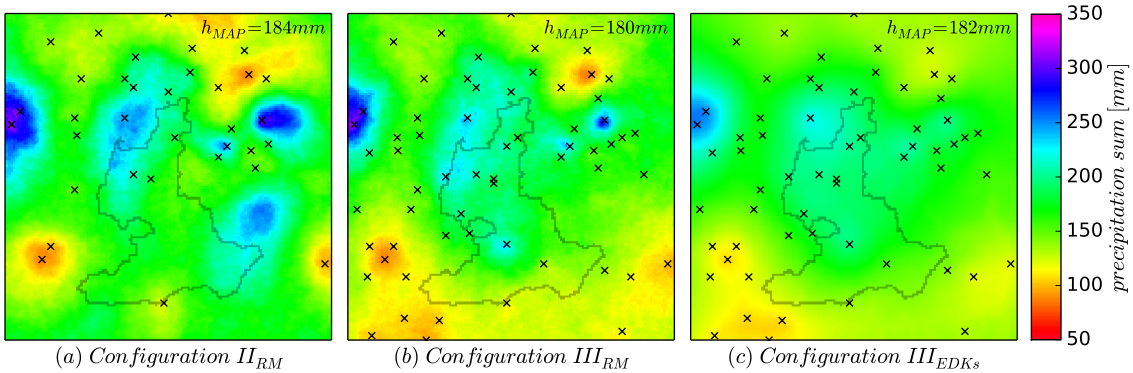
(a) Hydrological year (b) Hydrological half-year

Figure 4.13: Comparison of kriged and simulated daily mean areal precipitation depths (h_{MAP}) for configuration *II*.

fields for each time step are simulated within a moderate computing time. In addition, a fast hydrological model is required for processing many time series of simulated input data within a reasonable time. The SUPERFLEX model (Subsection 2.2.2) has these attributes and is chosen for this purpose.

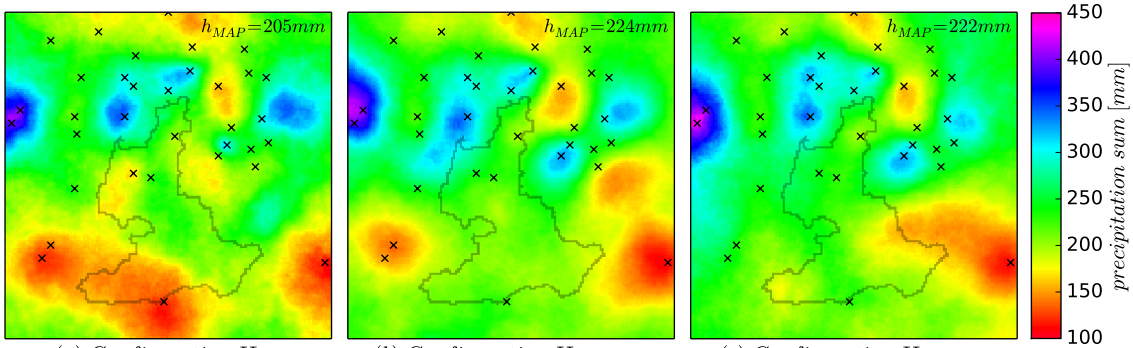
The time series of daily mean areal precipitation (h_{MAP}) for rain gauge configuration *II* generated by EDK_s and RM is compared in Figure 4.13. The data refer to the hydrological years and half-years from 2005 until 2012 and the values of h_{MAP} to the investigated catchment area. Both scatterplots show that the values of h_{MAP} simulated by RM are larger than the values generated by EDK_s for higher precipitation amounts. This tendency is appearing stronger for the winter half-years, even for smaller precipitation amounts. The mean annual precipitation sum of the catchment area is $1587\,mm$ if the precipitation is interpolated by EDK_s. Applying RM a mean annual value of $1784\,mm$ is calculated which is $11\,\%$ more than the value obtained with the conventional kriging method. Separating the time series by its hydrological half-years, the proportion of higher mean annual precipitation generated by RM is $8\,\%$ for the summer and $16\,\%$ for the winter half-years. Nevertheless, checking the accumulated precipitation sums of individiual precipitation events does not necessarily show the tendency to higher values of h_{MAP} for all cases. Figure 4.14 presents the accumulated seven-day precipitation patterns of a noteworthy flood event in June 2010. The maps are generated by RM and EDK_s for the rain gauge configurations *II* and *III*. The shown maps of RM are the mean of 100 realizations. Differences of the spatial structure are visible but the value of h_{MAP} hardly differs between the interpolation methods as well as the rain gauge configurations.

The potential error sources of measuring precipitation data are pointed out in Subsection 4.1.1. Regarding to the catchment topography and the distribution of rain gauges in the study area, a tendency of underestimated precipitation data can be assumed. This might



(a) *Configuration* II_{RM} (b) *Configuration* III_{RM} (c) *Configuration* III_{EDKs}

Figure 4.14: Accumulated seven-day precipitation patterns of RM and EDK_s for configuration *II* and *III*.



(a) *Configuration* II_{RM} (b) *Configuration* $II_{RM,\geq 1000}$ (c) *Configuration* $II_{RM,\geq 700}$

Figure 4.15: Comparison of accumulated five-day precipitation patterns of RM with and without *greater-equal* constraints for configuration *II*.

be especially the case for observation data of upper located stations. Thus, *greater-equal* constraints are applied to gauging stations above certain levels of elevation. The levels are defined by elevations above $1000\,m.a.s.l.$ and $700\,m.a.s.l.$

The effect of considering *greater-equal* constraints is analyzed for an accumulated five-day precipitation event causing the highest flood within the investigation period in June 2013. The resulting precipitation patterns, shown in Figure 4.15, vary considerable due to their spatial structure. The calculated value of h_{MAP} increases about $20\,mm$ considering *greater-equal* constraints. Especially, the higher located stations southwestward have a strong decreasing influence on the precipitation amount for the configuration without *greater-equal* constraints. Independently of the application of *greater-equal* constraints to all stations above $1000\,m.a.s.l.$ or $700\,m.a.s.l.$, the value of h_{MAP} is approximately equal. The precipitation pattern of configuration $II_{RM,\geq 700}$ is stronger influenced by a lower station located in the southeastern hinterland.

4.5 Summary

Firstly, the sources and potential quantities of uncertainty of the input variables are summarized in this chapter. The main driving forces of a hydrological model are precipitation and evapotranspiration. These data are often subject to systematic errors that are difficult to detect and correct satisfactorily in advance (*Bergström*, 1991). Thus, a careful quality control (QC) should be the first step before data are used for further processing. A QC avoids the use of apparently erroneous input data. Furthermore, the knowledge of potentially inadequate input information is helpful for the interpretation of the outputs in the model evaluation.

The analysis of the rain gauge data shows the difficulty of determining systematic measurement errors with sufficient certainty. Especially in such a study area, the relationship or trends of the precipitation patterns due to the highly variable topography are not easily distinguishable from systematic measurement errors and data inadequacies. Thus, the uncertainty of inadequate input data needs to be considered via other approaches. In this study, random mixing (RM) is applied to consider the uncertainty of precipitation data. Additional *greater-equal* constraints are included for different levels of station altitudes. This approach should account for a possible underestimation of precipitation amounts due to the exposed location of upper stations. The available temperature data indicate a good linear relationship between station altitude and mean annual values. Consequently, a spatial interpolation using external drift kriging (EDK) is reasonable. The QC of the discharge data shows no obvious inconsistencies of the time series at gauge *Unterjettenberg*. Some identified potential inadequacies in the time series refer to the other two available gauges upstream and downstream of the gauge *Unterjettenberg* and do not influence the hydrological modeling. Only, the time series of the year 2008 shows some irregularities which could be related to inaccurate observations of gauge *Unterjettenberg*. This has to be analyzed in the later evaluation of the model results.

In addition, the available meteorological point measurements must be transferred to a spatial representation. For this, conventional kriging methods and a conditional simulation approach are used for the interpolation of precipitation. The interpolation of temperature and evapotranspiration data is performed by EDK. The analysis of the interpolations using ordinary kriging (ODK) shows that the mean annual areal precipitation (h_{MAP}) is relatively similar for all three rain gauge configurations. Considering elevation as secondary information applying EDK or EDK_s leads in a considerable increase of h_{MAP} for rain gauge configuration *III* compared to the other two configurations. The application of RM generates higher values for the daily mean areal precipitation and consequently for the mean annual sum. Especially, for higher precipitation intensities and during the hydrological winter half-years an increase of h_{MAP} is observable using RM. Nevertheless, an event-based analysis of accumulated precipitation patterns shows that RM does not necessarily account for more h_{MAP} than the kriging methods. The modification of RM implementing *greater-equal* constraints allows to simulate higher precipitation amounts at the conditional points above certain altitude levels. The effect of including such constraints is exemplary shown for an accumulated intense precipitation event in Figure 4.15. There, the use of *greater-equal* constraints clearly increases h_{MAP} for such a single event. Thus, it can be a useful tool for

considering measurement uncertainties due to difficult measurement conditions of higher located observation stations.

The water balance has a noteworthy total deficit of $22\,\%$ on average for the *Saalach* catchment. This result is obtained regardless of the rain gauge configuration and the kriging method used. A deficit of less than $9\,\%$ is obtained calculating the water balance with the mean areal precipitation simulated by RM. Applying *greater-equal* constraints to stations above $1000\,m.a.s.l.$ reduces the deficit to $4\,\%$. This may be an indicator for partly inadequate observation data as well as a limited spatial representation by the station distribution and applied interpolation methods. Another reason for the deficient water balance seems to be an overestimation of the potential evapotranspiration (PE). Mean annual values of about $660\,mm$ for PE are estimated for the catchment area. Such an amount of PE as areal mean value is too high according to values found in literature (*BMU*, 2003) and reported in Subsection 3.1.2. In addition, the discrepancy between surface and subsurface catchment area can lead to a deficient water balance. Geohydrological properties such as karst forms can also influence the runoff characterisitics in the subsurface.

5 Effects of Model and Process Complexity

In this chapter, the choice of an appropriate level of model and process complexity is investigated and discussed. Usually, the more complex a model is, the more input information is required. Thus, the potential of overparameterization or equifinality problems due to an inadequate representation of the hydrological processes may also increase (*Perrin et al.*, 2001). In this study, two models with different levels of complexity are applied. The obtained results are presented in this chapter. This approach should support the investigation of structural model adequacy. A poor estimation of certain flow characteristics with similar input information appearing just in one model configuration can be an indicator for an imperfect model structure. In such a case, the distinction between model structural and input data error might be feasible. Such structural shortcomings can be caused by an inappropriate choice of model variables and processes, an inadequate process formulation or an insufficient spatial and temporal resolution of the model (*Del Giudice et al.*, 2015). The choice of an appropriate model complexity should be executed in a systematic approach. For example, the model complexity is increased stepwise that the positive as well as the negative effects of the modifications are clearly assignable (see Subsection 5.1.2). Such an approach supports the identification of a sound balance between model complexity and the quantity as well as the quality of available input data (*Schoups et al.*, 2008; *Fenicia et al.*, 2008).

In the following step, it is tested if outsourcing of certain flow periods to an external more complex process description supports a better representation of the model output. Here, the snow module of the SUPERFLEX model consisting of a lumped snow process representation is disabled. Instead, a highly spatially distributed snow model is applied externally prior to the actual model calculations. The stored and released water of the snow accumulation and melt processes is externally calculated and respectively removed or added to the precipitation time series and passed as a *liquid water* time series to the SUPERFLEX model.

Section 5.3 deals with model and process complexity due to computational effort. Depending on the modeling target, an appropriate temporal resolution of the model input and output is required. For instance, modeling flood processes in small to mid-size catchments needs a higher temporal resolution than the widely-used daily modeling time step. Nevertheless, modeling approaches with high temporal resolution also have a higher level of complexity in the sense of time-consuming pre- and postprocessing procedures as well as distinct longer computational times for ,e.g., hourly model calculations. The extent to which higher temporal resolution is beneficial in terms of improved model performance and robustness on the one hand, and processing time and effort on the other, is investigated.

5.1 Complexity of Hydrological Model

The model outputs gained by different complex model structures are investigated using the HBV-IWS and SUPERFLEX model introduced in Section 2.2. In the first part of this section, the differences in model complexity due to the model type are compared. The HBV-IWS model belongs to the more complex model types. Here, all input variables are distributed on a regular grid, whereas, the SUPERFLEX model has a lumped structure and uses mean areal values as input variables. In addition to the different spatial representations of input variables and hydrological processes, the models also vary in the required amount of input information. The second part of this section deals with the differing complexity representing the hydrological processes. For this purpose, the SUPERFLEX framework is taken. It has the ability to easily add and remove reservoirs accounting for different hydrological processes, such as interception, snow, slow and fast reacting outflows. By analyzing several reservoir configurations, it is checked whether the results of more complex model structures outperform the more simple approaches.

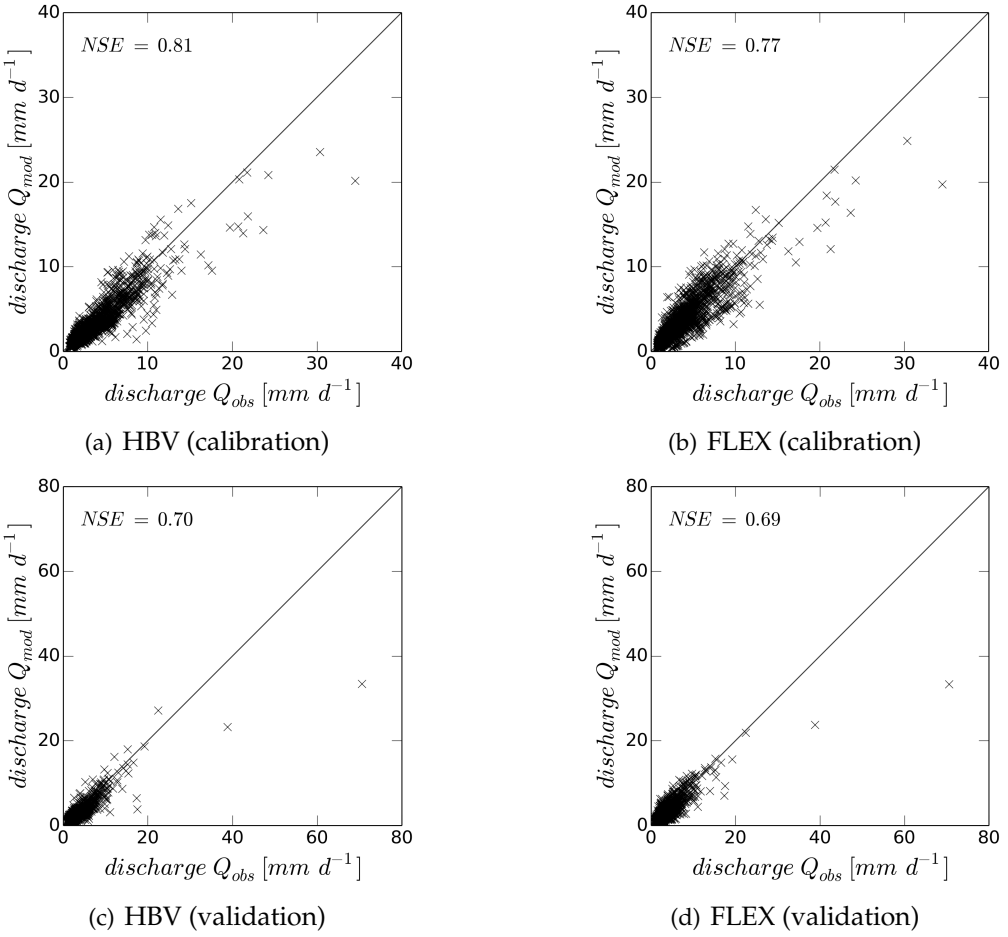
The representation of the model results in Chapter 5 and 6 of this study is divided in four parts due to different time periods and flow conditions. The model output and performance is generally analyzed on the *total* time series as well as on the *snowmelt* periods, the periods without snowmelt and some single flood events. The *snowmelt* period is assumed from the 1st of March until the 31th of May. The period without snowmelt is named *no snow* and disregards the assumed three months of *snowmelt*. The analysis of the *peak* flows contains the four highest observed flood events within the corresponding investigation time period (Figure 3.3). There are always two events part of the calibration as well as of the validation procedure.

In order to distinguish the performance of different model configurations, the commonly used Nash-Sutcliffe efficiency (NSE) is highlighted in bold in the following tables. Furthermore, the SUPERFLEX model is labeled as *FLEX* and the HBV-IWS just as *HBV* for a simplified representation.

5.1.1 Distributed or Lumped Model Structure

The comparison between the distributed and lumped model structure refers to the rain gauge configuration *I* and the total period from 2004 until 2013. The precipitation patterns are interpolated using ODK. The temporal resolution of the input and output variables is daily. The first year is used as warm-up period and is disregarded for the performance evaluation for both models. The models are calibrated for a period of five years (2005 - 2009) and validated for the subsequent four years (2010 - 2013). The FLEX model is run with the reservoir configuration *C1* (Table 5.3) and is calibrated using the *SLS* regression scheme.

In Figure 5.1, the results of the model runs generated by the HBV and the FLEX model are compared to the observed discharge data. Here, the *total* time series is considered separately by the calibration and validation period. The scatterplots show that both models tend to underestimate higher discharge conditions. The model results generated by the FLEX model



(a) HBV (calibration)

(b) FLEX (calibration)

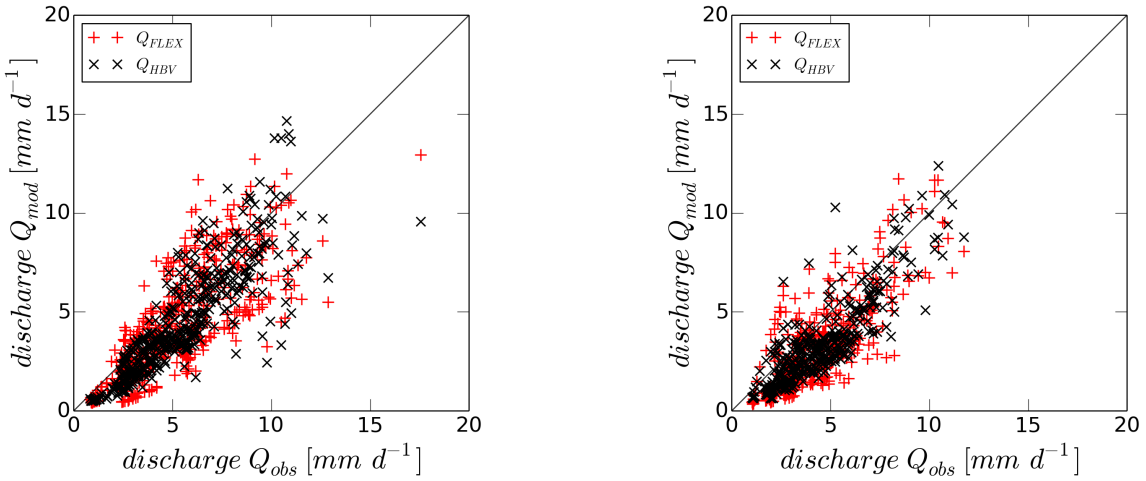
(c) HBV (validation)

(d) FLEX (validation)

Figure 5.1: Scatterplots of observed against modeled daily discharge for the total time series separated by the calibration and validation period.

indicate a greater spread of discharge values for the low and middle flow ranges. The model output for the *snowmelt* periods is shown in Figure 5.2. Both models mostly underestimate the discharge within the *snowmelt* period whereas the results of the HBV model fit slightly better for low flow conditions.

The corresponding model performance due to the different time periods and flow conditions is summarized in Table 5.1 and 5.2. The evaluation also includes the model results obtained for the precipitation inputs interpolated by EDK and EDK_s. The influence of different processed precipitation input data sets is marginal for the used rain gauge configuration on both models. Solely, the HBV model shows better performances for the *snowmelt* periods if EDK or EDK_s are applied. The *total* time series is represented slightly better by the HBV model according to the NSE. R^2 and $RMSE$ are in the same range for both models. Looking at the $PBIAS$ and KGE_m, the FLEX model outperforms the HBV model. Especially, the validation results of FLEX show very poor outputs for the *snowmelt* period. Here, the HBV model performs distinctly better than the FLEX, even though the *snowmelt* periods are mod-



(a) Snowmelt (calibration)

(b) Snowmelt (validation)

Figure 5.2: Comparison between observed and modeled daily discharge of the snowmelt period for the HBV and the FLEX model.

eled on a comparatively low performance level with both modeling approaches. The *peak* flows are considerably underestimated by both models, especially in the validation period (Figure 5.3). Nevertheless, the model performance of the *total* time series indicates good results in the calibration period for both models. But, the performance measures also show a clear gap between the results of the calibration and validation. Figure 5.4 shows the model behavior between the daily discharge Q at time t and the subsequent time step $t + 1$. This visualization represents the model's ability of reproducing the discharge dynamics within two successive days. The scatter of the model results has always a narrower spread than the scatter of the observed discharge data. This indicates that both models have problems to account for the ranges with high discharge dynamics.

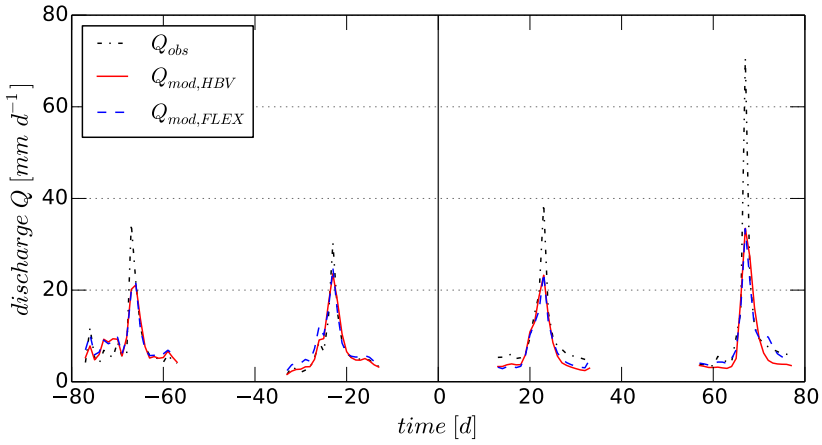


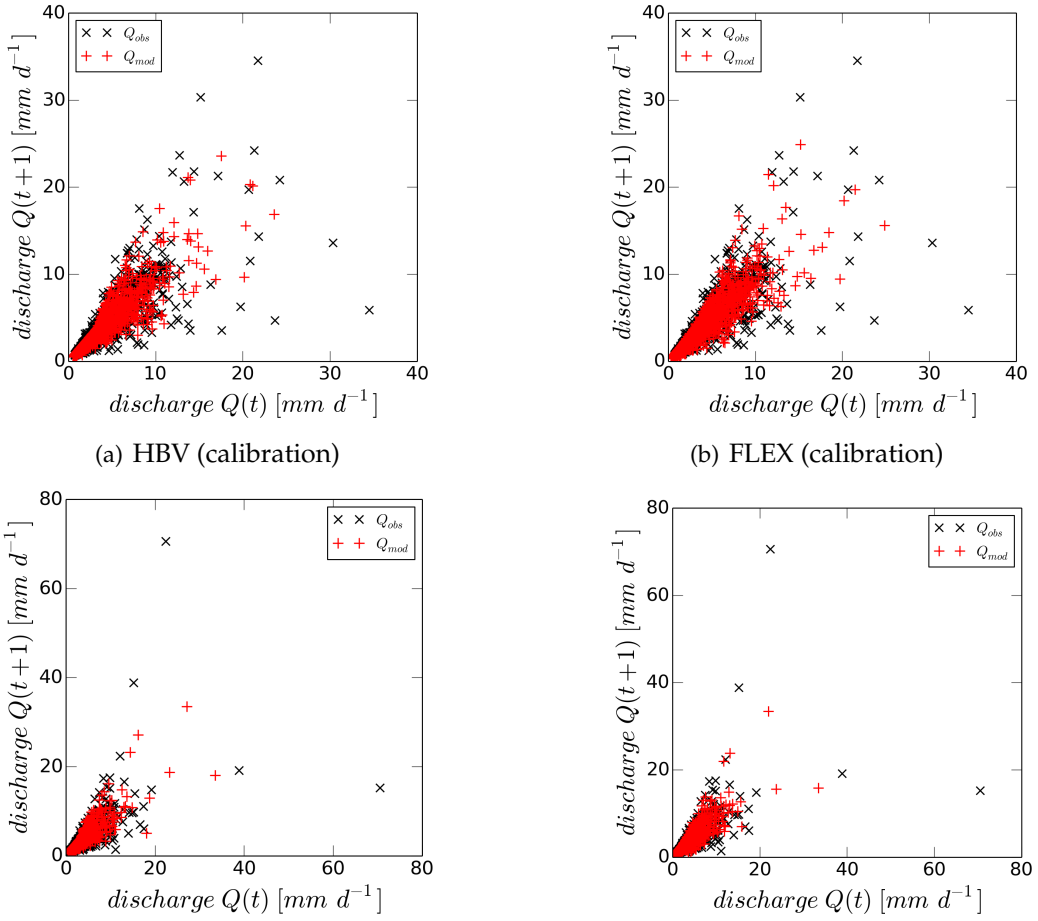
Figure 5.3: Hydrographs of the highest observed flood events reproduced by the HBV and the FLEX model (d < 0: calibration and d > 0: validation).

Table 5.1: Model performance of the HBV model evaluated by different measures for rain gauge configuration *I*.

cal.	total			snowmelt			no snow			peak		
	ODK	EDK	EDK_s	ODK	EDK	EDK_s	ODK	EDK	EDK_s	ODK	EDK	EDK_s
R^2	0.83	0.84	0.84	0.73	0.76	0.77	0.86	0.86	0.86	0.83	0.85	0.85
$RMSE$	1.3	1.2	1.2	1.8	1.6	1.6	1.1	1.1	1.1	2.9	2.8	2.7
$PBIAS$	11.8	8.4	8.1	18.4	13.8	12.7	7.9	5.2	5.4	-1.7	-3.2	-5.9
NSE	**0.81**	**0.82**	**0.83**	**0.57**	**0.65**	**0.66**	**0.85**	**0.85**	**0.86**	**0.81**	**0.82**	**0.83**
KGE_m	0.92	0.96	0.96	0.81	0.89	0.90	0.96	0.98	0.98	0.99	0.99	0.98
val.	total			snowmelt			no snow			peak		
	ODK	EDK	EDK_s	ODK	EDK	EDK_s	ODK	EDK	EDK_s	ODK	EDK	EDK_s
R^2	0.76	0.76	0.75	0.74	0.73	0.73	0.77	0.77	0.76	0.78	0.78	0.79
$RMSE$	1.7	1.7	1.7	1.6	1.5	1.5	1.7	1.7	1.8	6.6	6.7	6.7
$PBIAS$	20.8	19.4	19.4	24.0	22.8	20.7	19.3	17.9	18.7	25.3	26.1	26.6
NSE	**0.70**	**0.70**	**0.69**	**0.47**	**0.49**	**0.50**	**0.72**	**0.72**	**0.70**	**0.67**	**0.65**	**0.65**
KGE_m	0.77	0.79	0.79	0.69	0.72	0.76	0.80	0.82	0.81	0.67	0.64	0.63

Table 5.2: Model performance of the FLEX model evaluated by different measures for rain gauge configuration *I*.

cal.	total			snowmelt			no snow			peak		
	ODK	EDK	EDK_s	ODK	EDK	EDK_s	ODK	EDK	EDK_s	ODK	EDK	EDK_s
R^2	0.78	0.79	0.79	0.60	0.61	0.62	0.84	0.84	0.85	0.87	0.87	0.87
$RMSE$	1.4	1.4	1.4	2.0	2.0	2.0	1.1	1.1	1.1	2.9	2.9	2.9
$PBIAS$	4.5	4.1	4.3	14.2	14.1	14.2	-1.1	-1.6	-1.4	-8.1	-7.5	-8.5
NSE	**0.77**	**0.78**	**0.78**	**0.46**	**0.47**	**0.47**	**0.84**	**0.84**	**0.85**	**0.81**	**0.81**	**0.82**
KGE_m	0.98	0.98	0.98	0.85	0.85	0.85	0.99	0.99	0.99	0.96	0.97	0.96
val.	total			snowmelt			no snow			peak		
	ODK	EDK	EDK_s	ODK	EDK	EDK_s	ODK	EDK	EDK_s	ODK	EDK	EDK_s
R^2	0.72	0.70	0.70	0.62	0.58	0.58	0.76	0.75	0.75	0.86	0.85	0.86
$RMSE$	1.7	1.8	1.8	1.8	1.9	1.9	1.6	1.7	1.7	6.4	6.7	6.8
$PBIAS$	13.2	14.5	15.0	24.9	26.5	26.8	7.7	9.0	9.6	21.5	24.9	25.7
NSE	**0.69**	**0.67**	**0.67**	**0.27**	**0.21**	**0.21**	**0.75**	**0.73**	**0.73**	**0.69**	**0.65**	**0.65**
KGE_m	0.89	0.87	0.86	0.64	0.59	0.59	0.95	0.94	0.94	0.76	0.68	0.66



(a) HBV (calibration)

(b) FLEX (calibration)

(c) HBV (validation)

(d) FLEX (validation)

Figure 5.4: Scatterplots of discharge Q(t) and Q(t+1) for observed and modeled daily discharge.

5.1.2 Different Reservoir Compositions of SUPERFLEX

In a next step, different levels of complexity within the SUPERFLEX framework are investigated (Table 5.3). The four applied configurations have four to six different reservoirs. The combinations of active reservoirs are chosen in a physically meaningful way. Thus, all model set-ups account for processes due to interception (S_I) and snow (S_W) as well as slow (S_S) and fast (S_F) reacting flow components are considered. The model results of the basic reservoir configuration *C1* are already introduced in the previous section. The other three configurations differ in the alternating involvement of the unsaturated soil reservoir (S_U) and the riparian reservoir (S_R). The configuration *C4* includes all available reservoirs and is consequently the most complex model set-up. The FLEX model is calibrated with the regression schemes SLS and WLS for all investigated reservoir compositions.

The following analysis refers to the rain gauge configuration *I* and the precipitation patterns are interpolated by EDK_s. The model performance is evaluated for the *total* time series and

Table 5.3: Reservoir compositions of the FLEX model. The cross denotes the involvement of the individual reservoir.

conf. ID	number of reservoir states	number of active parameters	reservoir modules S_I	S_W	S_U	S_F	S_S	S_R
$C1$	6	22	X	X		X	X	
$C2$	7	23	X	X		X	X	X
$C3$	7	27	X	X	X	X	X	
$C4$	8	28	X	X	X	X	X	X

for an explicit examination the measures NSE (Table 5.4) and KGE_m (Table 5.5) are used. The tables also show the mean values (avg) and the absolute differences (Δ) of the performance measures for the calibration (cal) and validation (val) period. For a better comparability, the different FLEX reservoir compositions are ranked (RK) according to their mean performance (avg) as well as the absolute performance difference (Δ) between calibration and validation run.

The evaluation of both performance measures indicates a distinct gap between the results of the calibration and the validation for all reservoir configurations. The values of NSE are all in a good range for the calibration but the results of the validation are just in a satisfactory order. Due to the absolute values of NSE, the calibration method using SLS performs better than the approach with WLS. The differences in model performance between the reservoir configurations are comparatively small. Checking the results for a more robust configuration which means a smaller gap between the performance of the calibration and validation, the set-ups of $C3$ and $C4$ calibrated with WLS outperform the other configurations. Analyzing the model results with the KGE_m which accounts more for a good fit of the water balance, the calibration method using WLS always outperforms the method using SLS. The more complex the FLEX reservoir compositions, the better the model performance tends to be.

Table 5.4: Model performance of different FLEX reservoir compositions and calibration methods evaluated by NSE.

	C1		C2		C3		C4	
	SLS	WLS	SLS	WLS	SLS	WLS	SLS	WLS
NSE_{cal}	0.78	0.76	0.79	0.76	0.79	**0.75**	0.80	**0.73**
NSE_{val}	0.67	0.62	0.68	0.62	0.69	**0.65**	0.69	**0.64**
NSE_{avg}	0.73	0.69	0.73	0.69	0.74	0.70	0.74	0.69
NSE_{Δ}	0.11	0.14	0.11	0.15	0.10	**0.10**	0.11	**0.09**
$RK_{NSE,avg}$	4	6	3	7	2	5	1	8
$RK_{NSE,\Delta}$	5	7	4	8	3	**2**	6	**1**

Table 5.5: Model performance of different FLEX reservoir compositions and calibration methods evaluated by KGE_m.

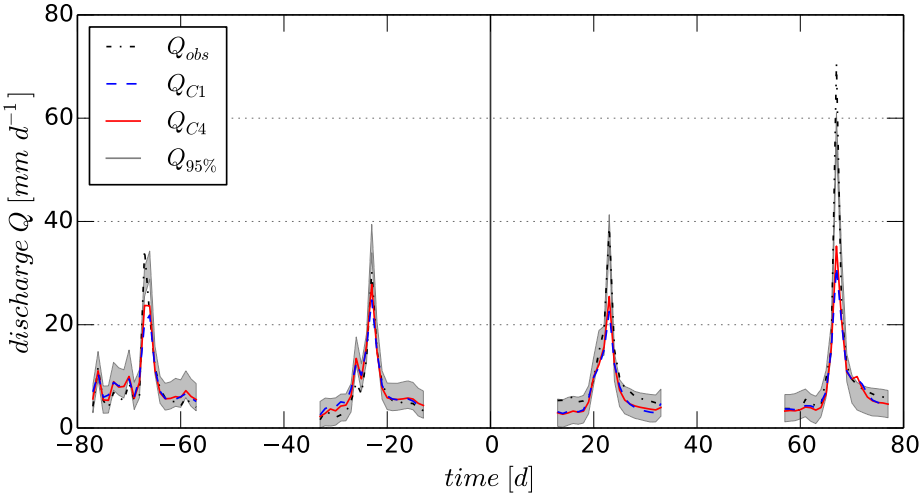
	C1		C2		C3		C4	
	SLS	WLS	SLS	WLS	SLS	WLS	SLS	WLS
KGE_{cal}	0.98	0.98	0.98	0.98	0.98	**0.98**	0.99	**0.98**
KGE_{val}	0.86	0.91	0.87	0.91	0.87	**0.93**	0.90	**0.94**
KGE_{avg}	0.92	0.95	0.92	0.95	0.92	**0.96**	0.94	**0.96**
KGE_{Δ}	0.12	0.07	0.11	0.07	0.11	**0.05**	0.09	**0.04**
$RK_{KGE,avg}$	8	4	7	3	6	**2**	5	**1**
$RK_{KGE,\Delta}$	8	4	7	3	6	**2**	5	**1**

5.1.3 Summary

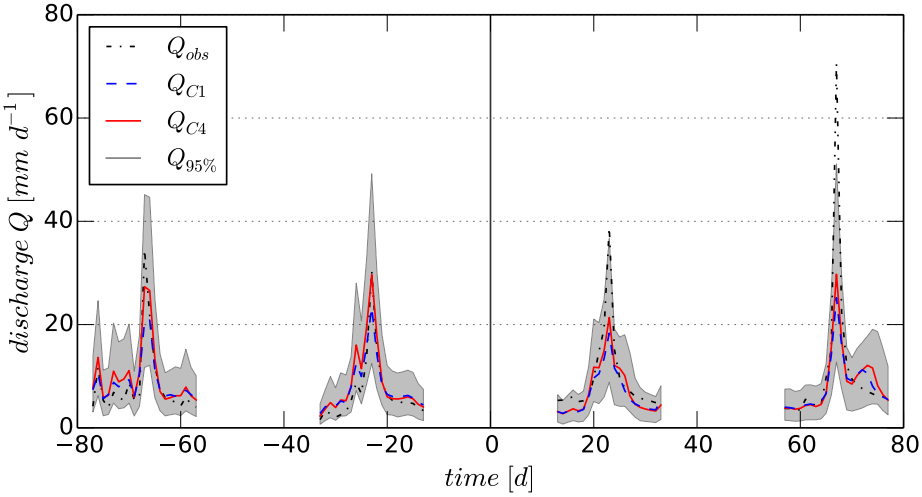
In this section, the purpose of the investigation is to determine whether model complexity influences model performance and robustness. Here, model complexity distinguishes between distributed or lumped model types on the one hand and different complex reservoir compositions of the FLEX model on the other hand. Furthermore, the model results give a first impression on which performance level the basic applications range and which shortcomings are identified.

Independently of the used model structure and reservoir configuration, a distinct gap of model performance between calibration and validation is distinguishable. The model performance for the calibration periods always indicate good model results while the validation periods show a reduction of about $10\,\%$ based on the NSE. Both models tend to overfit during calibration, so robust model results cannot be achieved. The analysis of different time periods and flow conditions reveal distinct shortcomings representing the *snowmelt* periods and *peak* flows in all model set-ups. Thus, inadequacies concerning the model structure cannot be distinguished at this point. Furthermore, a more complex modeling approach due to a spatially distributed consideration of the input variables with the HBV model does not necessarily show an improved and more robust model performance.

The investigation of different reservoir compositions of the FLEX model suggests that an increased complexity of model set-up tends to slightly better and more robust model results. Especially, if the WLS regression scheme is applied for the calibration. The reservoir configuration *C4* matches the *peaks* better than the less complex compositions for both regression schemes (Figure 5.5). Thereby, the model runs calibrated by the SLS regression scheme generate higher *peaks*. The regression scheme SLS prefers the fitting of high flows, whereas, WLS also puts a substantial weight to the low flows by approximating the heteroscedasticity of the errors (*Kavetski and Fenicia*, 2011). Figure 5.5 shows that the *peaks* are better captured by the uncertainty bands using WLS.



(a) FLEX (SLS)



(b) FLEX (WLS)

Figure 5.5: Peak flow hydrographs for different reservoir compositions and regression schemes of the FLEX model ($Q_{95\,\%}$: Uncertainty bands of $FLEX_{C4}$).

The analysis of the relative volume error (EV) and peak error (EP) is represented in Table 5.6. The EV of the *total* flow period indicates similar results for all three considered model configurations. Only the modeling results of set-up $FLEX_{C4}$ calibrated by WLS show a slight overestimation of the observed discharge volume looking at the *total* time period of the calibration. All other results indicate a volume underestimation for the *total* calibration as well as the *total* validation period. The validation period even has a larger volume deficit. Due to the absolute values of EV, the model results obtained by both FLEX set-ups present smaller volume deficits than the results of the more complex HBV model. The results of the *snowmelt* period modeled by a lumped structure are comparable to the distributed approach. All considered model set-ups underestimate the discharge volume for the calibration and the validation period. Nevertheless, the EV of the *total* and the *snowmelt*

period are in a similar range for the HBV model. The FLEX model set-ups show a clear decline for the *snowmelt* periods compared to the EV of the *total* period. Regarding the *peaks*, the EV of the calibration period presents a volume overestimation for all three model configurations whereas the modeled *peak* volume of the validation time period underestimates the observed *peak* volume in all cases.

Looking at the peak error EP, the FLEX calibration runs show better results for the *peaks* than the HBV model. But, the clear decline in model performance between calibration and validation results is again visible. Here, model set-up $FLEX_{C4}$ shows the poorest results due to robustness.

Table 5.6: Relative volume (EV) and peak (EP) error for different model structures and flow conditions.

	HBV		$FLEX_{C4}\ (SLS)$		$FLEX_{C4}\ (WLS)$	
	cal.	val.	cal.	val.	cal.	val.
EV_{tot}	0.08	0.19	0.02	0.13	-0.04	0.07
EV_{snow}	0.13	0.21	0.16	0.32	0.11	0.24
EV_{peak}	-0.06	0.27	-0.09	0.25	-0.23	0.19
EP	0.73	0.50	0.79	0.55	0.87	0.47

In summary, all investigated model structures show the same shortcomings like a lack of robustness between calibration and validation period, inadequate representation of the *snowmelt* periods and a distinct underestimation of the *peak* flows. Possible reasons for some of these observed inadequacies might be the quality of available input data, e.g., the overall water deficit of the system, systematic precipitation measurement errors of certain gauging stations or an overestimated assumption of the potential evapotranspiration (see Section 4.2). This can result in parametrization and overfitting problems of the models. The sensitivity of model parameters due to the temporal model resolution can also be a problem. Consequently, other approaches regarding the description of certain hydrological processes, data properties and temporal resolution are taken into account to investigate potential deficits in the next steps.

5.2 Process Complexity - External Snow Modeling

The first section shows that the representation of *snowmelt* processes is insufficient by both investigated models. Especially, the results obtained by the FLEX model configurations indicate a strong decline of model performance between calibration and validation (Table 5.2). Calibration and validation period show a model bias towards underestimation as indicated by the values of the $PBIAS$. The analysis of the relative volume error (EV) shown in Table 5.6 also denotes a deficit of water volume during the *snowmelt* periods. The lumped representation of snow accumulation and melt processes as used in the FLEX model seems to be unsatisfactory. Therefore, the outsourcing of the snow processes to a spatially distributed external snow model is tested in this section.

5.2.1 General

Modeling snow processes is very complex and is often underrepresented in common hydrological models. Thus, snow modeling can be seen as a separate discipline in hydrology and ranges from sole snow pack modeling to complete snowmelt-runoff models (SRM). Generally, temperature-based and energy balance snowmelt models of varying complexity are available. The calibration of snow models is a challenge due to the typical lack of adequate data such as snow depth, values of the snow water equivalent and well-distributed temperature data. Further difficulties of snow modeling are the drift of snow and different intense ablation processes depending on shadowing, slope inclination or orientation (*Hock*, 2003). Different studies of *Braun and Lang* (1986), *Blöschl and Kirnbauer* (1991), *Kirnbauer et al.* (1994), *Franz et al.* (2010) and *Kult et al.* (2014) introduce different complex snow model approaches and investigate the resulting uncertainties for hydrological predictions.

Regarding lumped hydrological models, such as the used FLEX model, a major drawback is the uniform consideration of the input variables for the whole catchment area. Snow accumulation and melt processes are also represented uniformly. This causes problems in catchments with a high topographic variability. The high range of elevations leads to meteorological conditions where snowfall and rainfall can occur at the same time within a catchment. Consequently, the simultaneous occurrence of snow accumulation and snowmelt influences the water availability for runoff generation. Also, the temporal extension of snow cover at lower elevations is usually shorter. Thinner snow packs melt faster than dense snow packs at higher elevations. Another issue regarding snow modeling is the temporal resolution. Smaller catchments require input data and modeling time steps on a higher temporal resolution (*Braun and Lang*, 1986). A significant increase or decrease of the temperature with a simultaneous appearance of precipitation can result in changing physical states during one time step. For modeling snow, e.g., at daily resolution, the total precipitation amount may be assumed to be rain or snow even though a change of the state has occurred within that time step. Such meteorological conditions within a time step determine if available water is stored as snow and result in reduced runoff conditions or the reverse case of fast melting snow due to so called rain-on-snow (ROS) events and the involved fast and large availability of water generating high runoff conditions.

5.2.2 Degree-Day Approach

The applied distributed external snow model (ESM) uses the degree-day approach equally to the snow routine in the HBV-IWS model. As no spatial observations of snow depth or snow water equivalent exist, an automatic calibration of the snow model using observation data for the parameter estimation is not feasible. For this reason, the used ESM is calibrated manually through several runs with varying sets of snow parameters in a prescribed range (Table 2.1). Subsequently, the *liquid water* time series is passed to the FLEX model for the calibration and validation procedures. The snow reservoir (S_W) is disabled. Finally, the best performing set of snow parameters is taken for further investigations.

The ESM is based on a regular grid with a spatial resolution of 500 m by 500 m. The distributed values of *liquid water* are averaged over the whole catchment area and transferred as

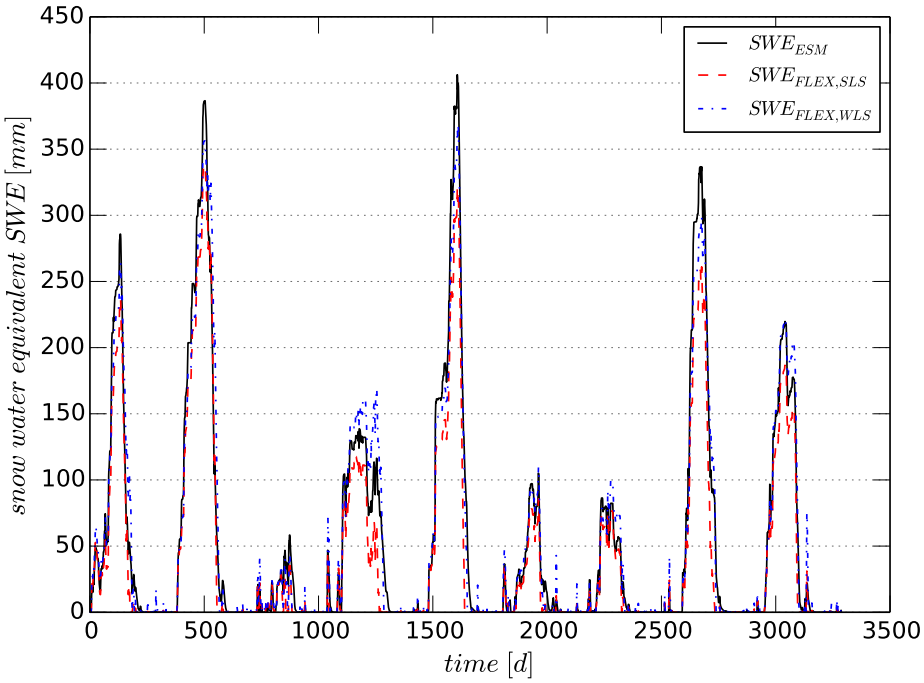


Figure 5.6: Snow water equivalent time series of the external snow model (ESM) and two runs of the FLEX model.

mean areal precipitation values (h_{MAP}) to the FLEX model. The scheme of the snow model is described with the subsequent equations where the threshold temperature (T_0) controls if it rains or snows.

Snowmelt occurs, if $T > T_0$:

$$\begin{aligned} S_{melt} &= Min\{SWE, DD \cdot (T - T_0)\} \\ SWE &= SWE - S_{melt}, \end{aligned} \tag{5.1}$$

otherwise snow is accumulated, if $T \leq T_0$:

$$\begin{aligned} SWE &= SWE + P \\ S_{melt} &= 0 \\ P &= 0, \end{aligned} \tag{5.2}$$

where S_{melt} is the amount of snowmelt, SWE the snow water equivalent, P the precipitation (all in mm) and T the temperature ($^{\circ}C$). The degree-day factor (DD) controls the melting rate of snow in mm per $^{\circ}C$ and time step t.

Alpine catchments are prone to floods caused by ROS events that occur mainly at the end of the winter season. Rainfall on an existing snow pack releases additional energy through long wave radiation and latent heat inputs. This can generate enormous discharge conditions due to the simultaneous occurrence of rainfall and the mobilization of stored water in

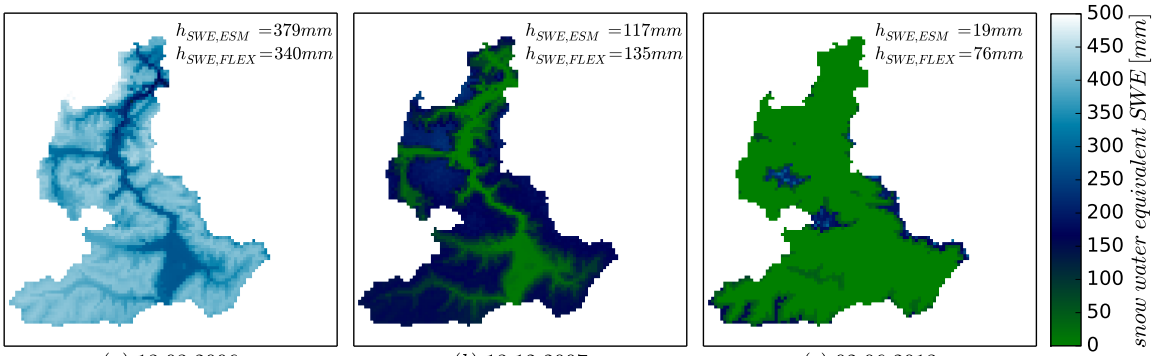
the snow pack (*Merz and Blöschl*, 2003). Therefore, a modified degree-day factor (DD_m) is incorporated to the snow model which accounts for enhanced melting rates during ROS:

$$DD_m = DD + DD_w \cdot P \tag{5.3}$$

DD is the degree-day factor for periods without rainfall and DD_w is a weight which controls an increased snowmelt rate during ROS events (*Bárdossy and Singh*, 2008).

5.2.3 Results

The modeling of *liquid water* time series and the subsequent application of the FLEX model is performed on daily resolution for this analysis. Rain gauge configuration I interpolated by EDK_s and the reservoir composition *C4* are used for the model runs. The FLEX model is calibrated using both regression schemes SLS and WLS. The evaluation of the FLEX model performance is compared between the internal snow processing of reservoir configuration *C4* and the best performing configuration of ESM. The best obtained parameter set for the ESM is $T_0 = 0.5\,°C$, $DD = 2\,mm\,d^{-1}\,°C^{-1}$ and $DD_w = 0.02\,d^{-1}\,°C^{-1}$.



(a) 13.03.2006 (b) 13.12.2007 (c) 02.06.2013

Figure 5.7: Comparison of accumulated spatial snow patterns generated by an external snow model (ESM).

Different generated time series of SWE for the years 2005 - 2013 are shown in Figure 5.6. Partly, the SWE courses differ considerably from each other. The ESM mostly calculates the highest SWE during the winter periods, while the FLEX configuration calibrated by SLS has the smallest amounts. The temporal distribution of SWE is quite similar for all three runs. The complete melting of snow cover last a bit longer for the ESM. The version $FLEX_{WLS}$ usually assumes much higher SWE during the summer months, when short periods of snow cover occur. Figure 5.7 compares three spatial patterns of SWE under different boundary conditions in terms of their seasonal occurrence and duration of accumulation. The spatial snow distributions are calculated with the ESM. The mean areal snow water equivalent (h_{SWE}) resulting from the two approaches to modeling snow processes

($FLEX_{WLS}$ and ESM) is also given in the figure. The left figure shows the snow distribution in late winter just before the beginning of snowmelt in one of the snowiest winters during the investigation period. The whole catchment area is covered with snow. Nevertheless, the underlying topography with the wide river valleys is very visible. The accumulated SWE is partly very high, which is also reflected in the values for h_{SWE}. The distribution of the first noteworthy formed snow cover in the early winter of a year is given in the middle. Here, the SWE is much lower and the lowlands are free of snow. Looking at these two events, it is apparent that h_{SWE} does not differ too much between the two modeling approaches. Furthermore, the selected events also show that one approach does not generally assume higher h_{SWE} than the other. The third pattern presents the snow situation that resulted in the highest observed flood event during the total investigation period. The most significant differences between the two model approaches occur here. The ESM only represents noteworthy SWE at the highest elevations of the catchment, which is also reflected in a relatively low value for h_{SWE}. In contrast, the internal snow model of FLEX assumes a much higher value for h_{SWE}, which appears to be very high as an mean areal value for this time of the year.

Table 5.7: FLEX performance using the internal snow routine or a *liquid water* time series.

cal.	total				snowmelt				no snow			
	SLS		WLS		SLS		WLS		SLS		WLS	
	$C4$	ESM	$C4$	ESM	$C4$	ESM	$C4$	ESM	$C4$	ESM	$C4$	ESM
R^2	0.80	0.84	0.76	0.83	0.60	0.78	0.59	0.79	0.88	0.87	0.83	0.86
$RMSE$	1.3	1.2	1.5	1.2	2.0	1.6	2.0	1.5	1.0	1.0	1.3	1.1
$PBIAS$	2.0	3.9	-3.8	-0.1	16.3	14.0	10.6	13.7	-6.2	-1.8	-12.1	-8.0
NSE	**0.80**	**0.84**	**0.73**	**0.83**	**0.46**	**0.66**	**0.42**	**0.68**	**0.87**	**0.87**	**0.79**	**0.85**
KGE_m	0.99	0.99	0.98	0.99	0.82	0.89	0.89	0.89	0.98	0.99	0.92	0.96
val.	total				snowmelt				no snow			
	SLS		WLS		SLS		WLS		SLS		WLS	
	$C4$	ESM	$C4$	ESM	$C4$	ESM	$C4$	ESM	$C4$	ESM	$C4$	ESM
R^2	0.72	0.79	0.65	0.76	0.49	0.62	0.60	0.63	0.80	0.84	0.68	0.82
$RMSE$	1.7	1.6	1.8	1.6	2.1	2.0	1.8	1.9	1.5	1.4	1.8	1.5
$PBIAS$	12.7	15.5	7.2	12.2	31.7	32.0	24.0	30.1	3.9	7.8	-0.6	3.9
NSE	**0.69**	**0.74**	**0.64**	**0.73**	**0.02**	**0.17**	**0.28**	**0.23**	**0.77**	**0.82**	**0.68**	**0.80**
KGE_m	0.90	0.87	0.94	0.91	0.41	0.44	0.66	0.50	0.98	0.96	0.97	0.98

Table 5.7 and 5.8 show the model results evaluated by different performance measures. The used *liquid water* time series generated by the ESM generally outperforms the FLEX version with its internal snow routine for calibration. The values of NSE are higher for the *total* period. The *snowmelt* period shows an improved NSE of more than $20\,\%$. The other performance measures indicate a similar trend for the calibration as well as validation period

using a *liquid water* time series as model input. The more volume-based $PBIAS$ partially leads to a slightly different result. Independently of the snow process consideration and the applied FLEX error model, the $PBIAS$ is in the same range for all time periods and flow conditions. The distinct gap of about $11-12\,\%$ (*total*), $13-18\,\%$ (*snowmelt*) and $10-12\,\%$ (*no snow*) between the calibration and validation runs is also visible. The $PBIAS_{tot}$ has values close to zero and thus similar observed and modeled volumes for calibration. The modeled values during validation tend to a volume underestimation for all set-ups. The $PBIAS_{snow}$ indicates a stronger underestimation of the modeled volume for calibration and validation. This consistent behavior for all set-ups at least suggests that no discharge volume disappears in certain model reservoirs through the model parameterization. In conclusion, the $PBIAS$ returns a slightly worse performance overall for the data processed by the ESM. Generally, an improved model performance is obtainable for the used *liquid water* time series due to the absolute values of the measures. Nevertheless, a distinct performance gap between calibration and validation remains. Looking at the NSE, the gap between the calibration and validation becomes less for the *total* and for the *no snow* periods compared to the results of the HBV model and the basic FLEX reservoir configuration, as shown in Table 5.1 and 5.2. There, the difference in model performance expressed by the NSE varies between $11-16\,\%$ for calibration and validation. The application of the ESM reduces this gap to $5-10\,\%$ for the *total* and for the *no snow* periods. The model runs with a *liquid water* time series also show slightly better results based on the performance gap than the model setup with an internal snow reservoir ($C4$).

Table 5.8: Evaluation of peak flows for different considered snow processes.

	$C4_{SLS}$		ESM_{SLS}		$C4_{WLS}$		ESM_{WLS}	
	cal.	val.	cal.	val.	cal.	val.	cal.	val.
EV_{peak}	-0.09	0.25	-0.13	0.23	-0.23	0.19	-0.15	0.24
EP	0.79	0.55	0.84	0.63	0.88	0.47	0.79	0.58
NSE_{peak}	0.88	0.72	0.89	0.80	0.82	0.60	0.86	0.75

Table 5.8 summarizes the analysis of the *peak* flows. The EV_{peak} reflects an overestimated modeled discharge volume for calibration and a distinct underestimation during validation for all set-ups. The evaluation of the *peaks* regarding the EP and the NSE_{peak} shows that the model runs using input data preprocessed by the ESM outperform the FLEX version with an internal snow reservoir, especially in validation. Figure 5.8 presents that configuration $C4_{WLS}$ generates higher *peaks* as well as a greater overestimation of the previous discharge hydrograph for calibration. The second *peak* of the calibration is quite well fitted by both approaches, while the model output using the *liquid water* time series generates higher *peaks* for both validation events. The second *peak* validated by $C4_{WLS}$ shows a second smaller *peak* during the descending flood event. The second *peak* can be explained by the fact that the model accumulates too much snow at the time of the main flood event (Figure 5.7 (c)) and releases it as discharge with a delay. These circumstances indicate that this event is influenced by snow and is better represented if the snow processes are calculated spatially distributed in advance.

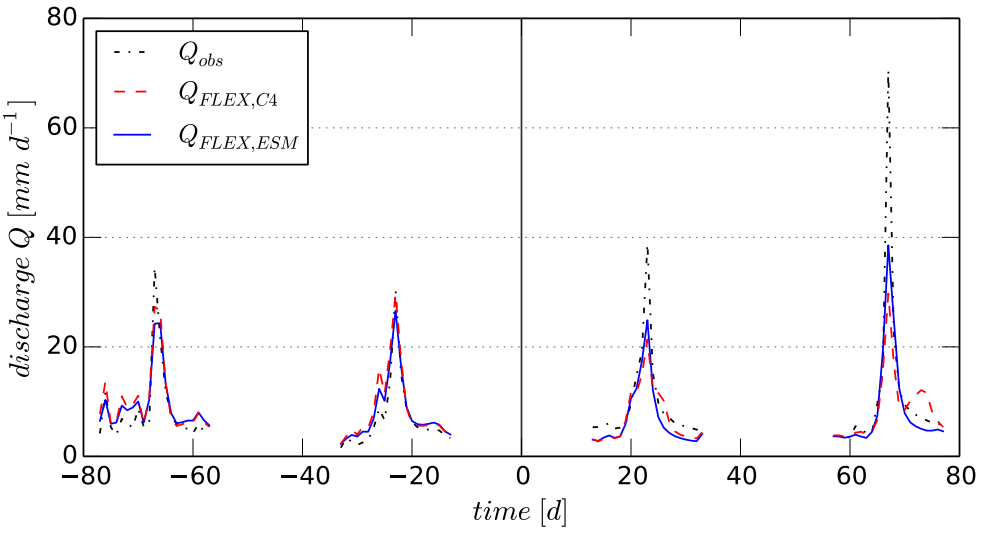


Figure 5.8: Peak flow hydrographs modeled by $FLEX_{WLS}$ with an internal snow routine (C4) and a *liquid water* time series (ESM).

The preceding application of an external distributed snow model generally leads to improved model performances. However, the reproduction of *snowmelt* processes can still be classified as unsatisfactory due to the robustness. The *no snow* periods indicate more robust model results compared to the internal consideration of snow processes of FLEX as well as to the model output of the HBV model analyzed in Subsection 5.1.1. The representation of *peak* flows is more promising using an external processed *liquid water* time series due to robustness and maximum discharge in validation. The temporal distribution of SWE during the investigation period (Figure 5.6) shows considerable differences between the results of the FLEX configurations and the ESM. ESM usually accumulates more snow during the winter months than the internal snow routine of FLEX. The presentation of spatial snow patterns and the comparison of the mean areal h_{SWE} support the assumption that FLEX accumulates too much snow during shorter events occurring outside the winter months. Spatially distributed snow processes seem to be beneficial for a more accurate consideration of variable topography and consequently improves model outputs. The issue of changes from snow to rain and the reverse case due to strong temperature changes around the threshold temperature during a day are considered in Chapter 7. Such changes of precipitation types can occur if the temperature increases above the threshold value for a couple of hours during noon or if a front system leads to a significant jump in temperature. Therefore, it is analyzed whether an hourly model resolution describes such snow processes more accurately.

5.3 Computational Effort - Different Temporal Resolutions

Depending on the modeling target, an appropriate temporal resolution of the model input and output is required. Thus, it has to be clarified which kind of catchment and discharge processes are of main interest. For instance, in the case of estimating discharge volumes for

reservoir management of water supply or hydro power lower temporal resolutions such as weekly or even monthly may be sufficient. In this study, the focus is on a robust parameter estimation for flood predictions. For such modeling purposes, higher temporal resolutions are generally necessary. Many studies applying hydrological models are based on a daily time step. For the dimensioning and operation of flood control reservoirs, model results on a daily resolution can be sufficient if the time to peak and the volume of a flood event are more important than the absolute peak height. But, for modeling flood events in small to mid-size catchments higher temporal resolutions are required. In such catchments, hydrological processes causing flood events mostly occur within less than a day. Maximum discharge conditions typically last only a few hours, which often results in considerable smoothing of the peak discharge for data in daily resolution. However, exactly the absolute maximum discharge is needed as information for flood protection measures and potential evacuation plans if the investigation area is prone to inundations.

In the first part of this section, the daily model output of the FLEX model is aggregated to lower temporal resolutions such as 2, 5 and 10 days. The aim of this analysis is to check if the aggregation of the model output to lower temporal resolutions improves the model performance and robustness. Available hourly data from precipitation and temperature stations are sparse due to their quantity and spatial density. Therefore, in a second approach, daily precipitation data are disaggregated to an hourly resolution. In this way, it is investigated whether the model performance and robustness benefit from densifying the measurement network with disaggregated data from daily stations as additional information. Both approaches have a higher level of process complexity in the sense of time-consuming pre- and postprocessing procedures and distinct longer computational times for the hourly model calculations. Against this background, it has to be investigated in which sense such an effort is justified regarding improved model performance and robustness.

5.3.1 Temporal Aggregation of Model Output

The results of aggregating the daily FLEX model output to temporal resolutions of 2, 5 and 10 days are exemplary shown on rain gauge configuration *II*. The reservoir composition *C4* and the regression scheme WLS is applied. Table 5.9 shows the resulting model performance measures for the different aggregation time steps.

Except for the $PBIAS$, the investigated performance measures show improving results for higher aggregation time steps looking at the *total* period. The $PBIAS$ remains on a constant value for all aggregation levels. The model bias of the calibration period tend to a slight overestimation and the results of the validation to an underestimation. The gap between the model performance of the calibration and validation decreases with increasing aggregation time step. Here, the model performance described by R^2 and NSE improves by 4 % for the *total* calibration period and about 10 % for the *total* validation period from a daily to an aggregated ten days resolution. The KGE_m shows a quite constant performance level independently of the temporal resolution and evaluated time period. Looking at different aggregation time steps for the *snowmelt* period, the trend of the used performance measures is not that clear and consistent. R^2 and KGE_m show constant values for all ag-

Table 5.9: FLEX performance evaluated by different measures and aggregation time steps.

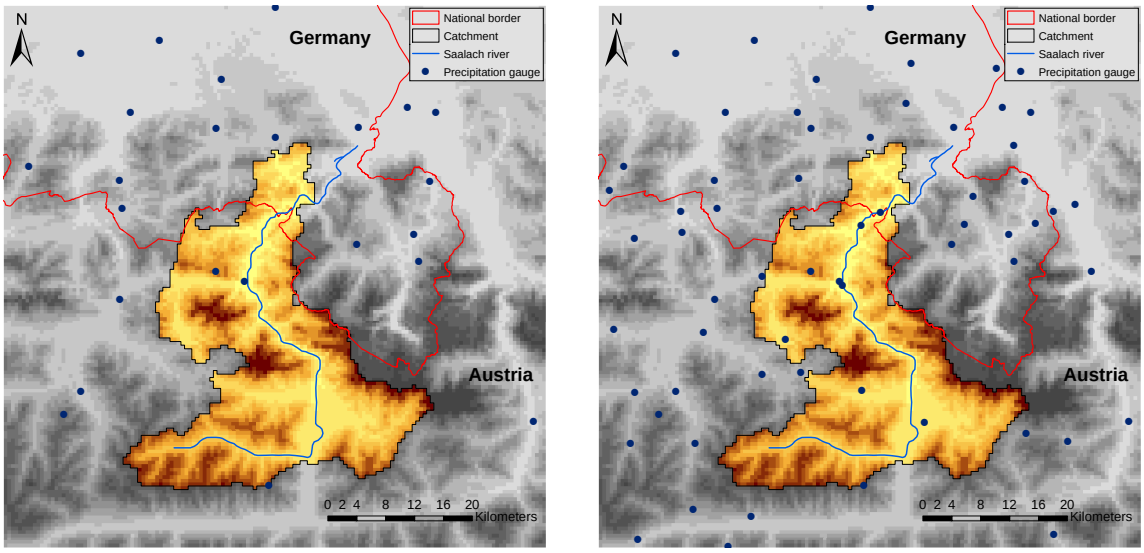
	total				snowmelt				no snow			
cal.	$1d$	$2d$	$5d$	$10d$	$1d$	$2d$	$5d$	$10d$	$1d$	$2d$	$5d$	$10d$
R^2	0.76	0.77	0.79	0.80	0.59	0.59	0.58	0.59	0.83	0.85	0.88	0.91
$RMSE$	1.5	1.4	1.3	1.1	2.0	2.0	1.9	1.7	1.3	1.2	1.0	0.8
$PBIAS$	-3.8	-3.8	-3.8	-3.7	10.7	10.8	11.2	10.9	-12.1	-12.2	-12.4	-13.1
NSE	**0.73**	**0.74**	**0.76**	**0.77**	**0.43**	**0.41**	**0.35**	**0.31**	**0.79**	**0.80**	**0.83**	**0.85**
KGE_m	0.98	0.98	0.98	0.98	0.89	0.89	0.88	0.89	0.92	0.92	0.92	0.91
	total				snowmelt				no snow			
val.	$1d$	$2d$	$5d$	$10d$	$1d$	$2d$	$5d$	$10d$	$1d$	$2d$	$5d$	$10d$
R^2	0.65	0.68	0.71	0.76	0.59	0.64	0.65	0.83	0.68	0.72	0.76	0.76
$RMSE$	1.8	1.6	1.3	1.1	1.8	1.8	1.7	1.7	1.8	1.6	1.2	0.8
$PBIAS$	6.8	6.8	6.8	6.8	23.6	23.2	22.9	21.9	-1.1	-1.0	-0.9	-2.3
NSE	**0.64**	**0.67**	**0.70**	**0.74**	**0.28**	**0.37**	**0.36**	**0.68**	**0.68**	**0.72**	**0.76**	**0.70**
KGE_m	0.94	0.95	0.95	0.96	0.67	0.69	0.70	0.75	0.97	0.98	0.98	0.98

gregation levels for the calibration period, while the model performance for the validation period clearly improves with increasing temporal aggregation. In contrast, the NSE shows a decreasing trend for the calibration period and an increasing trend for the validation period. The $RMSE$ indicates a slight improvement of the model performance for higher aggregation levels for the calibration as well as the validation. Regarding the $PBIAS$, the model tends to underestimate the discharge during the *snowmelt* period for the calibration and even more for the validation. Especially, R^2, NSE and KGE_m have a distinct jump of model performance between the aggregation level of 5 and 10 days for the validation. These results show that interpreting the model performance describing the snow processes is not straightforward. Snow processes are very complex and influenced by many factors. Thus, errorenous snow accumulation can already cause a considerable deficit of stored discharge volume which would confirm the magnitude of the $PBIAS$. The model bias tends to an underestimation during the *snowmelt* period and an overestimation during the *no snow* period. *Snowmelt* is a temporal extensive process. Melting water can be stored in the remaining snow pack and released at a later date. The *snowmelt* rate also varies depending on the depth and density of the snow pack. Such factors are difficult to capture at higher temporal resolutions and could explain the jump in model performance at an aggregation time step of 10 days.

5.3.2 Temporal Disaggregation of Model Input

It is now investigated whether a higher temporal resolution provides more information for the hydrological catchment dynamics. An hourly input data and computation time step of

the model is chosen. A higher temporal resolution may be necessary for representing fast reacting discharge conditions such as *peak* flows which occur within a few hours in the study area. A better representation of precipitation state changes from snow to rain and vice versa during a day can also provide an improvement of the model results.



(a) Hourly stations (Config. IV) (b) Disaggregated stations (Config. III_h)

Figure 5.9: Different rain gauge configurations of stations with hourly and disaggregated hourly data.

As described in Section 3.2 and shown in Figure 5.9, the availability and consequently the density of gauging stations with hourly precipitation data is more sparse than on daily resolution. For this reason, it is tested if additional information of disaggregated precipitation data of daily gauging stations leads to an effect for model performance and robustness. The disaggregation of the daily precipitation time series is performed with the *nearest-neighbor* method. The assumption is made that the hourly precipitation distribution at the daily rain gauge station is similar to that of the nearest hourly station. First, for each individual daily station of rain gauge configuration *II* (10 extra stations) and *III* (37 extra stations), the geographically nearest hourly station is determined using the Euclidean distance. Then, the relative precipitation distribution over the respective day is derived from the hourly precipitation values. For this purpose, the hourly value is divided by the daily precipitation sum, which is calculated from the hourly data. Finally, the relative hourly precipitation values are multiplied by the daily precipitation sum from the nearest daily gauging station. In this way, a disaggregated hourly time series is obtained for each individual daily station. For rain gauge configuration *II*, the minimum distance to the nearest hourly station is $2.9\,km$, the maximum $10.4\,km$ and the average distance is $6.0\,km$. The minimum distance to the nearest hourly station is $0.7\,km$, the maximum $20.6\,km$ and the average is $9.0\,km$ for rain gauge configuration *III*. Configuration *III* has greater maximum and average distances between a daily and the nearest hourly gauging station. This can be explained by the fact that considerably more daily stations are available in the central part of the investigation area.

Table 5.10: Hourly model performance of the FLEX model evaluated by different measures for rain gauge configuration IV, II_h and III_h.

cal.	total			snowmelt			no snow		
	IV	II_h	III_h	IV	II_h	III_h	IV	II_h	III_h
R^2	0.77	0.78	0.79	0.57	0.58	0.58	0.84	0.86	0.88
$RMSE$	0.07	0.07	0.07	0.09	0.09	0.09	0.06	0.05	0.05
$PBIAS$	-4.5	-4.1	-5.2	10.3	9.0	8.6	-13.3	-12.0	-13.3
NSE	**0.72**	**0.73**	**0.74**	**0.38**	**0.36**	**0.39**	**0.78**	**0.81**	**0.80**
KGE_m	0.97	0.98	0.97	0.89	0.90	0.91	0.91	0.92	0.91
val.	total			snowmelt			no snow		
	IV	II_h	III_h	IV	II_h	III_h	IV	II_h	III_h
R^2	0.72	0.75	0.75	0.60	0.61	0.60	0.76	0.78	0.80
$RMSE$	0.07	0.07	0.07	0.08	0.07	0.08	0.07	0.07	0.07
$PBIAS$	9.5	8.9	-0.3	24.7	21.1	13.6	2.5	3.2	-6.7
NSE	**0.71**	**0.74**	**0.74**	**0.30**	**0.36**	**0.33**	**0.76**	**0.78**	**0.79**
KGE_m	0.93	0.94	0.98	0.64	0.73	0.86	0.98	0.98	0.97

There are neither daily nor hourly gauging stations available for the other configurations (Figure 3.4 and 5.9). Thus, the central catchment area remains underrepresented if daily gauging stations of configuration *II* are included.

The notation of the hourly rain gauge configuration with 24 stations is *IV*. The name of the hourly rain gauge configurations including disaggregated daily data is identical to the daily versions added with an index h. The mean observed discharge ($Q_{obs,m}$) is $0.15\,mm\,h^{-1}$ for the calibration as well as for the validation period. The FLEX model is run with reservoir configuration *C4* (Subsection 5.1.2). Regression scheme WLS is applied for optimization. The results in Table 5.12 refer to the hydrographs of the four highest *peak* flows during the chosen investigation period shown in Figure 5.3.

Table 5.10 shows the model performance evaluated by different measures resulting from the three investigated hourly rain gauge configurations. The evaluation includes different flow periods. Regarding all performance measures, the three different rain gauge configurations perform almost similar for the calibration of the *total* period. Solely, for the R^2 and the NSE the model performance gets slightly better with a larger quantity of considered rain gauge stations. The $PBIAS$ indicates a model bias toward overestimation for all input configurations. The $RMSE$ of all configurations is equal for the calibration as well as the validation period. Looking at the *total* validation period, the tendency of a better model performance with an increasing number of rain gauge stations is given for all measures except for the $RMSE$. The NSE suggests robust model results for all rain gauge configurations. Taking the measures R^2 and KGE_m into account, only configuration III_h indicates good results due to model robustness. The same trend is given for the $PBIAS$ where configuration

Table 5.11: Evaluation of peak flows for different rain gauge configurations of hourly precipitation data.

	IV		II_h		III_h	
	cal.	val.	cal.	val.	cal.	val.
EV_{peak}	-0.17	0.22	-0.15	0.21	-0.22	0.11
EP	0.97	0.63	0.93	0.63	0.85	0.78
NSE_{peak}	0.90	0.78	0.91	0.78	0.88	0.85

IV and II_h indicate a model bias toward underestimation. In contrast, configuration III_h shows a model bias toward a very slight overestimation, which is the same trend like the model results of the calibration period. Analyzing the other flow periods, the model performance of the *no snow* period shows the same tendency regarding the different configurations like the results of the *total* time series. All configurations even perform better with respect to the NSE and the R^2 than with the *total* time series. In contrast, the *snowmelt* period is still the flow characteristic with the poorest model representation. All configurations have the same level of performance and suggest model robustness for the *snowmelt* period due to the R^2 and $RMSE$. But, the values of NSE are on a low level for all three configurations with more or less large gaps between calibration and validation period. The $PBIAS$ of all configurations indicates a model bias toward underestimation for the calibration as well as the validation period. Configuration III_h clearly outperforms the other configurations due to model performance and robustness taking the $PBIAS$ and the KGE_m into account.

Table 5.11 compares the *peak* flow performance of the investigated hourly rain gauge configurations. The *peak* flows partly present large differences between calibration and validation. Only configuration III_h shows a model performance on a high and robust level regarding the NSE. The EV_{peak}, on the other hand, indicates a strong opposing trend between calibration (overestimation) and validation (underestimation) for all configurations. The analysis of the EP indicates the best results for the calibration in the descending order of configuration IV, II_h and III_h. Configuration III_h clearly outperforms the other configurations due to the absolute value of EP during validation and model robustness. The largest deficits of modeled discharge volume occur during the *snowmelt* period and the *peak* flows. Again, it is apparent that the *snowmelt* period and the *peak* flows are the weakest points represented by the model. The evaluation of the different configurations illustrate that the configuration IV and II_h perform on a same level. Especially, the *snowmelt* period is better represented by configuration III_h. Configuration III_h also has the smallest gaps between calibration and validation for all performance measures regarding the *peak* flows.

The time until the occurrence of a peak is an important information for flood protection measures of fast-reacting catchments with a response time of less than a day. The time to peak error (ET) is considered for a comprehensive performance evaluation on an hourly temporal resolution. In Table 5.12, the ET of the different configurations are listed. Regarding the ET, all hourly rain gauge configurations perform equally well. Except for the second *peak* of the validation ($d = 67$), where configuration IV consisting of only 24 hourly

Table 5.12: Time to peak error (ET) resulting from different rain gauge configurations for the four highest peak flows (hourly resolution).

	IV	II_h	III_h
$Q_{peak,1}$	3	3	3
$Q_{peak,2}$	1	1	1
$Q_{peak,3}$	0	0	0
$Q_{peak,4}$	-4	0	0

rain gauge stations estimates the *peak* four hours too soon. Both configurations considering additional information of disaggregated daily rain gauge stations match the *peak* of the validation period exactly. The two *peaks* during the calibration period are represented identical by all three configurations. In all cases, the model estimates the *peaks* with a delay of three hours or one hour, respectively.

5.4 Summary

In this study, two different model types with a different consideration of the input variables due to their spatial representation as well as the required amount of input information are investigated on the background of model complexity. Varying model complexity is analyzed by a flexible model structure which is stepwise increased in its complexity. In the first part of this chapter, the objective is to determine if more model process complexity necessarily generates a better model output. The decision to use a more complex approach should only be made if there are noticeable improvements in terms of model robustness and performance level.

The model results in Subsection 5.1.1 show that a more complex distributed approach (HBV) does not necessarily outperform a lumped model structure (FLEX). The results of the FLEX model against the HBV model demonstrate that good model performance can be achieved even with a simple model structure. The comparison between the two models and afterwards between different complex reservoir configurations of the FLEX model present a distinct gap of model performance between calibration and validation. The model performance of the calibration period indicates good model results in all cases. The validation period has a clear decline in model performance. Independently of the used model, the modeled discharge output tends to overfit the observed discharge during the calibration. Consequently, the estimated parameter sets do not generate robust model outputs in validation. The analysis of the different time periods and flow conditions shows that all model set-ups have distinct problems with representing the *snowmelt* periods and *peak* flows. Solely, the *snowmelt* period is reproduced slightly better with the HBV model where the input variables such as precipitation and temperature are considered spatially distributed. According to these findings, clearly identifiable inadequacies concerning the model structure cannot be distinguished for none of the two models. Also, an improved and more robust representation of

the hydrological processes is not obtainable with a more complex modeling approach such as the spatially distributed HBV model.

The investigation of different reservoir compositions of the FLEX model indicates that a more complex model set-up tends to slightly better and more robust model results. A more robust parameter estimation for the FLEX framework could be achieved if other hydrological processes in terms of additional reservoirs and parameters are considered. This may lead to an increased complexity but avoids overfitting the model during the calibration period. But, this improvement is achieved by an increase of 22 (simplest configuration) to 28 (most complex configuration) calibrated parameters. If the improvement of model output is just marginal it has to be balanced which model configuration is finally selected. An appropriate model choice must focus on a preferably parsimonious modeling approach on the one hand and the best level of model performance on the other hand. It can be concluded that all investigated model structures have the same shortcomings such as a lack of robust parameterization, inadequate representation of the *snowmelt* periods and a distinct underestimation of the *peak* flows. Thus, the use of a simple lumped model structure like the FLEX model seems to be feasible. Within the different complex lumped model configurations, the most complex model set-up promises better and more robust model outputs. Nevertheless, it has to be emphasized that an insufficient quality of required input variables during the investigation period is problematic to obtain a good parameterization and good fits, independently of the applied model structure. The fact of inadequate but fixed input variables (observations) is deeper investigated in Section 6.2. Based on these conclusions, other approaches regarding the description of certain hydrological processes, data properties and temporal resolution are taken into account in the subsequent sections.

As mentioned above, one clear drawback of both models is the representation of snow processes. The *snowmelt* period is poorly estimated, especially by the lumped FLEX model. The analysis of the model output indicates a deficit of water volume during the *snowmelt* periods and consequently the modeled discharge series underestimates the observations. A lumped model structure considers a mean value of temperature and precipitation per time step for the whole catchment. Therefore, just one condition can be represented. In catchments with very variable topography, different states of precipitation and consequently accumulation and melt of snow can occur at the same time. A third scenario can be so called rain-on-snow (ROS) events. Such different simultaneously occurring conditions cannot be represented by a lumped catchment representation. For this reason, the snow processes of the lumped model structure are released to a spatially distributed external snow model (ESM). The ESM uses the degree-day approach equally to the snow routine in the HBV model. The output of the ESM is transferred as a *liquid water* time series to the FLEX model. Then, the model is calibrated and validated with the common procedure. The separation of the snow processes from the lumped model structure to a prior distributed accounting of snow accumulation and melt shows noticeable improvements of the model results due to the absolute performance values. The use of a *liquid water* time series represents the *snowmelt* periods better. The results are still unsatisfactory due to the model robustness of these flow characteristics. The *no snow* periods as well as the *peak* flows are better represented by this approach regarding performance level and robustness. Especially during the winter months, the ESM accounts for more snow than the internal snow routine of the FLEX model. The

consideration of spatially distributed snow processes appears beneficial for the present investigation area with a very variable topography. Additionally, the amount of optimizing parameters is reduced by seven and the model complexity is decreased through the deactivation of the snow reservoir. An ESM increases the processing effort slightly but just uses three parameters in this case. Through the assumption of equal snow parameters for each grid cell and the non-consideration of some smoothing functions the parameter demand is less. An improved model performance can be obtained with an efficient substitution of the available input information without adding more parameters.

In the last part of this chapter, another kind of modeling complexity due to computational effort is discussed. Depending on the used temporal resolution, the temporal and computational effort for data preprocessing, modeling procedures and postprocessing can increase enormously. Thus, the modeling purpose determines the required temporal resolution. Lower temporal resolutions such as daily to monthly might be sufficient if short-time discharge dynamics are not important for the modeling objective. Modeling of flood events usually requires a temporal resolution of daily or even higher for representing the discharge dynamics. The subsequent aggregation of the model output to lower temporal resolutions analyzes if there is an improvement of the model output due to performance level and robustness. The modeling procedures are executed on a higher temporal resolution than the evaluation and presentation of the model results. This investigation shows that the performance level is generally better for higher aggregation time steps looking at the *total time* period. The gap between the model performance of the calibration and validation also decreases with increasing aggregation time step. The evaluation of the *snowmelt* period does not show such a clear and consistent trend for the performance measures. For some measures and lower aggregation time steps, the performance level decreases for the calibration time period and increases for the validation period. Thus, a clear trend of the model performance describing the snow processes is not recognizable. One reason of this behavior can be the complexity of snow processes and its many influencing factors. For instance, an errorenous snow accumulation can already cause a considerable deficit of stored discharge volume which is then missing during the *snowmelt* period. In summary, the analysis of the aggregated model output shows that there is an increase in performance level and even robustness for lower aggregation time steps. This can be due to the compensation of under- and overestimated discharge periods. One potential shortcoming of the model is the adjustment of the fast-reacting flow dynamics at higher temporal resolutions, which has a significant impact on model performance. This fact is smoothed out for aggregated time steps. An opposing trend of increasing model performance for the calibration and decreasing performance for the validation with increasing aggregation time steps can indicate that as well other error sources such as inadequate input variables, volume errors or wrong parameter estimates influence the model output.

The disaggregation of daily precipitation data to an hourly resolution deals with the question if the model application benefits by the use of additional input information due to performance level and robustness. Catchment areas with limited high-resolution gauging stations and consequently with a sparse spatial coverage suffer from insufficient point information. These conditions can result in inadequate spatial interpolations of the model input variables. The disaggregation of daily precipitation data is performed by the *nearest-*

neighbor method. Independent of the considered rain gauge quantity, the model output on an hourly resolution shows a more robust behavior between the calibration and validation. The *no snow* period has a distinct better model performance compared to the *total* time series for all considered configurations. The *snowmelt* period is still the flow characteristic with the poorest model representation due to the performance level. But, the model robustness is also improved for the *snowmelt* period. The *peak* flows still indicate the largest gap between the results of calibration and validation which is especially the case for the two configurations with less gauging stations. Taking all performance measures into account, it can be concluded that the configuration with most additional information outperforms the other configurations. Therefore, the consideration of additional information, through ,e.g., the disaggregation of available daily data to a higher temporal resolution, is beneficial for obtaining more robust model outputs.

In summary, a more complex spatially distributed model is not necessarily better than a simpler lumped model approach. The most complex reservoir configuration is suggested within the lumped model framework for obtaining better and more robust model outputs. The outsourcing of more complex hydrological processes such as snow modeling is recommended for a better spatial representation. Here, a lumped modeling framework has its limitations. In smaller catchments, such as the study area, a high temporal modeling resolution is necessary for catching the flow dynamics. Incorporating additional information by disaggregating lower resolution input data improves the level of performance and, in particular, the robustness of the model output. This can be a beneficial approach for catchment areas where input data on high temporal resolution are very rare.

6 Assessment of Data and Model Adequacy

Bergström (1991) stated that the confidence of models as powerful instruments are at risk if poor modeling ethics, not suitable model structures or inadequate model calibration and verification are applied. *Refsgaard and Henriksen* (2004) also argued that the reliability of models is often questioned when there is insufficient evidence of predictive ability. Thus, the interaction between data, model structure, parameter sets and predictive uncertainty has to be examined (*Beven*, 1993; *Kavetski et al.*, 2006b).

Chapter 4 gives a review of sources and ranges of uncertainty concerning the required input data. Measured data are often inaccurate to a certain extent and consequently the data quality has to be analyzed. Such a quality control avoids the use of apparently erroneous input data. The prior knowledge of potentially inadequate input information supports a differentiated evaluation and interpretation of the model output. Chapter 4 also introduces different spatial statistical methods for the interpolation of point information. Chapter 5 investigates and compares different complex model structures, preprocessing and data aggregation approaches. There, shortcomings concerning model robustness and performance between the different applied methods are just analyzed and evaluated. In Chapter 6, the possible reasons and potential impact of the identified shortcomings due to the data availability and quality as well as the application of different data preprocessing methods are presented. Finally, the model performance due to time-dependent and event-based criteria as well as the model parameterization is evaluated.

6.1 Input Data Analysis

Models are a simplification of a real natural system. Nevertheless, models cannot be operated without observed data. In this context, *Silberstein* (2006) remarked appositely:

> *"Data are science, models are a complement to them, but not a replacement for them."*

This citation evokes the need for sufficient data collection and their quality control. In the following subsections, the influence of data quantity and quality on the model robustness and performance is analyzed and discussed. Chapter 5 shows that the results of widely accepted and used hydrological models suggest a good representation of the reality due to the performance level of the *total* calibration periods. But, a careful validation of the calibrated models with different preprocessed input data as well as a comprehensive analysis of several performance measures on different discharge periods and characteristics indicate distinct shortcomings of robust model outputs. Consequently, the potential to use such a

model for reliable predictions is limited. These findings also confirm the citation above that models are not able to compensate every lack of missing information due to incomplete or inadequate input data.

6.1.1 Rain Gauge Density and Distribution

The first indications of potential data shortcomings due to the quality and spatial availability in the investigation area are already shown in Chapter 3. Orographic effects at the north side of the Alps indicate increasing mean annual precipitation depths with increasing station altitude. Table 3.1 compares the mean annual precipitation for four rain gauge stations on different elevations. The simple presentation of the measured data confirms the expected trend for three considered stations until an elevation of about $1000\,m.a.s.l.$ The highest available station on almost $2000\,m.a.s.l.$ has a clear decrease of approximately $400\,mm$ of mean annual precipitation compared to the station on $1000\,m.a.s.l.$ This station is also the most southern and very solitary located station of rain gauge configuration I and II (Figure 3.4). Especially, such a fact can have a distinct influence on model calibration and finally on model robustness and performance if such a station provides poor data quality. Other basic conditions, such as number of available stations, spatial station density, vertical distribution and the minimum distance between two stations (Table 6.1) may also influence an adequate model calibration. An insufficient density of rain gauge networks can be a problem. In particular, regions with a high variability in precipitation (e.g., due to a very heterogeneous topography) and a sparse observation network can cause large uncertainty if areal precipitation is estimated by interpolation.

Table 6.1: Basic information of the daily rain gauge configurations due to quantity and spatial distribution.

description	unit	daily rain gauge configuration			
		IV_d	II	I^*	III
total number of stations	$[-]$	24	34	59	61
station density of investigation area	$[1/1000\,km^2]$	1.1	1.5	2.6	2.7
station density of catchment area	$[1/1000\,km^2]$	3.2	4.3	7.5	10.8
minimum distance of two stations	$[km]$	3.9 - 28.4	2.9 - 28.4	0.8 - 23.6	0.7 - 13.0
average minimum distance	$[km]$	8.7	6.9	4.5	5.6
average station altitude	$[m.a.s.l.]$	774	748	807	767

In this subsection, the FLEX model output of four different rain gauge configurations is compared. The basic conditions listed in Table 6.1 show considerable differences in the total number of stations, available stations in the catchment area, the minimum distance between

Table 6.2: Flex performance for different discharge periods using varying daily rain gauge configurations as precipitation input (calibration: 2005 - 2009, validation: 2010 - 2012).

	total				snowmelt				no snow			
cal.	IV_d	II	I^*	III	IV_d	II	I^*	III	IV_d	II	I^*	III
R^2	0.74	0.76	0.74	0.78	0.58	0.59	0.58	0.58	0.81	0.83	0.80	0.85
$RMSE$	1.6	1.5	1.6	1.4	2.1	2.0	2.1	2.0	1.4	1.3	1.4	1.1
$PBIAS$	-4.5	-3.8	-3.9	-1.6	11.5	10.7	9.8	10.8	-13.6	-12.1	-11.7	-8.6
NSE	**0.71**	**0.73**	**0.71**	**0.77**	**0.42**	**0.43**	**0.40**	**0.45**	**0.75**	**0.79**	**0.76**	**0.84**
KGE_m	0.97	0.98	0.97	0.98	0.88	0.89	0.89	0.88	0.90	0.92	0.92	0.96
	total				snowmelt				no snow			
val.	IV_d	II	I^*	III	IV_d	II	I^*	III	IV_d	II	I^*	III
R^2	0.63	0.67	0.66	0.73	0.45	0.51	0.50	0.59	0.70	0.73	0.73	0.79
$RMSE$	1.6	1.5	1.5	1.3	2.0	1.8	1.9	1.6	1.4	1.3	1.3	1.2
$PBIAS$	6.0	5.9	7.5	1.6	25.8	22.1	23.8	13.2	-3.1	-1.6	0.0	-3.8
NSE	**0.58**	**0.64**	**0.63**	**0.73**	**0.06**	**0.19**	**0.16**	**0.41**	**0.67**	**0.72**	**0.72**	**0.78**
KGE_m	0.94	0.95	0.94	0.98	0.56	0.67	0.63	0.86	0.97	0.98	0.98	0.98

a station pair, etc. Rain gauge configuration II and III (Figure 3.4) are already known from the former chapters. Configuration IV_d only includes 24 daily stations for which also hourly observations are available. The fourth analyzed rain gauge configuration I^* differs from configuration I in the total amount of 59 stations. Here, some additional stations from the national park *Berchtesgaden* are considered which are located south of the city of Berchtesgaden and are only clustered in the corner at the German side (Figure 3.1). These gauging stations have data until 2007. Consequently, the additional information is spatially and temporally very inhomogeneous. Table 6.2 shows the FLEX model performance of the different rain gauge configurations. As the additional stations of rain gauge configuration III have no data for the year 2013, the evaluation disregards this year for an objective comparison of the results.

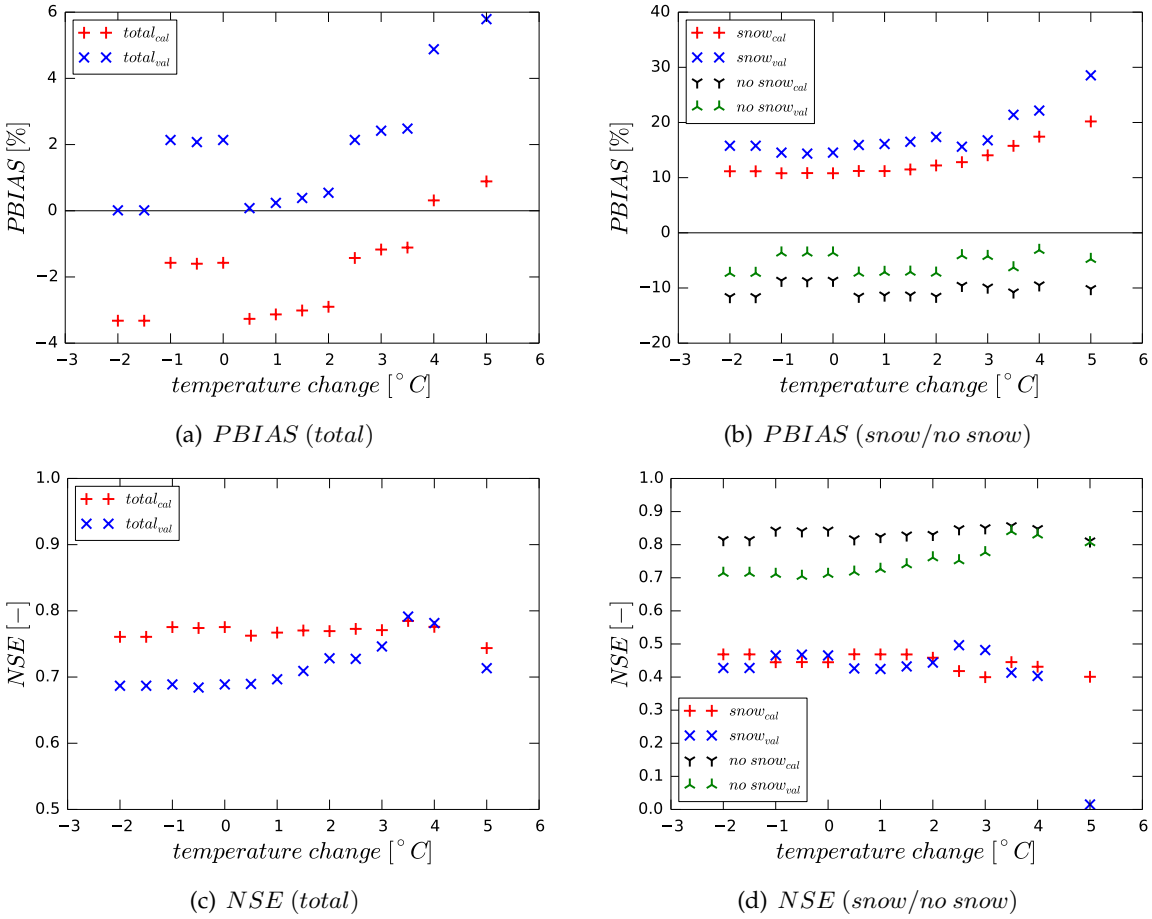
The different performance measures of the calibration indicate that there is no distinct trend for better model results due to an increased amount of gauging stations. Only rain gauge configuration III shows clearly better model results compared to the other three configurations for almost all performance measures. The improvements become particularly apparent for the *total* and the *no snow* periods. The differences in model performance of the *snowmelt* period are slight for all measures and rain gauge configurations. There are considerably more differences between the rain gauge configurations if the model output of the validation is analyzed. Here, an order due to more available stations is visible. The total evaluation indicates the worst model performance on a satisfactory to unsatisfactory level for configuration IV. Then, configuration II and I^* are at the same satisfactory level. Configuration

III shows again the best performance on a very good level. Solely the representation of the *snowmelt* period also remains on an unsatisfactory performance level. But, configuration III is the only one which generates robust model outputs for the *snowmelt* period as well as the other investigated flow periods.

This analysis demonstrates in which manner areal precipitation input resulting from sparse or unbalanced rain gauge networks influences the model performance. The range of available gauging stations varies between 24 up to 61 stations. The water balance of the different rain gauge configurations in Subsection 4.3.4 shows a distinct deficit of about $20\,\%$. This is a first indication for an insufficient quality of certain input data which can also cause problems in model calibration. There, configuration III with the most available gauging stations is already the most promising input data set due to the lowest water deficit (Table 4.4). These findings are confirmed by the model output where the results obtained by configuration III show by far the best and the only robust model performance. Looking at the calibration period, the model output of configuration IV, II and I^* (24, 34 and 59 stations) has very similar model performances. Despite the very different quantity and distribution of the gauging stations, the FLEX model is obviously able to compensate the lack of information through its parameterization. However, the results of the validation period reveal that the individual parameter sets are not appropriate for generating robust model outputs. The stations have to be well spatially distributed within the study area. This indicates the model output of configuration I^* and III. Both configurations have the same amount of gauging stations which also means an equal station density for the total investigation area. Taking now the station density of the catchment area as well as the minimum distance between the individual station pairs into account (Table 6.1), configuration III has a much denser and well distributed gauging network. Figure 3.4 also shows the irregularly distributed gauging network of configuration I where the southern part of the investigation area only contains very few stations. If the conditions mentioned above are fulfilled, as for configuration III, the quantity of available gauging stations has a positive effect on model performance and robustness.

6.1.2 Temperature Influence on Different Flow Conditions

In addition to precipitation, temperature is another important model parameter, which is particularly relevant for the correct representation of snow processes. The review of data availability and control in Sections 3.2 and 4.2 shows that with 22 stations, there are even less temperature stations available than rain gauge stations. Only ten of the 22 available temperature stations fulfill the quality criterion of a maximum of one year of missing data related to the entire investigation time period. The highest available temperature station is at $1973\,m.a.s.l.$ On the other hand, the catchment area reaches altitudes of $2600\,m.a.s.l.$ and has an average elevation of $1260\,m.a.s.l.$ The station density of $0.5/100\,km^2$ is very sparse within the catchment. A value of $R^2 = 0.99$ suggests a linear relationship between station altitude and mean annual temperature (Figure 4.4). However, a large range of altitudes ($800\,m.a.s.l.$ to $1600\,m.a.s.l.$ and above $2000\,m.a.s.l.$) is not covered with stations and sufficient measurement data. There are only three stations above $800\,m.a.s.l.$ that fulfill the requirements of the quality control. This large deficit in vertical and horizontal information



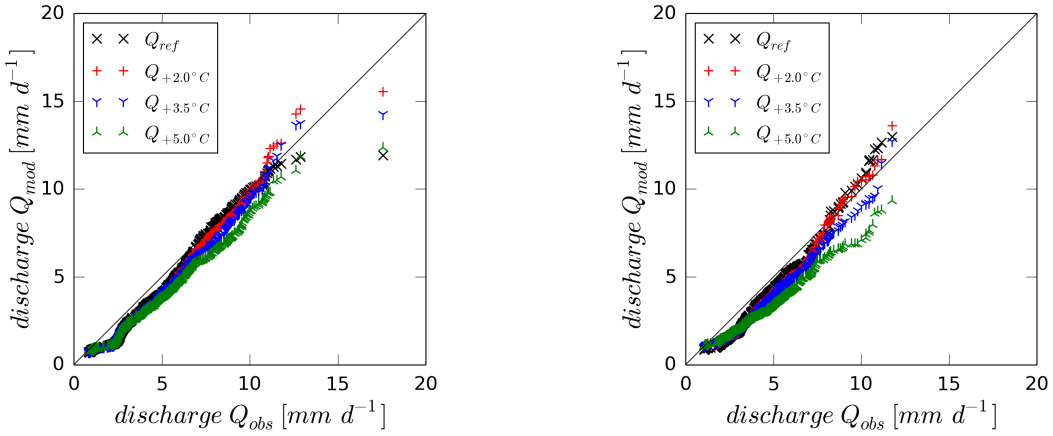
(a) $PBIAS$ $(total)$ (b) $PBIAS$ $(snow/no\ snow)$

(c) NSE $(total)$ (d) NSE $(snow/no\ snow)$

Figure 6.1: Model performance analysis of different modified temperature time series.

can lead to significant uncertainties in the interpolation of the spatially distributed temperature fields. Consequently, this can have a great influence on the subsequent averaging to derive the uniform temperature value as lumped model input. The major drawbacks of this uniform consideration of the temperature data, especially with respect to the snow processes, are already discussed in Section 5.2. In addition, weather conditions such as inversion or foehn effects can be mentioned here, which can reverse the temperature profile or cause distinctly different temperature values at same elevations at the same time within the catchment. Therefore, the influence of temperature on the various flow conditions is evaluated in more detail in this subsection. For the investigations, the lumped temperature time series is modified over the entire calibration and validation period (2004 - 2013), by reducing or increasing the daily value in $0.5\,°C$ steps. The temperature time series is reduced or increased in a range from $-2.0\,°C$ to $+5.0\,°C$. The precipitation time series is based on configuration III with data at all stations for 2013.

Figure 6.1 presents the evaluation of the individual model runs with the modified temperature time series using the performance measures $PBIAS$ and NSE. The left diagrams show the respective model performance for the *total* calibration and validation period. In



(a) snowmelt (calibration)

(b) snowmelt (validation)

Figure 6.2: Q-Q plots for assessing the predictive reliability of discharge for the snowmelt period of different modified temperature time series.

the diagrams on the right, the model performance is divided according to the flow conditions *snowmelt* and *no snow*. Regarding the $PBIAS$, the model results have no clear tendency of improvement or deterioration for the *total* period and for the *no snow* period. The $PBIAS$ is not very sensitive to temperature changes in a range of $1.0\,°C$ to $2.0\,°C$ and then changes quite abruptly. Furthermore, the gap in model bias between calibration and validation remains nearly constant. This difference increases slightly from a temperature change of $+4.0\,°C$. With respect to the *total* modeling period, a $PBIAS$ close to zero during validation always assumes a model bias toward overestimation during calibration. However, the general tendency for the *total* period is that the discharge is overestimated during calibration and underestimated during validation. Looking at the $PBIAS$ for the different flow conditions, it can be seen that for both calibration and validation, the model bias always tends to underestimate the discharge for *snowmelt* and overestimate it for *no snow*. The model bias is generally slightly smaller and relatively insensitive to the temperature changes for the *no snow* conditions. For *snowmelt*, the model bias towards underestimation increases from a temperature change of $+2.0\,°C$. In addition, the difference of the $PBIAS$ between calibration and validation is slightly smaller at a temperature deviation of $+2.5\,°C$ and $+3.0\,°C$. In summary, no noteworthy changes in model performance for modified temperature time series in a range between $-2.0\,°C$ and $+3.5\,°C$ can be identified using the $PBIAS$. The NSE is relatively constant during calibration for the *total* modeling period. A clear drop in performance occurs from a temperature deviation of $+5.0\,°C$. The NSE reproduces an increasing trend of the model performance during validation in a range from $+1.0\,°C$ to $+4.0\,°C$. This also reduces the gap between calibration and validation and is even equal for a temperature deviation of $+3.5\,°C$ and $+4.0\,°C$. The course of the NSE during the *no snow* period is very similar to the *total* modeling period at a slightly higher performance level. Due to the $PBIAS$ and NSE, a reduction in daily mean temperatures does not appear to cause a notable change in model performance. The application of modified temperature data also shows no clear positive effect for the *snowmelt* period.

Figure 6.2 assesses the predictive reliability of some selected configurations during the *snowmelt* period using Q-Q plots. Again, it can be seen that especially the smaller to medium discharges tend to be underestimated for all configurations during calibration and validation. No model output with modified temperature time series provides an improvement to represent *snowmelt* conditions. Only the configuration with a temperature change of $+2.0\,°C$ has a comparable distribution to the reference configuration. The two configurations with larger temperature deviations tend to underestimate the discharge distribution more clearly during calibration and validation.

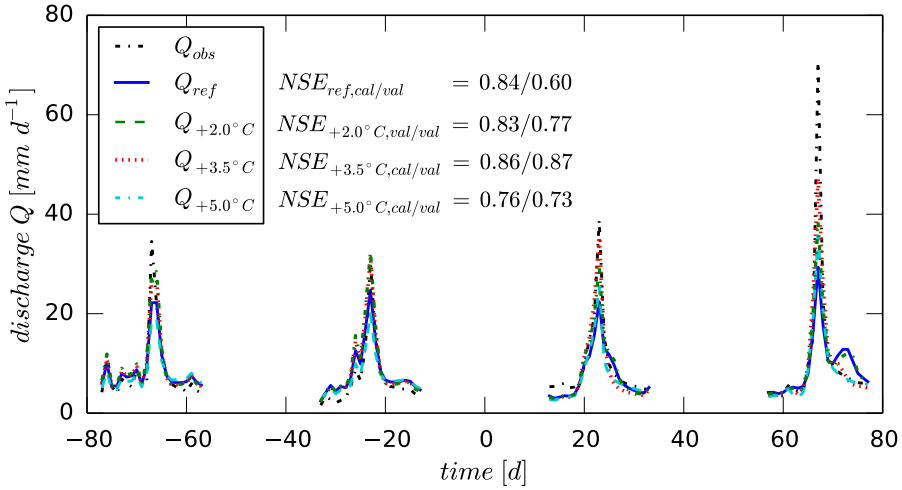


Figure 6.3: Peak flow hydrographs for different modified temperature time series.

Finally, the model representation of the *peak* flows for the selected configurations is shown in Figure 6.3. The most significant influences from the application of the modified temperature time series are seen here. Regarding the NSE, the model performance for validation is improved for all three configurations with modified temperature time series. Here, configuration $Q_{+3.5\,°C}$ provides the best and most robust model results for the *peaks*. The configurations $Q_{+2.0\,°C}$ and $Q_{+3.5\,°C}$ produce significantly higher *peaks* in both calibration and validation. At a temperature change of $+5.0\,°C$, a clear decrease in model performance is again visible in the calibration, although it is still above that of the reference run in the validation. In the second event of the validation, the smaller *peak* in the descending flood wave disappears for the configurations $Q_{+3.5\,°C}$ and $Q_{+5.0\,°C}$. This phenomenon can already be observed with the application of an external distributed snow model in Section 5.2. The possible influence of snow on the flood events is already addressed there. The time of the flood events within the year also suggests that the two events during validation (May/June) are more likely to be influenced by snow than the two events during calibration (July/August). This is confirmed by looking at the input data and by the model results with increased temperature time series. Especially the two events in the validation are influenced by snow, which cannot be represented satisfactorily by the uniformly assumed input data and the lumped model structure.

The analysis of the temperature influence on the different flow conditions confirms that besides precipitation and discharge, temperature should not completely disregarded as a

possible source of uncertainty. While a large impact on the *snowmelt* periods is not apparent, the influence can be significant for individual snow-affected events. The relatively steep and short-term increase in discharge and the usually long-lasting high discharge volume during *snowmelt* periods (Compare the hydrograph at the beginning of the second quarters in Figure 3.3) is obviously also dependent on other factors, with which the model has difficulties in correctly representing.

6.1.3 Discrepancy of Input Variables

This subsection focuses on the question if discrepancies in available time series of input variables have evident drawbacks on modeling purposes. Data discrepancy can appear in different ways. Input data can have a time offset if they are not synchronized on the time scale. Such an offset can have a significant effect on the goodness of fit of the model (*McCuen et al.*, 2006). These discrepancies occur due to record errors of the measurement devices, the aggregation time interval as well as the distribution of the rain gauges inside and outside of the modeled catchment. The effect of different aggregation time intervals on the input variables precipitation and discharge is shown in the following item. The second item of this subsection deals with potential inadequacies of indirect measurements such as the potential evapotranspiration (PE). This variable is usually determined with more or less complex empirical equations. Here, discrepancies can occur between the calculated and the physically reasonable values for a certain region.

Time Offset of Input Data
Daily precipitation data provided by weather services are usually used. These data are often not measured for a "full" day which means from 12:00 midnight to 12:00 midnight. For instance, the German Weather Service (DWD) measures at 6:00 UTC and takes the last 24 hours as daily value for the previous day. Other observation data like discharge and temperature are often assumed for a "full" day. This can cause a time offset between the input variables because of different aggregation time intervals. The model representation of flood events occurring in the first half of the day is also influenced if the precipitation of the first hours is already considered in the previous day. Then, the amount of precipitation is "missing" for the actual time step and for the discharge conditions being modeled.

The mean areal precipitation depth (h_{MAP}) of configuration IV_d and IV_{agg} are plotted against each other in Figure 6.4. Configuration IV_d includes 24 daily stations with the aggregation time interval of the DWD. Configuration IV_{agg} are the aggregated hourly data to a "full" day for the corresponding 24 gauging stations. The mean annual precipitation is about $1500\,mm$ for both configurations. The red plus represents the daily h_{MAP} and the black cross the daily h_{MAP} aggregated for two days. Especially for the daily resolution, a clear variation between configuration IV_d and IV_{agg} is visible. The precipitation amounts are influenced by the different aggregation time intervals bringing a time offset. The assumption that relevant precipitation events usually do not take more than 48 hours should smooth out the effect of time offset for the data set of aggregated two-days (h_{2d}) mean areal precipitation. This fact is also seen in Figure 6.4 where the spread of the black crosses is much smaller. The calculation of the coefficient of determination (R^2) for the daily as well

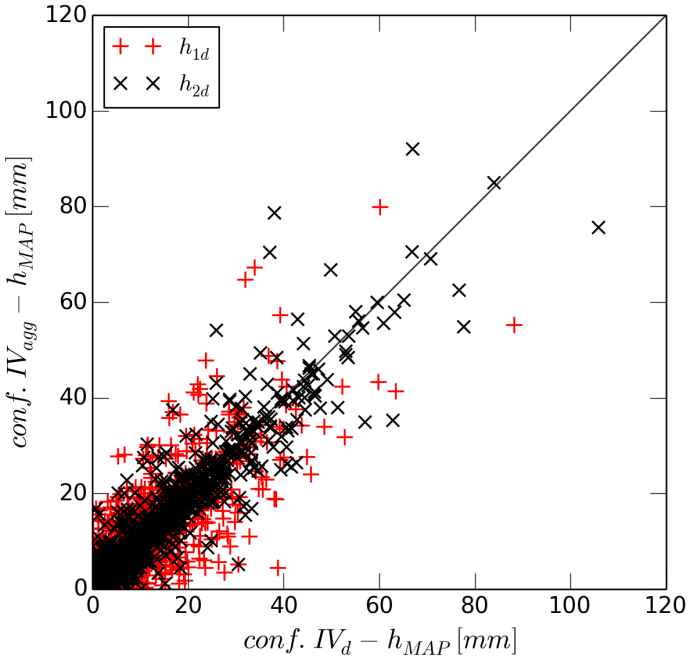


Figure 6.4: Comparison between station configuration IV_d and IV_{agg} for daily (h_{1d}) and for aggregated two-days (h_{2d}) mean areal precipitation depth (h_{MAP}).

as the aggregated two-days data set verifies this assumption. R^2 is 0.74 for the daily and 0.90 for the aggregated two-days data set. The resulting model output evaluated by the NSE of these two configurations summarized in Table 6.3 does not reproduce this effect of data discrepancy due to the time offset between discharge and precipitation data.

Table 6.3: Model performance (NSE) of daily (IV_d) and aggregated hourly data (IV_{agg}) for investigating the effect of input data time offset.

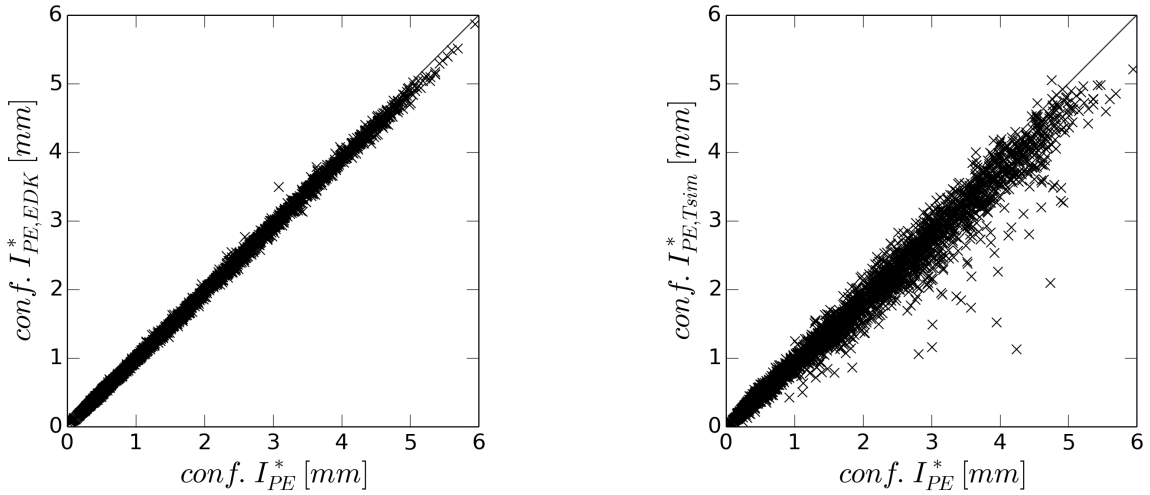
	total		snowmelt		no snow		peak	
	IV_d	IV_{agg}	IV_d	IV_{agg}	IV_d	IV_{agg}	IV_d	IV_{agg}
NSE_{cal}	0.71	0.75	0.42	0.48	0.75	0.79	0.81	0.82
NSE_{val}	0.60	0.58	0.17	0.16	0.65	0.63	0.61	0.49
NSE_{Δ}	0.11	0.17	0.25	0.32	0.10	0.16	0.20	0.33

Overall, it can be stated that the supposedly correct synchronized precipitation time series performs worse than the input times series with a time offset. Even though configuration IV_{agg} outperforms configuration IV_d for the calibration period, worse results for the validation period are obtained for configuration IV_{agg}. The largest decline of model performance for configuration IV_{agg} occurs for the *peak* flows. Due to robust model outputs, configuration IV_d shows better results for all evaluated time periods and flow conditions. In this case, a time offset of input variables does not necessarily deteriorate the model output.

Effects of Overestimated Evapotranspiration Data

Maps of annual long-term mean potential evapotranspiration (PE) suggest that the used input values are obviously overestimated for the investigated region. In Subsection 3.1.2, mean annual values of PE found in literature between $350\,mm$ and $550\,mm$ are mentioned. In connection with possible reasons for the deficient water balance, an estimated and for modeling purposes used mean annual value of about $660\,mm$ is reported in Section 4.5. To quantify the influence of such a discrepancy on the model output, different modifications of PE are applied to check if the model reacts sensitive on problematic PE values.

The first modification of the PE time series is the processing procedure of the PE values. Three different configurations are analyzed. All PE values are calculated by the approach of Hargreaves and Samani shown in Subsection 2.2.1. Except for the analysis of overestimated PE data, the basic data set (I^*_{PE}) is always used for FLEX model runs in this study. The calculation of the PE values uses mean areal temperature values which are interpolated by EDK. The configuration $I^*_{PE,Tsim}$ considers mean areal temperature values simulated by RM. For the third configuration $I^*_{PE,EDK}$, an individual PE time series is firstly generated for all available 22 temperature stations. Afterwards, a mean areal PE value is interpolated by EDK for each time step.



(a) lumped vs. distributed

(b) kriged vs. simulated

Figure 6.5: Comparison of different potential evapotranspiration (PE) time series.

The scatterplots in Figure 6.5 compare the three different PE data sets. The left plot shows the configuration I^*_{PE} and $I^*_{PE,EDK}$. The values of both configurations are close together and the differences are apparently small. Nevertheless, it is recognizable that the distributed assumption of PE values accounts for slightly lower values in the higher ranges. In contrast, between the configurations I^*_{PE} and $I^*_{PE,Tsim}$ exists a greater spread with higher values. Here, the PE time series generated by simulated temperature values tends to have lower values. This tendency is also confirmed by the mean annual potential evapotranspiration (PE_a) and the deficit of water balance in Table 6.4. Depending on the processing of the PE

Table 6.4: Some facts and the resulting FLEX performance of different modified potential evapotranspiration (PE) time series.

description	unit	configuration		
		I^*_{PE}	$I^*_{PE,Tsim}$	$I^*_{PE,EDK}$
mean annual PE_a	$[mm]$	662	599	646
deficit of water balance	[%]	19.2	16.6	18.6
NSE_{cal}	[-]	0.709	0.708	0.715
NSE_{val}	[-]	0.638	0.639	0.642

time series, the PE_a is decreased up to $60\,mm\,a^{-1}$ which can reduce the deficit of water balance by almost $3\,\%$. The model performance for the *total* time series presents almost identical results for all three configurations. In order to see any difference in model performance, the third decimal place is given for this case. Then, configuration $I^*_{PE,EDK}$ suggests the "best" results.

Table 6.5 shows the parameters which basically influence the evapotranspiration processes of the FLEX model. Apart from one parameter, all parameters are identical. Only the parameter controlling the maximum unsaturated storage capacity $S_{U,max}$ changes a bit. Obviously, the differences of the three PE time series are negligible for parameter settings or the model is able to compensate slight differences by adjusting just single parameters. A distinct overestimation of PE during the summer month can be the main reason for the deficit in water balance. The water deficiency in the system can influence the model parameterization, which especially controls the base flow. For this reason, the effect of potentially overestimated PE values is tested by the corruption of PE sequences before heavy flood events. The antecedent moisture conditions are changed by taking PE sequences of other flood events as well as very dry periods.

Table 6.5: FLEX model parameters referring to processes driven by potential evapotranspiration (PE).

model parameter	unit	configuration		
		I^*_{PE}	$I^*_{PE,Tsim}$	$I^*_{PE,EDK}$
maximum percolation capacity P_{max}	$[mm\,t^{-1}]$	4.0	4.0	4.0
transpiration threshold ET_{par}	[-]	30	30	30
maximum unsaturated storage capacity $S_{U,max}$	$[mm]$	17.4	14.7	19.8
maximum interception threshold I_{max}	$[mm]$	10	10	10

Two flood events during the year 2005 (calibration) and 2013 (validation) are taken for the analysis. The PE sequences are replaced 30 days before the flood event. The first configuration is the respective sequence of the other event. In addition, three more configurations

are used. The configurations with modified PE sequences before the flood event in 2005 are recalibrated by the FLEX model. The evaluation of the flood event in 2013 is done with constant model boundary conditions (i.a. input variables and parameter set). Solely, the model is revalidated with the modified PE sequences before the flood event.

The analysis of the resulting model output for the flood events in 2005 and 2013 are given in Table 6.6. The evaluated parameter sets of the recalibrated model which basically influence the evapotranspiration processes of the FLEX model are all identical. Thus, such a short corrupted sequence has absolutely no effect on the parameter settings. The absolute *peak* flow has higher discharges for the configurations with fewer total potential evapotranspiration (PE_{tot}) for both investigated flood events. Nevertheless, the two configurations ($I^*_{HW05,1}$ and $I^*_{HW05,3}$) present worse results looking at the model performance measures of the event in 2005. Independently of calibration or validation period, the impact of a modified PE sequence before a single flood event has almost no influence to the model performance of the *total* time series. Here, all modified configurations perform equal or slightly worse compared to the original configuration I^*. The calculation of the *peak* model performance (NSE_{peak} and $PBIAS_{peak}$) considers the corrupted 30 days before and 10 days after the flood event. A more considerable impact is visible for the model performance of both *peak* flows. Always the opposing PE condition performs distinct worse. The NSE has a decline of $3 - 5\,\%$ and the $PBIAS$ up to $7\,\%$. The discharge conditions of the flood event in 2005 indicate a model bias toward overestimation due to the $PBIAS$. The flood event of 2013 shows a model bias toward underestimation.

Table 6.6: Analysis of modified potential evapotranspiration (PE) sequences before the flood events in 2005 ($Q_{peak,obs} = 34.5\,mm$) and 2013 ($Q_{peak,obs} = 70.6\,mm$).

HW 2005	unit	configuration				
		I^*	$I^*_{HW05,1}$	$I^*_{HW05,2}$	$I^*_{HW05,3}$	$I^*_{HW05,4}$
PE_{tot}	$[mm]$	109	77	114	69	116
$Q_{peak,mod}$	$[mm]$	27.1	28.5	28.0	28.2	26.6
$NSE_{tot,cal}$	[-]	0.71	0.70	0.71	0.70	0.71
$NSE_{tot,val}$	[-]	0.64	0.64	0.64	0.64	0.64
$NSE_{peak,cal}$	[-]	0.82	0.77	0.83	0.78	0.83
$PBIAS_{peak,cal}$	[%]	-14.2	-21.1	-14.3	-21.0	-12.3
HW 2013	unit	configuration				
		I^*	$I^*_{HW13,1}$	$I^*_{HW13,2}$	$I^*_{HW13,3}$	$I^*_{HW13,4}$
PE_{tot}	$[mm]$	77	109	114	69	116
$Q_{peak,mod}$	$[mm]$	30.2	29.0	29.5	29.9	28.4
$NSE_{tot,val}$	[-]	0.64	0.63	0.63	0.64	0.62
$NSE_{peak,val}$	[-]	0.59	0.56	0.57	0.58	0.55
$PBIAS_{peak,val}$	[%]	22.7	26.8	27.3	22.3	28.2

For systematically erroneous PE values, *Oudin et al.* (2006) found a noticeable impact on the model performance, especially if the PE was overestimated. Other studies demonstrated that the model performance is almost not affected by inadequate or rather randomly corrupted PE values, or that the model parameters are able to compensate PE biases. According to PE values found in literature for this area, it is evident that the used input time series consider distinct overestimated PE values. The different processing approaches of the PE time series are able to reduce the mean annual PE_a as well as the deficiency in water balance up to 3 %. But, these facts are apparently compensated by the model and do not bring any effect on the model output. The results of modified PE sequences confirm the variation of about $1\,\%$ in model performance evaluated by the NSE for the *total* time series which is also reported by other studies in Subsection 4.1.3. The model performance of single flood events shows a more distinct influence on corrupted PE sequences, but the impact on the absolute *peak* discharge is almost negligible. The recalibrated model parameters are absolutely not affected by such short modified sequences. The correction of systematically overestimated PE time series is a difficult task and is not part of this investigation. At least, this short analysis shows that the influence on the *total* time series and the absolute *peak* discharges is negligible in the context of this study.

6.2 Effects of Data Processing

The Sections 4.3 and 4.4 introduce and discuss different interpolation (kriging) and simulation (random mixing) methods as well as their effects on processing input variables. The influence on the modeling process and output is now described within the next subsections. Initially, the three different applied kriging methods and the conditional simulation method random mixing (RM) are compared among each other. Then, the incorporation of linear inequality constraints to RM (Subsection 4.4.2) is tested for investigating measurement uncertainties of gauging stations on high elevations. In the last part of this section, the uncertainty of input data and its effects to the model output is analyzed with the use of many simulated precipitation time series as model input.

The indices of the different modifications of RM are used as follows: Only RM denotes the configuration where the averaged precipitation field of 100 different simulation runs is used as single model input. The superscript *i* refers to 100 different simulated precipitation fields and the corresponding individual model runs. The index *m* explains the mean result of 100 model runs. The index *min* and *max* similarly denotes the worst and the best model runs out of 100. The incorporation of *greater-equal* constraints is labeled with a *greater* sign. The respective number stands for the level of elevation ($e.g.,\ 1000\,m.a.s.l.$) above which *greater-equal* constraints are applied to the gauging stations.

6.2.1 Comparison of Spatial Interpolation and Simulation Methods

Table 6.7 compares the resulting FLEX model performance of the different input configurations gained by the three interpolation methods (ODK, EDK and EDK_s) and the simulation method (RM) on the *total*, *snowmelt* and *no snow* period. The performance measures

Table 6.7: Performance of the FLEX model evaluated by different measures and interpolation/simulation methods (rain gauge configuration III).

	total				snowmelt				no snow			
cal.	ODK	EDK	EDK_s	RM	ODK	EDK	EDK_s	RM	ODK	EDK	EDK_s	RM
R^2	0.78	0.78	0.78	0.78	0.60	0.58	0.58	0.59	0.85	0.85	0.85	0.85
$RMSE$	1.5	1.4	1.4	1.4	2.0	2.0	2.0	2.0	1.3	1.1	1.1	1.2
$PBIAS$	-3.6	-1.6	-1.6	-3.1	11.9	10.8	10.8	9.4	-12.5	-8.7	-8.6	-10.3
NSE	**0.75**	**0.77**	**0.77**	**0.76**	**0.46**	**0.44**	**0.45**	**0.44**	**0.80**	**0.84**	**0.84**	**0.83**
KGE_m	0.98	0.98	0.98	0.98	0.88	0.88	0.88	0.90	0.92	0.96	0.96	0.94
	total				snowmelt				no snow			
val.	ODK	EDK	EDK_s	RM	ODK	EDK	EDK_s	RM	ODK	EDK	EDK_s	RM
R^2	0.69	0.67	0.67	0.70	0.67	0.65	0.65	0.67	0.71	0.69	0.69	0.73
$RMSE$	1.8	1.8	1.8	1.7	1.6	1.6	1.6	1.7	1.8	1.8	1.8	1.7
$PBIAS$	-3.2	3.0	3.1	-2.2	13.7	15.2	15.3	13.0	-11.1	-2.7	-2.6	-9.3
NSE	**0.67**	**0.67**	**0.67**	**0.68**	**0.44**	**0.46**	**0.46**	**0.39**	**0.69**	**0.68**	**0.68**	**0.72**
KGE_m	0.97	0.96	0.96	0.97	0.87	0.85	0.84	0.88	0.91	0.97	0.97	0.94

show very similar values for the *total* calibration period. Taking the validation period into account, the differences between the interpolation and simulation methods as well as the various performance measures are more apparent. Due to absolute performance level and robustness, the results of RM slightly outperform the input configurations generated by kriging methods for all measures. The analysis of the *snowmelt* period indicates a preference for the kriging methods due to NSE. The KGE_m which accounts more for a balanced volume tends to the set-up generated by RM. The *no snow* period suggests a preference to the both interpolation methods using an external drift if the absolute performance level is taken, especially for the calibration period. Configuration RM and even ODK show better results due to model robustness. The $PBIAS$ for the *total* period shows values close to zero during the calibration and validation for all configurations. The configurations ODK and RM slightly overestimate the volume for calibration as well as validation. The configurations EDK and EDK_s show an opposing trend with a marginally overestimated volume during calibration and a volume underestimation during validation. The $PBIAS$ of the *snowmelt* period clearly underestimates the volume during calibration and validation for all configurations. The *no snow* period indicates a volume overestimation during calibration and validation for all configurations. The discrepancy of underestimated volume during the *snowmelt* and overestimated volume during the *no snow* period is obvious. It seems to be enough total precipitation as model input due to the almost balanced volume for the *total* period. But, the available precipitation amount is wrongly distributed between the *snowmelt* and *no snow* period. Apparently, the snow reservoir of the model does not store or release enough precipitation volume during the *snowmelt* period.

Table 6.8: Different measures evaluating the peak flows for interpolation/simulation methods (rain gauge configuration III).

	ODK		EDK		EDK_s		RM	
	cal.	val.	cal.	val.	cal.	val.	cal.	val.
EV_{peak}	-0.26	0.06	-0.11	0.19	-0.11	0.19	-0.26	0.04
EP	0.90	0.57	0.73	0.44	0.72	0.44	0.94	0.57
NSE_{peak}	0.79	0.69	0.84	0.55	0.84	0.55	0.81	0.70

A more detailed analysis of the *peak* flows is given in Table 6.8. Looking at the representation of the *peak* flows, all performance measures show a great gap between the results of calibration and validation. Only the NSE_{peak} of configuration ODK and RM is in an acceptable range like the other analyzed time periods. The EV_{peak} again shows a clear trend to overestimated volumes during calibration and underestimated volumes during validation. Both configurations of ODK and RM outperform the configurations of EDK. The presented model results in Table 6.7 and 6.8 have the identical shortcomings due to performance level and model robustness as already shown on the tested approaches in Chapter 5. The performance differences between the various interpolation and simulation methods are generally not very significant. The conclusion of a marginal influence due to different processed precipitation input data sets is already done for different kriging methods in Subsection 5.1.1. Even though the now investigated rain gauge configuration III contains more and spatially well-distributed stations, the shortcomings of the model output remain. At this point, the denser and spatially well-distributed station arrangement might bring no benefit in terms of model robustness. Focusing on improved model robustness, the applied simulation method (RM) reveals the most promising results.

6.2.2 Data Modification with Random Mixing

In Chapter 4, the data quality is controlled using simple plausibility checks such as physical relationships and water balance. One result is the almost always negative water balance of the system. In the previous subsection, the discrepancy of input variables is investigated on overestimated potential evapotranspiration (PE) data. The result of this analysis is that the used input time series considers distinct overestimated PE values according to PE values found in literature for this area. This finding is certainly the main part of the deficient water balance. But, inadequate rain gauge data also contribute to a lack of water in the balance. In the first part of Chapter 4, losses through wind are named as the main reason for systematic errors measuring precipitation. Consequently, larger wind-induced measurement errors occur in high mountain regions due to higher wind speeds and proportion of snow. The QC of the rain gauge data in Section 4.2 shows no clear trend of increasing mean annual precipitation sums for higher station altitudes. The relatively low mean annual precipitation sums for the higher located stations suggest underestimated rain gauge data due to systematic measurement errors. Another indicator for erroneous precipitation data from higher located rain gauge stations is the calculation of the correlation coefficient (R) in Table 4.1.

If all stations are considered almost no relationship between station altitude and precipitation sum is visible. Taking just the hydrological winter half-years into account, the results even show a light anticorrelation. This fact may also indicate data quality problems due to snowfall. Disregarding rain gauge stations above certain altitudes shows a clear increase of R. A neglect of all stations above $1000\,m.a.s.l.$ means an increase of R from 0.06 to 0.54 depending on the station configuration.

For a better investigation of the above-mentioned problems due to potentially erroneous precipitation data from higher located rain gauge stations, the conditional simulation method random mixing (RM), introduced in Section 4.4 is used. Different precipitation time series are simulated incorporating *greater-equal* constraints to rain gauge stations above altitude levels of $1000\,m.a.s.l.$ and $700\,m.a.s.l.$ In Section 4.5, the water balance by different generated precipitation time series shows a reduction of the deficit from about $22\,\%$ to $4\,\%$ applying *greater-equal* constraints to stations above $1000\,m.a.s.l.$ This subsection shows the influence of simulated precipitation time series using *greater-equal* constraints on the model output. The set-up of rain gauge configuration *II* is the base of the following model results. The total amount of stations is 34 whereby four stations are located above $1000\,m.a.s.l.$ and the total number of stations above $700\,m.a.s.l.$ is 13. The denotation of the indices is explained in the introduction of Section 6.2.

Table 6.9: Performance of the FLEX model evaluated by NSE and EV for interpolated and simulated input data as well as incorporated *greater-equal* constraints (rain gauge configuration II).

	EDK_s		RM		$RM_{>700}$		$RM_{>1000}$	
	cal.	val.	cal.	val.	cal.	val.	cal.	val.
NSE_{tot}	0.73	0.64	0.76	0.65	0.67	0.67	0.68	0.67
NSE_{snow}	0.43	0.28	0.42	0.35	0.31	0.38	0.34	0.41
$NSE_{no\ snow}$	0.79	0.68	0.83	0.68	0.73	0.70	0.73	0.70
EV_{tot}	-0.04	0.07	-0.03	0.08	-0.05	0.08	-0.05	0.08
EV_{snow}	0.11	0.24	0.07	0.20	0.06	0.20	0.06	0.20
$EV_{no\ snow}$	-0.12	-0.01	-0.09	0.03	-0.12	0.03	-0.11	0.03

The results in Table 6.9 show that the model results obtained by RM slightly outperform EDK_s according to the NSE. Due to the measure EV, both input configurations are quite similar except for the *snowmelt* period where RM again suggests better results. The performance gap between calibration and validation period still remains for both configurations. The model results which are calculated with simulated input data and incorporated *greater-equal* constraints have a slight lower performance level in the calibration period according to the NSE. But, the validation period performs better and more robust model outputs are generally obtained. Considering the EV for different investigated periods, all input configurations have the same weakness. The calibration as well as the validation underestimates the volume of *snowmelt* period. Here, the use of simulated input data can obviously reduce this deficit. Independently of the input configuration, the other evaluated flow peri-

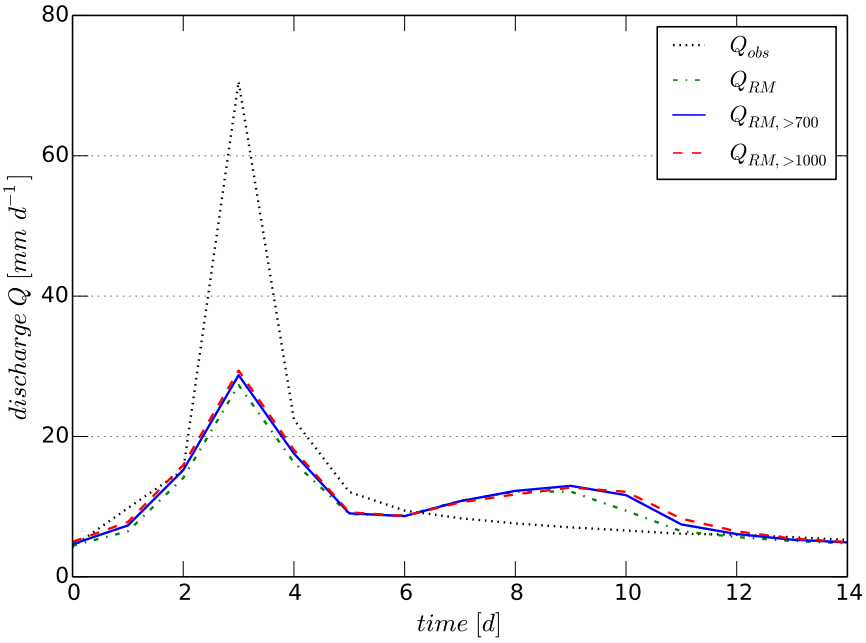


Figure 6.6: Hydrograph of flood event 2013 modeled by the FLEX model with different simulated precipitation time series (rain gauge configuration II).

ods show almost similar tendencies of overestimated flow volumes for the calibration and an underestimation for the validation.

Figure 6.6 and Table 6.10 exemplify the results of the heaviest observed flood event within the investigation period. The different corresponding simulated and accumulated precipitation patterns are already shown in Figure 4.15. All modeled hydrographs generally have the same trend where the *peak* is considerably underestimated and a second clear smaller but smoother *peak* can be noticed. The second *peak* probably results from an inaccurate representation of rain and snow in the internal lumped snow model. Two to three days before the flood event, precipitation amounts of about 40 to $50\,mm$ are recorded, which are temporarily fallen as snow down to $1000\,m.a.s.l.$ Snow is usually very wet and build just a temporary snow pack which melts fast for this season. The constant snowmelt rate of the model is probably too small for representing such fast occurring as well as disappearing snow events during the summer period. The daily resolution of the modeling processes can also have its influence. Such a delayed response of discharge is picked up in a subsequent section. The *peak* hydrographs of the three configurations only present marginal differences in Figure 6.6. But, the configurations using simulated precipitation time series with incorporated *greater-equal* constraints performs better. This tendency is also validated by the performance measures in Table 6.10. The configuration with applied *greater-equal* constraints to all gauging stations above $1000\,m.a.s.l.$ performs best for this event. Configuration $RM_{>1000}$ assumes the most precipitation as well as discharge volume. All three configurations have a clear positive balance of available water. Nevertheless, the model performance of all configurations can be ranked on a moderate to satisfactory level. The EV indicates a model bias toward underestimated discharge volumes which is already seen in Figure 6.6. Neverthe-

Table 6.10: Performance analysis of a single flood event in 2013 generated by different simulated precipitation time series ($Q_{tot,obs} = 191.2\,mm$).

	unit	configuration		
		RM	$RM_{>700}$	$RM_{>1000}$
$h_{MAP,5d}$	$[mm]$	205.4	221.7	224.4
$Q_{tot,mod}$	$[mm]$	148.1	157.7	161.5
NSE	[-]	0.47	0.50	0.52
KGE_m	[-]	0.74	0.84	0.87
EV	[-]	0.23	0.18	0.16
EP	[-]	0.39	0.41	0.42

less, between configuration RM and $RM_{>1000}$ an performance improvement of $7\,\%$ for the EV and $5\,\%$ for the NSE is achievable.

Using such simulated time series, which consider measurement uncertainties due to the location of gauging stations, for instance with incorporating *greater-equal* constraints, is an approach to address and verify these problems. Here, it can be seen that the use of *greater-equal* constraints gives more robust model results. Nevertheless, volume problems still remain which can originate from other error sources such as overestimated input time series of PE, investigated in Subsection 6.1.3. This fact can also result in modeling problems such as determining good parameter sets.

6.2.3 Analysis of Input Data Uncertainty

An explicit reduction of the uncertainty of input data can support a better estimation of other uncertainty sources. Uncertainty of interpolated precipitation estimations contribute a noteworthy portion to the gaps between observed and modeled discharge. Thus, more reliable and improved model predictions are hardly achievable without a clear decrease of precipitation estimation errors (*Moulin et al.*, 2009). An improved quality as well as a more appropriate temporal and spatial resolution of the input data can emphasize inadequate model structures (*Neumann and Gujer*, 2008). Erroneous precipitation data can have a significant declining influence on model performance (*Moulin et al.*, 2009). Depending on the model, *Oudin et al.* (2006) specified a mean decrease of the NSE from $35\,\%$ up to $50\,\%$ if randomly corrupted precipitation values were used. They still estimated a decrease in model performance of $15\,\%$ to $25\,\%$ for a systematic underestimation of precipitation. *Bárdossy and Singh* (2008) investigated the effect of observation errors on the identification of robust parameter sets by corrupting discharge and temperature data. A further point is the time dependence of interpolation errors (*Moulin et al.*, 2009). This can happen if single stations have data gaps or provide inaccurate values due to measurement errors for single time steps or periods. Depending on the distribution of rain gauge stations, spatially limited precipitation patterns can be captured insufficiently and result in under- or overestimation.

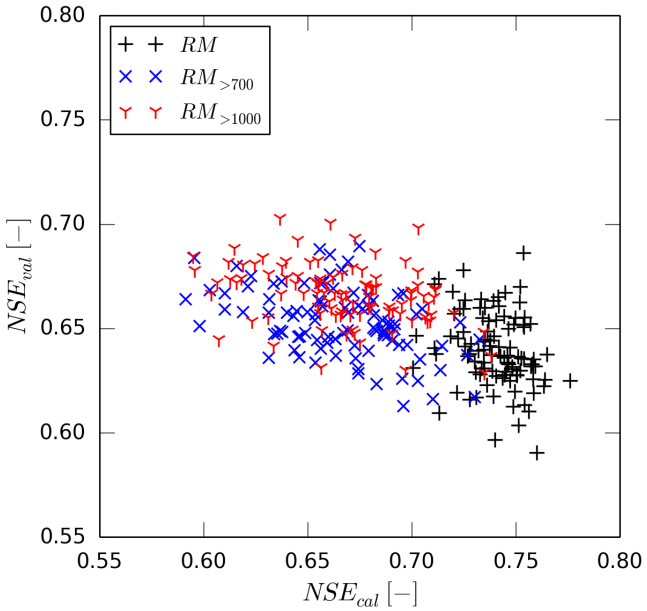


Figure 6.7: FLEX model performance (NSE) for 100 simulated precipitation time series generated by RM and incorporated *greater-equal* constraints.

For this reason, the following subsection discusses how uncertainty of input data can be taken into account for the calibration of hydrological models and the representation of the model results. RM is used to simulate many precipitation fields which are able to describe the variability of real precipitation fields. Each realization of RM maintains the observed values at the conditioning points as well as the spatial variability of the observations. Here, the model is calibrated by 100 different precipitation fields generated by RM as well as incorporated *greater-equal* constraints. The model is validated by the received 100 individual parameter sets and respective precipitation fields. The model results of these input configurations are shown in Figure 6.7 and Table 6.11. The first entry of Table 6.11 is the reference run which is the averaged input precipitation field of 100 simulation runs. The denotation of the indices is explained in the introduction of Section 6.2.

The scatterplot in Figure 6.7 suggests that the configuration of RM has the best results for the calibration runs. Both configurations with incorporated *greater-equal* constraints generally show lower values as well as a greater spread of the NSE. The validation runs reflect other results. Here, the configuration of $RM_{>1000}$ outperforms the others. Both configurations of $RM_{>700}$ and $RM_{>1000}$ show more robust model results on a slight lower performance level than the configuration RM (Table 6.11). Due to the standard deviation (S) of all 100 NSE, the scattering of the model performance increases from RM to $RM_{>1000}$ for calibration. The performance results for the validation have a reverse relationship. The model results generated by 100 different input precipitation fields show that on average the model performance is almost equal to the single model run. But, the range of model performance due to the variability of the precipitation fields varies between $8\,\%$ to $10\,\%$ for RM. The configurations with modified input data sets $RM_{>700}$ and $RM_{>1000}$ have ranges from $7\,\%$ to $14\,\%$. This evaluation suggests an impact of input data uncertainty between $7\,\%$ to $14\,\%$

Table 6.11: Performance analysis for different simulated input data as well as incorporated *greater-equal* constraints (rain gauge configuration II).

	RM		$RM_{>700}$		$RM_{>1000}$	
	cal.	val.	cal.	val.	cal.	val.
NSE	0.76	0.65	0.67	0.67	0.68	0.67
$NSE_m^{i=100}$	0.74	0.64	0.67	0.65	0.67	0.67
$NSE_{min}^{i=100}$	0.70	0.59	0.59	0.61	0.60	0.63
$NSE_{max}^{i=100}$	0.78	0.69	0.73	0.69	0.74	0.70
$S_{NSE}^{i=100}$	0.015	0.018	0.030	0.016	0.033	0.014

to the model performance due to the spatial variability of precipitation fields and potential measurement drawbacks of topographically influenced gauging stations.

Figure 6.8 exemplarily shows the range of possible *peak* flow hydrographs due to 100 different precipitation time series simulated by RM without incorporated constraints. Some parts of the hydrograph are not represented very well by the mean of all simulations. More sections of the observed discharge hydrograph are captured if the range of model output which originate by the uncertainty of the precipitation input is considered. Some clear discrepancies between observed and modeled hydrographs remain, independently of the taken simulation run. Here, other reasons are responsible for the modeling problems. For instance, an insufficient representation of snow accumulation or melt can cause a delayed discharge response or a volume shift.

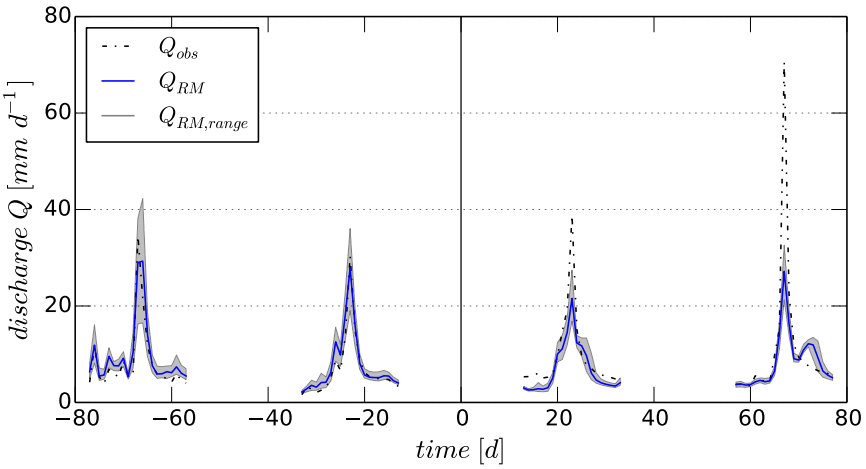


Figure 6.8: Peak flow hydrographs for 100 simulated precipitation time series (RM).

Another option to investigate the uncertainty of input data is the consideration of multipliers for precipitation and evapotranspiration (*Sikorska et al.*, 2012). The SUPERFLEX framework enables such an analysis. Nevertheless, the assumption of multiplicative precipitation errors has its limitations, especially if the errors originate from time offsets between the precipitation and discharge time series (*Kavetski et al.*, 2006b,c). Another restriction of such mul-

tipliers is that just systematic variations with a constant value are applicable for the whole time series. This approach needs the assumption of systematic uniform input data errors. This study focuses more on the quantity and spatial distribution of rain gauge stations and its influence on the spatial variability of precipitation fields. For this reason, the approach with multipliers on lumped input time series is not further investigated.

6.3 Model Performance Analysis

One important issue of the analysis of model performance and finally the reliability of predictions is the detection of reasons for errors and uncertainty. These perceptions support further proceeding in improving model predictions. They indicate if the predictions benefit from more adequate model structures or accurate input information (*Del Giudice et al.*, 2015). Input and output data, model structures and their parameterization are sources of uncertainty (*Wagener*, 2003; *Kavetski et al.*, 2006b). The main part of prediction uncertainty arise from inaccurate precipitation information and structural model deficiencies (*Sikorska et al.*, 2012). In a comprehensive model evaluation, the reliability and uncertainty of the obtained output has to be assessed. Nevertheless, a general assumption of statistical uncertainty estimation is that the applied model is true and the errors follow a stochastic process. This can be problematic as models due to their conceptual process representation cannot be considered as true and measurement devices often produce (systematic) errors (*Neumann and Gujer*, 2008). Good or rather reliable model results can be assumed if the model performance ranges within the obtained uncertainty bands (*Refsgaard and Henriksen*, 2004). The magnitude of uncertainty bands are also useful measures for the reliability of the model results (*Willmott et al.*, 1985). Other studies indicate model predictions as reliable if at least 95 % of the validation data is covered by the 95 % uncertainty band (*Del Giudice et al.*, 2015).

The separation of error components and thus the estimation of the different uncertainty sources is a difficult task and can impede a profound analysis of model accuracy (*Kavetski et al.*, 2006b). One of the challenges is the variety of potential error sources on the one hand, and the availability of just one data pair (observed and modeled) for measuring the deviations or residuals at every time step on the other hand (*Beven*, 2006). Especially, the uncertainty originating from structural model deficiencies is particularly vague and difficult to separate (*Wagener*, 2003). Structural inadequacy can originate from missing state variables and processes. Another source due to numerical errors might be avoided by implementing robust numerical methods (*Gupta et al.*, 2012). Sources of model uncertainty can also interact with each other. An improved model structure often involves supplementary process descriptions and accordingly an increase of complexity due to more parameters. A reduction of model structural uncertainty may be expected on the one hand, but the parameter uncertainty can increase on the other hand (*Silberstein*, 2006). Biased and deceptive estimated parameter sets can be caused by input errors which again can differ from catchment to catchment due to different data availability and quality as well as topographic diversity. This can lead to difficulties especially if the parameters are intended for predictions in ungauged basins (*Kavetski et al.*, 2006b). Another limitation for predictive reliability are biased parameter sets arising from input uncertainty structures which are different in calibration

from those in validation (*Kavetski et al.*, 2006b). *Doherty and Welter* (2010) argue that model calibration often fails in simulating all characteristics of a discharge time series equally well. Depending on the calibration process, it may occur that, e.g., low flows are better estimated than high flows or the other way round. The study of *Gan et al.* (1997) ascertained that wet years outperform dry years as calibration data. Obviously, wet years contain more sufficient information of low and high flow conditions to activate all model parameters.

This section of the study provides a model performance analysis due to time-dependent and event-based criteria. The time-dependent analysis includes different lengths of time series, hydrological seasons and various modeling intervals. Unusual events are used for the event-based performance analysis. Subsequently, the model parameterization is considered with different model configurations.

6.3.1 Time-dependent Performance Analysis

In *Brath et al.* (2004) a period not less than three months is suggested for model calibration with hourly time step. A few other studies are mentioned which assume that such a calibration length is sufficient with hourly resolution. In Bayesian analysis an increasing length of the calibration time period can maintain a reduction of the estimated model parameter uncertainty (*Doherty and Welter*, 2010). *Bergström* (1991) is of the mind that there is no general rule for the required length of a calibration period as long as an adequate range of significant events is included in the data. *Gan et al.* (1997) argues similar that the data should include a sufficient amount of average, dry and wet years. All model parameters characterizing different flow conditions are thus activated during calibration. He proposes a time length of three to five years. If a bad model fit is expected, the time interval should be selected as short as possible for the most accurate performance index. Else, an insensitivity of the index can occur to the critical time interval (*McCuen et al.*, 2006).

This study uses a time period of ten years (2004 - 2013). As mentioned in Section 3.2, the time period is chosen because of the required flexible data set-up for higher temporal resolutions and the sparse availability of hourly data before 2004. Due to the general conditions of the research project, the investigated time period is adapted and extended during this study. For the first part of the investigations, just a time series from 2004 until 2009 is available. In a second step, the time series is extended until 2013. The only drawback of the second data set-up is the missing data for the additional rain gauge stations of configuration III for the year 2013. In the meantime, precipitation data are available for the 27 additional stations of configuration III. Where the required preprocessing and subsequent model runs are possible with justifiable efforts, the new data set is considered for the model evaluation in Section 6.3. The flow duration curve in Figure 3.6 shows the non-exceedance probability of the selected (2004 - 2013) and the long-term (1901 - 2006) time period. Both curves match fairly well. Thus, the selected time series of ten years is representative for the long-term discharge characteristics of this river. The selected time period even provides more relevant and higher flood events. That should be beneficial for investigations of high discharge conditions. Flood events which exceed the long-term mean of the $99.7\,\%$ duration of non-exceedance occurred in 2005, 2010 and 2013.

Length of Time Series

This paragraph presents the results of different long time series. The first modeling approaches just consider a data set from 2004 until 2009. Consequently, two of the highest observed flood events are not included within the short time series until 2009. The calibration (validation) length differs from 3 (2) to 5 (4) years to the regular investigation time period from 2004 until 2013. The first year is always used as warm-up period and is disregarded for the performance evaluation. Non of the four highest *peak* flows occur during the validation period of the short time series. For this reason, the analysis of the *peak* flow performance is skipped. Both time series are based on rain gauge configuration III.

Table 6.12: Observed and modeled discharge statistics for different long time series (1: 2004 - 2009 and 2: 2004 - 2013).

	total [mm]				snowmelt [mm]				no snow [mm]		
cal.	Q_m	Q_{max}	Q_{sum}	S	Q_m	Q_{max}	Q_{sum}	S	Q_m	Q_{sum}	S
obs_1	3.6	34.5	3942	3.2	5.1	17.6	1405	2.9	3.1	2537	3.1
III_1	3.6	29.0	3957	3.0	4.2	15.0	1172	2.8	3.4	2785	3.0
obs_2	3.7	34.5	6667	2.9	5.3	17.6	2425	2.7	3.1	4242	2.8
III_2	3.7	24.7	6772	2.7	4.7	11.9	2163	2.8	3.4	4609	2.6
	total [mm]				snowmelt [mm]				no snow [mm]		
val.	Q_m	Q_{max}	Q_{sum}	S	Q_m	Q_{max}	Q_{sum}	S	Q_m	Q_{sum}	S
obs_1	3.7	20.7	2726	2.6	5.5	11.8	1020	2.4	3.1	1706	2.3
III_1	3.4	20.2	2514	2.3	4.2	11.4	769	2.3	3.2	1744	2.2
obs_2	3.6	70.6	5187	3.1	4.5	11.8	1646	2.2	3.2	3541	3.3
III_2	3.5	29.4	5076	2.5	3.8	13.0	1407	2.4	3.4	3670	2.5

The statistics in Table 6.12 provide an overview of the observed and modeled discharge conditions for each case. Q_{max} is always similar for the *total* and the *no snow* time period. The comparison of the observed discharge data shows that the mean discharge Q_m matches quite well between the two time series as well as between calibration and validation period for the *total* and the *no snow* period. Only the *snowmelt* period differs between the previously mentioned cases. Here, the short time series (obs_1) has a higher Q_m during validation whereas the long time series (obs_2) shows the opposite. The short discharge time series has only very low maximum discharges Q_{max} for validation. In terms of the standard deviation S, the observed discharge of the short time series shows slightly larger differences between calibration and validation. The comparison between the observed and modeled statistics indicates that the model is able to reproduce the basic statistics Q_m, Q_{sum} and S of the *total* period for both time series as well as for the calibration and validation. Only the reproduction of Q_{max} has the known deficiencies. This is particularly noticeable for the long time series. The observed Q_{max} is significantly higher during validation. Thus, the calibrated

Table 6.13: Model performance analysis of different long time series.

	2004 - 2009						2004 - 2013					
	total		snowmelt		no snow		total		snowmelt		no snow	
	cal.	val.	cal.	val.	cal.	val.	cal.	val.	cal.	val.	cal.	val.
R^2	0.83	0.70	0.73	0.51	0.89	0.81	0.78	0.69	0.58	0.66	0.85	0.72
$RMSE$	1.3	1.4	1.8	2.2	1.1	1.0	1.4	1.7	2.0	1.6	1.1	1.7
$PBIAS$	-0.4	7.8	16.6	24.6	-9.8	-2.3	-1.6	2.1	10.8	14.5	-8.6	-3.6
NSE	**0.83**	**0.69**	**0.62**	**0.13**	**0.88**	**0.80**	**0.78**	**0.69**	**0.44**	**0.47**	**0.84**	**0.71**
KGE_m	0.99	0.94	0.84	0.62	0.95	0.99	0.98	0.97	0.88	0.86	0.96	0.97

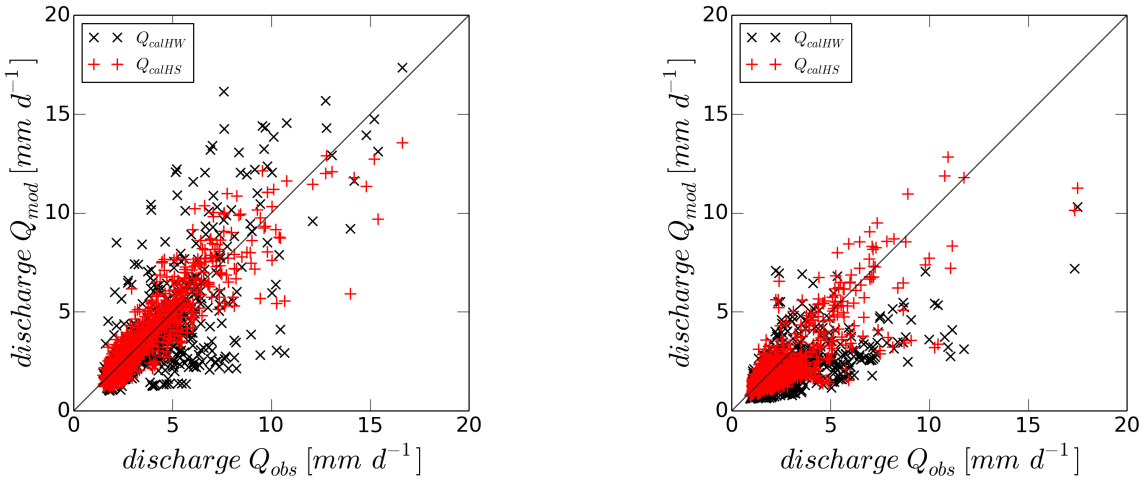
model extrapolates its information of the hydrological system into the unknown. Looking at Q_{sum} of the splitted times series, it can be seen that the missing discharge volume of the *snowmelt* period is contained in the *no snow* period.

The findings of analyzing the descriptive statistics are also represented in the model performance (Table 6.13). The overestimated discharge volume during the *no snow* period is reflected by the negative values of $PBIAS$. The *snowmelt* period shows more robust results for the long time series. Regarding the *total* and *no snow* periods, the short time series suggests a better model performance according to R^2 and NSE. $PBIAS$ and KGE_m tend more to the long time series. This investigation indicates that depending on the considered performance measure a decision to both time series is possible. In summary, the short time series seems to be the better approach. But, it should be noted that the short time series has no significant *peak* flows in the validation period. In contrast, the two *peak* flows in the validation of the long time series are significantly higher than those in the calibration. They are probably the decisive reason for the slightly worse model performance in the *total* and the *no snow* period.

Hydrological Seasons

This subsection examines the model behavior for different configurations regarding the hydrological year (HY) as well as the hydrological summer and winter half-years. One reason for using hydrological years as input time series is that the snow reservoir of a catchment system is assumed to be empty at the beginning of November. This should avoid a deficit of precipitation volume if the snow reservoir is initially empty starting at the beginning of a calendar year. These circumstances are relevant for mountainous catchments. In order to compensate for these effects, the first year is always used as warm-up period in this study, regardless of the model set-up. The considered time series for this analysis ranges from November 2004 to October 2013.

Figure 6.9 shows the validation results separated by the hydrological summer and winter half-years. The model validation is performed once with the optimized parameter set from the hydrological winter and summer half-years. The objective of this illustration is to check whether the model is capable to predict the dissimilar or the rather unknown hydrological condition. During the validation phase, only two events occur where the observed discharge



(a) Hydrological summer (HS)

(b) Hydrological winter (HW)

Figure 6.9: Validation of the hydrological seasons using the optimized parameter sets of the summer and winter half-years.

Q_{obs} is greater than $20\,mm$. Both events are measured during the hydrological summer half-year. Both parameter sets underestimate the *peak* flow almost identically. In the hydrological winter half-years, no events with an observed discharge Q_{obs} of more than $20\,mm$ occur. For this reason, the two discharge events are not shown in Figure 6.9 in order to be able to compare the results for the two hydrological half-years more clearly.

The scatterplot for the hydrological winter represents that the higher discharges ($> 5\,mm$) are completely underestimated by the winter parameter set. The modeled discharge by the summer parameters shows a behavior towards under- and overestimation for these discharge conditions. Looking at the left figure of the hydrological summer, the scattering of the modeled discharge by the winter parameters is much larger than the summer set-up. Especially for higher discharges, the modeled discharge by the winter parameter set tends to a large spread towards under- and overestimation. Only on the basis of the two scatterplots, it appears that the original parameter set produces the better model results for the hydrological summer half-years. For the hydrological winter half-years, only the scatterplot cannot be used to make such a clear statement. It is also difficult to say which parameter set represents the dissimilar hydrological condition better. For this reason, the corresponding model performance is considered in the next paragraph.

The model represents the mean discharge and the standard deviation very well in calibration (Table 6.14). The validation results show a quite good representation of the observed mean discharge according to the related parameter set. The standard deviation of the modeled discharge during the validation period is smaller for all set-ups. This fact is obviously due to events with a discharge of more than $20\,mm$, which are clearly underestimated by the model. If these events are disregarded or if the set-up of winter half-years is considered, the standard deviation of the modeled data also agrees well with the observed ones. The overall view of the performance measures shows that the hydrological summer half-years

Table 6.14: Model performance analysis of different calibration and validation set-ups regarding hydrological seasons.

		$Q_{obs,m}$	$Q_{mod,m}$	S_{obs}	S_{mod}	$RMSE$	$PBIAS$	NSE
HY	$cal.$	3.6	3.7	2.9	2.9	1.5	-3.6	0.73
	$val.$	3.6	3.5	3.0	2.6	1.8	1.8	0.65
HS	$cal.$	4.5	4.5	3.1	2.9	1.2	-0.5	0.85
	$val._{HS}$	4.3	4.3	3.7	2.8	1.8	-0.1	0.77
	$val._{HW}$	4.3	4.0	3.7	3.3	2.4	6.2	0.57
$HS_{<20mm}$	$cal.$	4.3	4.4	2.6	2.6	1.1	-1.5	0.83
	$val._{HS}$	4.1	4.2	2.4	2.3	1.0	-1.8	0.80
	$val._{HW}$	4.1	3.9	2.4	3.0	1.9	4.6	0.33
HW	$cal.$	2.7	2.8	2.3	2.3	0.9	-0.8	0.83
	$val._{HW}$	2.9	2.5	2.0	1.8	1.2	13.8	0.63
	$val._{HS}$	2.9	2.0	2.0	1.2	1.8	30.6	0.21

are best represented by all different model set-ups. If discharges $> 20\,mm$ are not taken into account for the performance calculations, even robust model results are obtained for the summer parameter set. The model bias ($PBIAS$) indicates a light overestimation of the model results for all calibration set-ups and the validation periods of the summer half-years with the corresponding parameter set. All set-ups including winter months or the winter parameter set show a trend to slightly overestimated model results for the calibration and always a tendency for underestimation during the validation phase. Both configurations of the hydrological summer half-years also show better model results when the NSE is evaluated. The use of the hydrological time series does not show an improvement to the model performance and robustness compared to time series regarding a calendar year (see also Table 6.7). The seasonal model evaluation shows that the hydrological summer half-years perform better than the *total* time period and the winter half-years. These worse results for the winter times correspond to the findings already gained during the evaluation of the *snowmelt* periods. It shows again, that the process representation of the very complex snow accumulation and melt seems to be a weakness of the lumped model structure. However, the winter parameter set better reproduces the dissimilar hydrological conditions in the summer half-years than the other way round. The more complex calibration of the winter period apparently better activates the various hydrological processes represented by the model structure and associated parameters.

Different Modeling Intervals

The last part of the time-dependent performance analysis deals with different modeling intervals (MI). The simple split-sampling method differs with regard to the respective length of the time period and swapped time intervals for calibration and validation (Table 6.15).

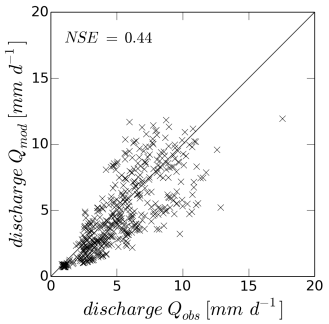
This approach investigates if there are clearly different conditions between calibration and validation period and to what extent these have an influence on model performance. The specific representation of the *peak* flows in calibration or validation receives special attention. The two considerably higher flood events are so far only taken into account in the validation. For configurations with swapped time series, these events are part of the calibration. This is used to investigate whether it has an influence on the previously significantly underestimated representation of the *peak* flows. Another point is the influence of a different division of the time intervals for calibration and validation. Can a longer calibration period provide additional input information that has an influence on model robustness and performance? The data for the period 2004 to 2013 are used for all configurations. The first year before the calibration interval is always a warm-up period and is not taken into account for the calculation of performance measures. The evaluation of *peak* flows includes events from the years 2005, 2006, 2010 and 2013 (Figure 6.12).

Table 6.15: Configuration overview and description for the different modeling intervals (rain gauge configuration III).

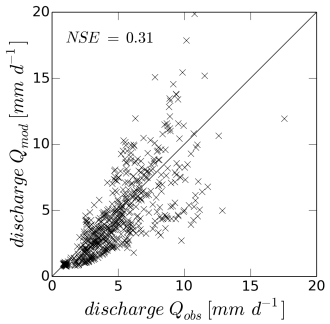
ID	description	modeling interval		no. of years	
		cal.	val.	cal.	val.
MI_1	basic configuration, precipitation data for 2013 included	'05 - '09	'10 - '13	5	4
MI_2	swapped calibration time period	'09 - '13	'04 - '07	5	4
MI_3	short calibration time period	'05 - '08	'09 - '13	4	5
MI_4	long calibration time period	'05 - '10	'11 - '13	6	3
MI_5	swapped long calibration time period	'08 - '13	'04 - '06	6	3

The following discussion of the descriptive statistics is solely qualitative for the different modeling intervals. The univariate measures mean discharge ($Q_{obs,m}$) and standard deviation (S) of all observed data sets show almost equal discharge conditions for the *total* and the *no snow* period of calibration and validation. The modeled mean matches fairly well to the observed mean for calibration as well as validation for all model set-ups. The modeled standard deviation tends to be slightly underestimated in both calibration and validation. The *snowmelt* periods show clearer differences in the univariate measures for both calibration and validation as well as the different configurations. $Q_{obs,m}$ as well as S of configuration MI_1 and MI_4 during calibration are much higher than the validation period. Configuration MI_5 shows the opposite case. Only configuration MI_2 and MI_3 have almost similar observed conditions for calibration and validation according to both univariate measures.

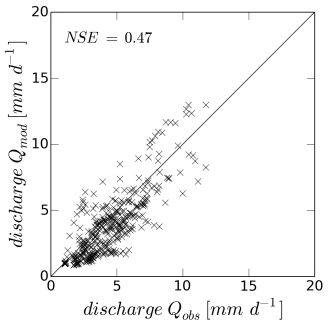
Figure 6.10 presents the calibration and validation results of configuration MI_1 and MI_4 for the *snowmelt* period. The scatterplot of MI_4 shows especially for higher discharges a wider spread and a trend to overestimation during calibration. The two scatterplots for the validation do not differ so much anymore. The parameterization of the model seems to be able to compensate this in the validation of MI_4. Regarding the swapped time series of



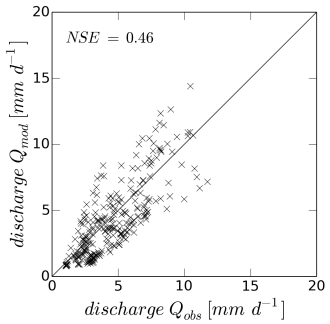
(a) MI_1 (calibration)



(b) MI_4 (calibration)

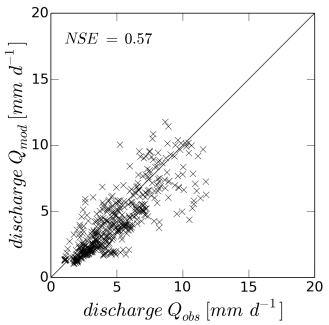


(c) MI_1 (validation)

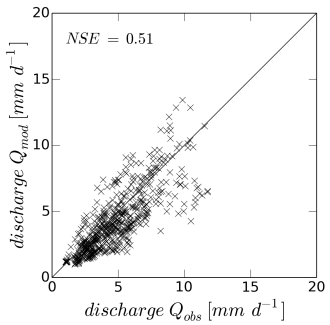


(d) MI_4 (validation)

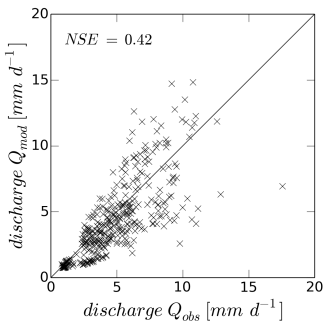
Figure 6.10: Scatterplots of different chronological modeling intervals (snowmelt period).



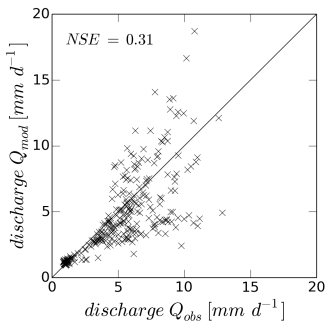
(a) MI_2 (calibration)



(b) MI_5 (calibration)



(c) MI_2 (validation)



(d) MI_5 (validation)

Figure 6.11: Scatterplots of different swapped modeling intervals (snowmelt period).

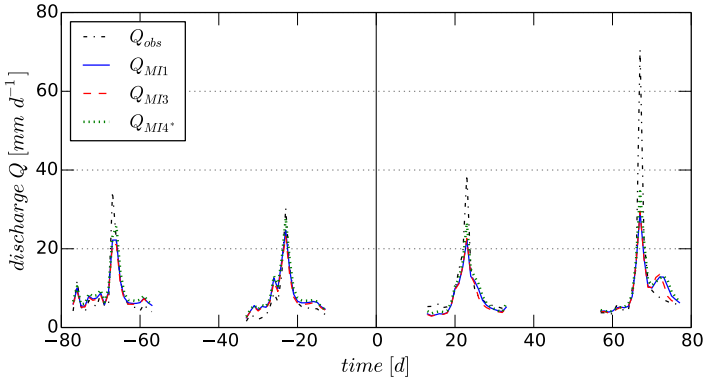
configuration MI_2 and MI_5, the differences between the scatterplots of the calibration as well as the validation are not so clear (Figure 6.11). Only some overestimated data points are recognizable for the higher discharges of MI_5. The scatterplots of the *snowmelt* period show that a longer calibration period does not necessarily get better performance results, especially with respect to the robustness of the model results. This behavior can be noted for both chronological and swapped time series.

Table 6.16: Model performance analysis of different calibration and validation modeling intervals. (*Only the event of 2013 is taken into account for validation.)

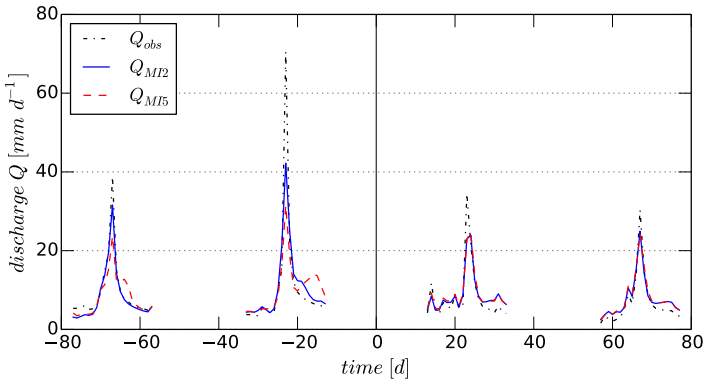
		MI_1		MI_2		MI_3		MI_4		MI_5	
		cal.	val.	cal.	val.	cal.	val.	cal.	val.	cal.	val.
total	$RMSE$	1.4	1.7	1.4	1.4	1.4	1.7	1.5	1.8	1.5	1.6
	$PBIAS$	-1.6	2.1	-0.7	-9.1	-1.1	7.1	-3.5	-2.0	-1.2	-10.2
	NSE	0.78	0.69	0.80	0.76	0.76	0.71	0.76	0.67	0.73	0.72
snow	$RMSE$	2.0	1.6	1.6	2.0	2.0	1.9	2.1	1.7	1.6	2.3
	$PBIAS$	10.8	14.5	14.3	9.9	10.2	17.4	8.7	7.0	12.2	11.0
	NSE	0.44	0.47	0.57	0.42	0.38	0.39	0.31	0.46	0.51	0.31
no snow	$RMSE$	1.1	1.7	1.3	1.2	1.1	1.6	1.2	1.8	1.5	1.2
	$PBIAS$	-8.6	-3.6	-8.0	-19.1	-7.5	2.0	-10.1	-6.4	-8.0	-23.2
	NSE	0.84	0.71	0.84	0.84	0.83	0.75	0.84	0.70	0.75	0.81
peak	$RMSE$	2.7	7.2	4.8	2.6	2.8	7.1	3.0	8.2*	7.0	2.7
	$PBIAS$	-11.2	14.5	4.8	-10.4	-5.3	17.7	-9.5	2.1*	5.5	-13.9
	NSE	0.84	0.60	0.82	0.85	0.83	0.61	0.83	0.66*	0.62	0.84

In a further step, the various configurations are evaluated on the basis of the model performance. Looking at the NSE, the configuration MI_1 seems to provide better and more robust model results for the *snowmelt* period than configuration MI_4 (Figure 6.10). The $RMSE$ shows an almost identical performance. The $PBIAS$ even suggests a better model performance of MI_4 because of a lower model bias toward underestimation for calibration as well as validation (Table 6.16). Configuration MI_2 performs better for the swapped time intervals of the *snowmelt* period. Analyzing the model performance of the five configurations for the *total* time period, configuration MI_2 performs best for $RMSE$ and NSE. Due to the $PBIAS$, configuration MI_4 is the best, followed by the basic configuration MI_1. Configuration MI_5 has the worst performance results for the *total* time period. Configuration MI_2 also performs best when evaluating the four *peak* flow events. A comparison of the two configurations MI_2 and MI_5 with swapped time period shows that the higher *peak* flows in calibration are still clearly underestimated (Figure 6.12). However, configuration MI_2 reproduces the two highest *peak* flows better than any other configuration in calibration or validation. The calibrated *peak* flows of configuration MI_5 are at the same low level as those configurations where these events are part of the validation. Despite the clear differences

in calibration of MI_2 and MI_5, the two validated *peaks* have almost identical hydrographs. The chronological configurations of MI_1 and MI_3 show hardly any differences in calibration and validation of the *peaks*. Only the configuration MI_4 generates higher *peaks*, where the first three events are part of the calibration.



(a) Chronological time period



(b) Swapped time period

Figure 6.12: Peak flow hydrographs of different modeling intervals. (*The first three peaks are part of the calibration.)

The evaluation of the different modeling intervals indicates that there is no configuration that consistently reproduces the best result for all discharge conditions and performance measures. Configuration MI_3 with a shortened calibration interval generates comparable results as the basic configuration MI_1. Only the $PBIAS$ has a slightly higher mean tendency towards an underestimated modeled discharge for the validation. In contrast, a longer calibration interval (MI_4) only has a positive influence on the validation of *peak* flows. For the remaining discharge conditions, the model performance is even worse. No improvement of the model results is apparent with a longer calibration interval (MI_5) for the swapped time period. Based on the performance measures $RMSE$ and NSE, as well as the model robustness, the swapped time period of configuration MI_2 tends to be the best modeling approach. However, the performance analysis also indicates that the $PBIAS$ has a tendency for overestimating all discharge conditions apart from the *snowmelt* period during valida-

tion. This may be due to the fact that the model weights the responsible parameters of the fast reacting processes more strongly due to the two higher flood events in calibration and thus also smaller *peaks* are rather overestimated. The investigations on the basis of different modeling intervals show that the length of calibration has no clear influence on the model performance for the selected time series. The configuration MI_4 with the longest calibration time period tends to have even worse model results. Regarding model robustness but also quality, the swapped time series generate the better results. Configuration MI_2 provides the best results for calibration and validation, especially for *peak* flows. For the chronological modeling approaches, the two highest *peak* flows are part of the validation period. This is the case of extrapolation, since the model must reproduce unknown discharge conditions in validation. The two highest *peaks* are part of the calibration for the swapped time series. Thus, the *peak* flows of validation are the case of interpolation. The model already knows these discharge conditions from calibration. In this case, all model parameters that describe the different discharge conditions are better activated during calibration. The fact of inter- and extrapolation cases must be taken into account when Chapter 7 attempts to clarify the question of applicability for the prediction of flood events.

6.3.2 Event-based Performance Analysis on Unusual Events

The higher the temporal resolution of model calculations, the higher are usually the computational times. Finding an appropriate compromise between temporal resolution and sufficient information output by the model is a challenging task. Thus, in many studies the model is just calibrated on single flood events or a series of events (*Brath et al.*, 2004). An alternative is the calibration on unusual events (UE) which can reduce the data demand for model calibration significantly. Unusual events are periods of a data time series (e.g., discharge or temperature) which contain more information than the rest of the time series. In the following, it is assessed if model calibration on unusual events is sufficient for estimating good and robust parameter sets. The theory of determining unusual events with the depth function is given in Subsection 2.3.3. Unusual events are selected from the total discharge time series (Q_{tot}), the discharge time series of the hydrological winter half-years (Q_w), the antecedent temperature index (ATI) and the daily mean temperature (T_m). In total, four different combinations of unusual event time series are considered as model input:

- UE_1: Q_{tot}
- UE_2: Q_{tot} & Q_w
- UE_3: Q_{tot} & ATI
- UE_4: Q_{tot} & T_m

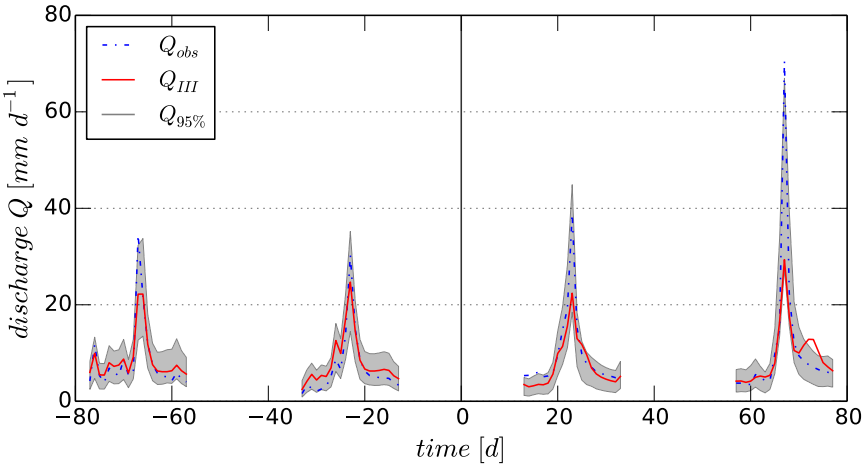
The response of a catchment to an unusual event usually takes more time than just the current and the subsequent day at which such conditions are identified. The antecedent hydrological situation also influences the flow behavior. Every five previous and subsequent days of an unusual event are considered for model calibration and evaluation. The used calibration data set varies between $9.2\,\%$ up to $19.4\,\%$ due to the total amount of data. The model validation always considers the complete time series.

Table 6.17: Model performance resulting from calibration on unusual event times series.

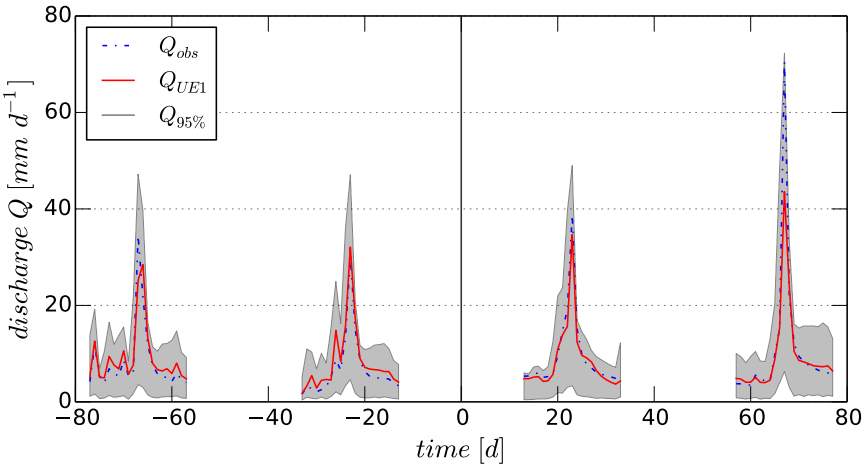
		III		UE_1		UE_2		UE_3		UE_4	
		cal.	val.	cal.	val.	cal.	val.	cal.	val.	cal.	val.
total	$RMSE$	1.4	1.7	1.8	1.8	1.3	1.8	1.5	1.7	1.4	1.7
	$PBIAS$	-1.6	2.1	-2.3	14.6	-0.2	11.7	0.6	11.9	-1.1	14.4
	NSE	0.78	0.69	0.92	0.64	0.93	0.66	0.91	0.69	0.91	0.70
snow	$RMSE$	2.0	1.6	1.3	2.7	1.1	2.7	1.5	2.4	1.4	2.7
	$PBIAS$	10.8	14.5	18.2	37.3	5.3	35.2	11.1	34.7	5.4	41.0
	NSE	0.44	0.47	0.89	-0.54	0.91	-0.55	0.78	-0.28	0.78	-0.57
no snow	$RMSE$	1.1	1.7	1.8	1.5	1.4	1.4	1.5	1.4	1.4	1.1
	$PBIAS$	-8.6	-3.6	-4.0	4.0	-4.0	0.8	-0.7	1.3	-2.6	2.0
	NSE	0.84	0.71	0.92	0.80	0.94	0.83	0.92	0.83	0.92	0.88
peak	$RMSE$	2.7	7.2	2.4	4.3	2.4	4.5	2.6	4.6	2.4	3.6
	$PBIAS$	-11.2	14.5	-13.3	9.1	-4.3	19.8	-1.1	14.1	-6.3	9.6
	NSE	0.84	0.60	0.87	0.85	0.88	0.84	0.85	0.84	0.87	0.90
data ratio [%]		100	100	9.2	100	16.9	100	16.1	100	19.4	100

The model performance in Table 6.17 indicates very good calibration results regarding the NSE for the *total* time series. However, the validation performance is similar or even slightly worse than the reference model set-up. The $PBIAS$ shows that nearly all configurations tend to minimal overestimation of the discharge during the calibration. The validation shows the opposing effect where all configurations of UE have a clear model bias toward underestimation. This fact can be explained if the model performance of the *snowmelt* periods is considered. Here, all measures indicate very poor model results for the validation period. Regarding the NSE, even the mean observed discharge is a better predictor than the model. A look on the other two evaluated flow conditions of *no snow* and *peak flow*, presents a different picture. There, very good and quite robust model results are obtained according to the performance measures for the calibration as well as the validation period. The best results are achieved if the unusual events are determined by two selection criteria. Here, the configuration UE_4 (Q_{tot} & T_m) outperforms the remaining configurations.

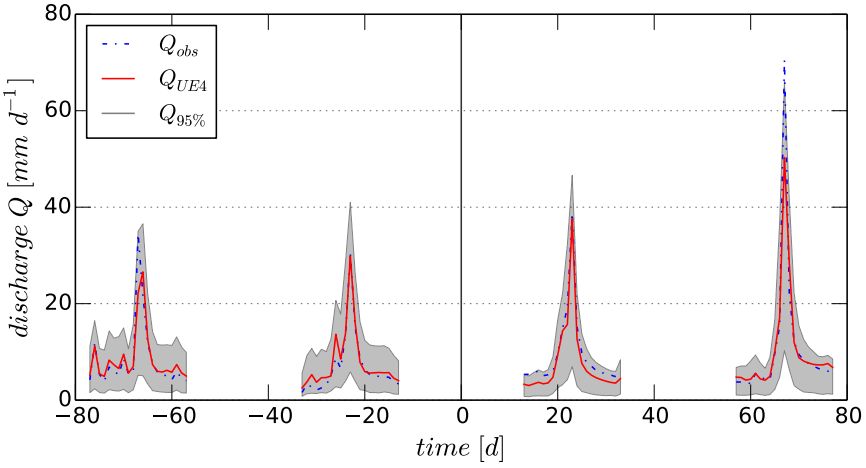
The hydrographs of the *peak* flows also show the improvements between the configurations of III (complete time series), UE_1 (only Q) and UE_4 (Q_{tot} & T_m) in Figure 6.13. The delayed second *peak* of the last event disappears for both UE time series. The match of observed and modeled *peak* flows is very good for the last three *peaks* of configuration UE_4. Only the first event causes difficulties in its reproduction for all set-ups. Configuration UE_1 has even for the lower discharges a relatively large range for the uncertainty bands. Here, configuration UE_4 also shows improved uncertainty bands, which suggests more robust model results.



(a) III



(b) $UE_1\ (Q)$



(c) $UE_4\ (Q\ \&\ T)$

Figure 6.13: Peak flow hydrographs of different unusual event configurations.

In summary, this investigation shows that the approach of model calibration on unusual events is obviously not applicable for *snowmelt* periods. Here, the used time series does not contain sufficient information and the calibrated model parameters are not transferable to the validation. A second temperature driven criterion for determining the unusual events is necessary. So, enough hydrological information is extracted from the *total* time series to activate the model functionalities and parameters describing the other flow conditions during the validation period. Except for the *snowmelt* periods, this approach is useful to calibrate a hydrological model because of its reduced computing time for estimating robust parameter sets.

6.3.3 Analysis of Model Parameterization

The robustness and the resulting reliability of a model arise from its capability to represent all occurring hydrological conditions as good as possible (*Perrin et al.*, 2001). Using the model performance of the calibration time period as only benchmark for the predictive capability of a model is a fundamental error (*Refsgaard and Henriksen*, 2004). It can often be seen that model results show a significant drop of model performance for the validation period. One reason can be the overparameterization of the model by too many degrees of freedom in contrast to the information contained in the observations (*Bergström*, 1991; *Seibert*, 1997). Poor input information due to data quality and data processing can also cause such a gap between the performance during calibration and validation. Hence, these reasons are an explicit indicator for the presence of either erroneous data or inadequate model structure (*Fenicia et al.*, 2008). The higher the difference in performance measures between the calibration and the validation period, the less transferable are the estimated model parameters. In contrast, higher performance measures during validation can be assumed to be purely random (*Bárdossy and Das*, 2008).

It can be seen that many factors influence the process of determining a robust model parameterization. Therefore, the aim of model calibration should not only focus on finding the best performing parameter set. Instead, different objectives should be achieved with the best possible set of parameters. This includes a good model performance for the investigation time period and a meaningful accounting for the hydrological processes. The obtained model results should be quite insensitive to small changes of the parameter values. Finally, the use of a transferred parameter set to other time periods does not decline the model performance (*Bárdossy and Singh*, 2008). Problems of under- or overestimated discharge conditions do not only arise from erroneous input data (mainly precipitation). They can also result from compensation effects in the model structure. A frequent underestimation of unusual *peak* flows could lead the model to increase the precipitation or speed up the outflow of the storages controlling the fast flow to reduce the model error. As a consequence of such compensation effects, the model can tend to overestimate generally smaller *peak* flows or wrongly adjust other storage parameters. Parameter uncertainty may be decreased by splitted calibration procedures for different seasonal or flow conditions (*Seibert*, 1997). Parameters with no relevant influence on the results increase the degree of freedom in a model. This can be a source of uncertainty. A sensitivity analysis supports finding such parameters and helps to identify a robust parameterization (*Bergström*, 1991). Another

Table 6.18: Model parameterization and performance of different FLEX reservoir compositions (validation: 2010 - 2013).

model parameter				model configuration			
symbol	unit	limit					
		lower	upper	$C1_{WLS}$	$C2_{WLS}$	$C3_{WLS}$	$C4_{WLS}$
I_{max}	$[mm]$	0.1	10	10	10	10	10
T_0	$[°C]$	-3	4	1.5	2.3	3.3	4.0
T_m	$[°C]$	-3	1.5	0.6	0.3	0.0	-0.4
k_W	$[mm\,t^{-1}\,°C^{-1}]$	0.01	150	1.3	1.3	1.3	1.2
k_F	$[t^{-1}]$	$5 \cdot 10^{-4}$	1	$5.0 \cdot 10^{-4}$	$5.0 \cdot 10^{-4}$	0.34	0.51
k_S	$[t^{-1}]$	$1 \cdot 10^{-5}$	0.5	$7.8 \cdot 10^{-3}$	$0.5 \cdot 10^{-5}$	$2.8 \cdot 10^{-3}$	$1.0 \cdot 10^{-5}$
NSE_{cal}	$[-]$	$-\infty$	1	0.76	0.76	0.75	0.73
NSE_{val}	$[-]$	$-\infty$	1	0.62	0.62	0.65	0.64
NSE_{Δ}	$[-]$	-	-	0.14	0.15	0.10	0.09
$NSE_{snow,cal}$	$[-]$	$-\infty$	1	0.47	0.45	0.44	0.42
$NSE_{snow,val}$	$[-]$	$-\infty$	1	0.31	0.32	0.27	0.28
$NSE_{snow,\Delta}$	$[-]$	-	-	0.16	0.13	0.17	0.14

source of uncertainty is the fact that the calibration on different time sequences can result in unequal "optimal" parameter sets (*Gan and Biftu*, 1996). In this context, *Beven* (1993) proposed the concept of equifinality where acceptable model predictions can be obtained by different model structures or parameter sets. Achieving good model simulations with varying values of one parameter can even occur if the simulations are sensitive to changes of this parameter. This can be an indication for compensating this parameter sensitivity by other parameters (*Seibert*, 1997). The range of obtained parameter values can be used as an indicator for parameter uncertainty. The parameter value may not spread over a wide range for similar good model simulations. Else, it is a hint for a poorly defined parameter (*Seibert*, 1997). The parameter set obtained by the calibration might be unique and should be independent of the calibration time sequences (*Gan et al.*, 1997; *Gupta et al.*, 1999). However, completely different meteorological conditions can result in poor discharge simulations for validation (*Seibert*, 1997). Another drawback of many actual catchment models is the incapability of reproducing low and high flow conditions with a single parameter set (*Wagener*, 2003). *Schoups et al.* (2008) discovered better model fits if temporally variable parameter sets are used. The parameters can be obtained by different seasons or by certain events.

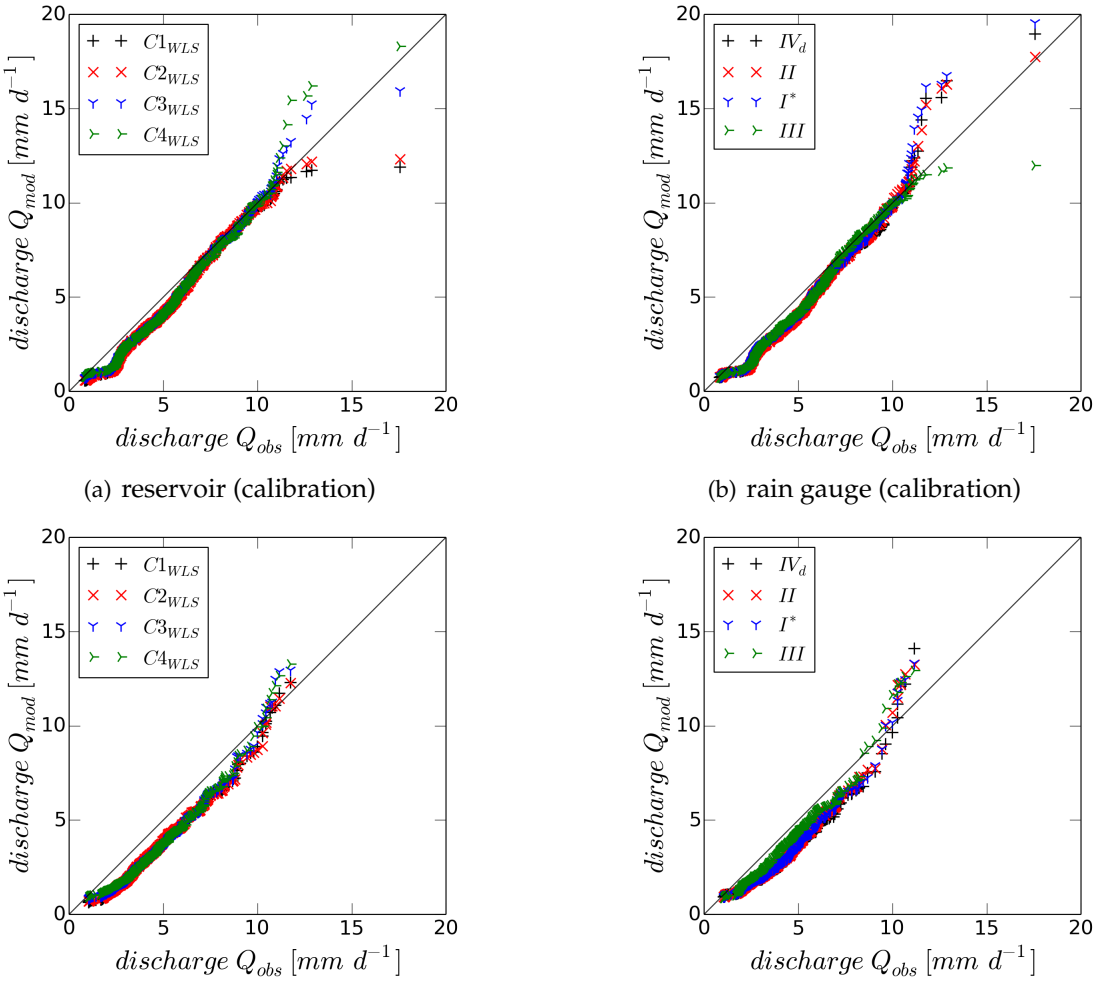
Here, the analysis of the model parameterization does not include all parameters. The selection of parameters refers to those which describe important hydrological processes and occur in all chosen model configurations. *Snowmelt* periods are identified as particular weak points of the previous model representation. For this reason, special attention is paid to the parameters that describe the snow processes. Table 6.18 shows the parameterization and performance (NSE) of different FLEX reservoir compositions which are presented in detail

in Subsection 5.1.2. The model parameterization of the four different rain gauge configurations is compared in Table 6.19. The model evaluation of this input data analysis is discussed in Section 6.1. The predictive reliability of discharge for the *snowmelt* period is shown with quantile-quantile ($Q-Q$) plots for both model set-ups in Figure 6.14.

Table 6.19: Model parameterization and performance of different rain gauge configurations (validation: 2010 - 2012).

model parameter				model configuration			
symbol	unit	limit lower	limit upper	IV_d	II	I^*	III
I_{max}	$[mm]$	0.1	10	10	10	10	10
T_0	$[°C]$	-3	4	4.0	4.0	4.0	2.2
T_m	$[°C]$	-3	1.5	-0.5	-0.4	-0.5	0.6
k_W	$[mm\,t^{-1}\,°C^{-1}]$	0.01	150	1.2	1.2	1.2	1.6
k_F	$[t^{-1}]$	$5 \cdot 10^{-4}$	1	0.46	0.49	0.83	$5.0 \cdot 10^{-4}$
k_S	$[t^{-1}]$	$1 \cdot 10^{-5}$	0.5	$3.0 \cdot 10^{-5}$	$3.0 \cdot 10^{-5}$	$2.6 \cdot 10^{-3}$	$4.0 \cdot 10^{-5}$
k_R	$[t^{-1}]$	0.05	6	0.06	0.06	0.06	0.05
$NSE_{tot,cal}$	$[-]$	$-\infty$	1	0.71	0.73	0.71	0.77
$NSE_{tot,val}$	$[-]$	$-\infty$	1	0.58	0.64	0.63	0.73
$NSE_{tot,\Delta}$	$[-]$	-	-	0.13	0.09	0.08	0.05
$NSE_{snow,cal}$	$[-]$	$-\infty$	1	0.42	0.43	0.40	0.45
$NSE_{snow,val}$	$[-]$	$-\infty$	1	0.06	0.19	0.16	0.41
$NSE_{snow,\Delta}$	$[-]$	-	-	0.36	0.24	0.25	0.04

Both model approaches have in common that the results have a significant drop in performance between calibration and validation for almost all model runs. This performance degradation can be observed in all reservoir compositions, regardless of their complexity and consequently the number of parameters (degrees of freedom). In relation to the model performance of the *total* time series, the more complex model structures even show slightly more robust model results (Table 6.18). Looking at the model performance for configurations with different numbers of rain gauge stations as input data, it can be seen that the more stations are available, the more robust model results are obtained (Table 6.19). In fact, during the *snowmelt* period, configuration III is the only model run that provides acceptable and robust predictions. These results indicate that rather a low quality of the input data due to, e.g., an insufficient spatial distribution of rain gauge stations than a model overparameterization or inadequate model structure is responsible for the drop in performance. A closer look at the parameter values shows that the model calibrates some parameters towards the value limits. For example, the parameter I_{max} always assumes the upper limit for all analyzed model runs. I_{max} is the maximum interception threshold of the overfall reservoir S_I. This reservoir controls which portion of precipitation evaporates and/or is available for the subsequent runoff processes after exceeding the threshold of I_{max}. The increase of the upper



(a) reservoir (calibration)

(b) rain gauge (calibration)

(c) reservoir (validation)

(d) rain gauge (validation)

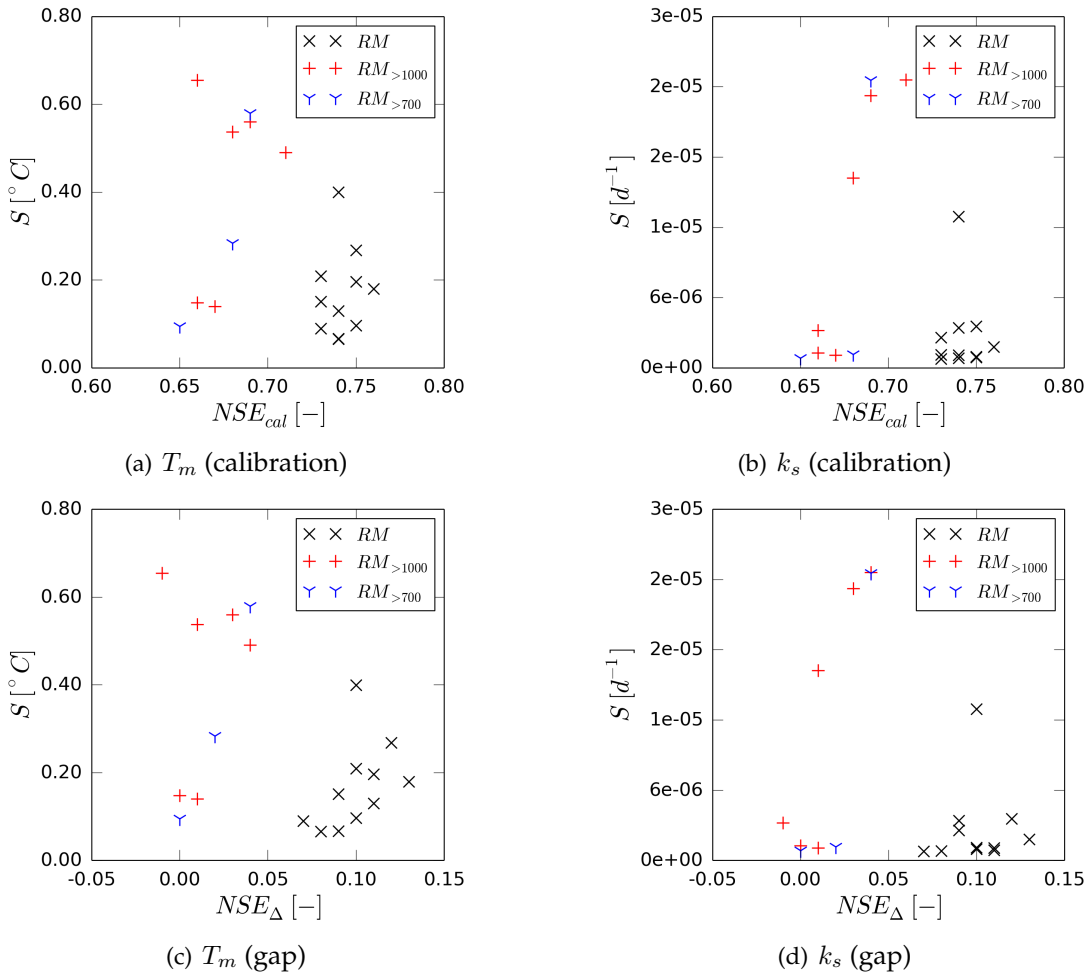
Figure 6.14: Q-Q plots for assessing the predictive reliability of discharge for the snowmelt period of different reservoir and rain gauge configurations.

parameter limit for I_{max} is not meaningful, since physically unrealistically high interception values are then accepted before runoff formation occurs. A sensitivity analysis with reduced or increased upper limits for I_{max} consequently leads to implausible reservoir characteristics. Other parameters, such as T_m, T_0 and k_F have partly clear varying values between the different model runs. The two reservoir compositions $C1_{WLS}$ and $C2_{WLS}$ achieve almost the same model performance despite considerably different values for the parameters T_m, T_0 and k_S (Table 6.18).

The Q-Q plots (Figure 6.14) show that the model calibrates the low to middle discharges ($< 10\,mm\,d^{-1}$) very similarly regardless of the configuration. However, it can be observed for the middle to higher discharges during the *snowmelt* period ($> 10\,mm\,d^{-1}$) that in some configurations the discharge is continuously under- or overestimated for calibration. The highest discharges are better represented for the model configurations with overestimated

discharges in the middle range. This can be seen even more clearly for the *total* time series, which is not presented graphically here. The Q-Q plots of the validation show a quite similar behavior for all model configurations. The discharges are slightly underestimated up to middle conditions, before a slight overestimation occurs. Here, contrary to the calibration results, all configurations behave identically. The storage coefficient k_F, which describes the fast outflow processes, is calibrated towards the lower limit in all three model runs $C1_{WLS}$, $C2_{WLS}$ and III, in which the middle to high discharges are continuously underestimated. For all other model runs, the calibration value of k_F is in a same reasonable range. However, it cannot be entirely excluded that the overestimation of middle discharges is caused by compensation effects of the fast outflow components in order to match the highest *peak* flows better. Rain gauge configuration III clearly outperforms all other listed model runs regarding the model performance and robustness. Configuration III provides the most plausible values for the snow parameters. This model set-up is the only one that achieves satisfactory performance results for the *snowmelt* period. For the remaining rain gauge configurations, the values of T_0 correspond to the upper limit and T_m always has negative values (Table 6.19). The approach with different reservoir configurations does not show significant differences in predictive capability. This is confirmed by the relatively equal and narrow value range of the NSE. On the other hand, the approach with different rain gauge stations clearly emphasizes the configuration III, which is also reflected by the values of NSE for validation.

In the first part of this analysis, individual runs of different model set-ups and resulting single parameter sets are compared. In a second step, the model results and the parameterization from simulated precipitation time series using RM as well as incorporated *greater-equal* constraints are analyzed. The evaluation is shown in Figure 6.15 using the two parameters snowmelt temperature T_m and the storage coefficient k_s. For this purpose, all 100 model runs of the three RM input configurations are sorted according to the calibration model performance (NSE_{cal}). Then, the groups of identical NSE_{cal} are sorted using the NSE for validation. NSE_Δ describes the gap between the model performance of calibration and validation. For the calculation of the standard deviation S, only combinations are taken into account that have at least four equal model runs based on the sorting of NSE. For instance, this means for configuration $RM_{>700}$ that there are only three combinations where at least four model runs achieve an identical model performance. The number of equal model runs within the combinations varies between four and nine. Figure 6.15 shows that configuration RM has the most combinations of the same model performance (11). The more rain gauge stations are simulated with *greater-equal* constraints, the less equal combinations are identifiable. The 11 combinations of RM include 59 of 100 model runs. In comparison, $RM_{>1000}$ contains 30 and $RM_{>700}$ only 17 model runs. For the configurations with *greater-equal* constraints, this is especially due to the clearly higher variation of the model performance (NSE) during calibration (see also Table 6.11). As already discussed in previous evaluations, the configuration RM generates the highest absolute values of NSE during calibration, but also the greatest gaps of NSE between calibration and validation. Despite the more robust model results based on NSE_Δ when using *greater-equal* constraints, a larger deviation of the parameters T_m and k_s can be seen for equal and similar model performances of the configurations $RM_{>1000}$ and $RM_{>700}$. The parameter uncertainty for identical model



(a) T_m (calibration)

(b) k_s (calibration)

(c) T_m (gap)

(d) k_s (gap)

Figure 6.15: Analysis of model parameter uncertainty using S and NSE based on model results generated by RM simulations.

performances of RM is greater for T_m than for k_s. The parameter uncertainty of T_m increases with increasing values of NSE_{Δ}. The variation of k_s is much smaller except for one outlier. Compared to the defined parameter limits (Table 6.19), the obtained parameter values, especially for k_s, are within an acceptable range for all three configurations of RM. The analysis of the model parameterization shows that some parameters always tend to their value limits. Other parameters indicate this behavior depending on the model set-up, which is an indication that model structure or complexity as well as the input configuration have a considerable influence on a sound parameterization. Nevertheless, test model runs for finding reasonable parameter ranges clearly show that adjusted or wider parameter limits do not support the calibration of optimal parameter sets. Continuously filled reservoirs are generated over the total modeling period. Consequently, the adjustment of the parameter limits can only be done to a very limited extent in order to obtain physically plausible reservoir characteristics. As mentioned in the discussion at the beginning of this chapter, a problem of model parameterization can be that a single parameter set is not able to represent all runoff

conditions sufficiently. In addition, in this case a lumped model approach must describe the hydrological processes of a highly variable topography. This fact is repeatedly underlined by the fact that the configuration with the most comprehensive horizontal and vertical distribution of rain gauge stations provides the best model results. The quality of the input data seems to have a stronger influence on a sound parametrization than the complexity of model structure. The evaluation of the model runs with simulated precipitation time series shows that identical model performances with unequal optimum parameter sets are possible. This already mentioned concept of equifinality has to be considered if a parameter set should be transferred to other time periods or hydrological conditions.

6.4 Summary

In this chapter, the issue of data and model adequacy is addressed from different perspectives. A comprehensive analysis of the input data and their different effects on the model behavior is the beginning of a successful modeling approach. In a second step, the effects of data processing are investigated and new simulation techniques are applied to better quantify the uncertainties of input data. Finally, different time-dependent and event-based modeling approaches are used to evaluate their impact on model performance and robustness. Model parameterization is also part of this assessment.

One of the main driving forces in hydrological modeling is precipitation as data input. The areal precipitation is often derived from point measurements using interpolation methods. Catchments with a high variability of precipitation due to a very heterogeneous topography have large uncertainties with respect to the estimation of areal precipitation. Sparse or unbalanced rain gauge networks therefore affect the model performance and robustness. The model shows that its parameterization is able to compensate for varying input information from very different station configurations. Almost identical and good model performances are achieved during the calibration, which can not be confirmed at all during the validation. In this context, good and reliable model results can only be obtained with configuration III, which has the most and spatially best-distributed rain gauge stations. The investigations with modified temperature time series show that a positive effect on the *snowmelt* periods, which have been insufficiently represented by the model so far, is not clearly apparent. However, the influence of increased temperature values during individual snow-affected events can be significant, as shown by the evaluation of the *peak* flows. This can also have an improvement in terms of robustness of the model performance for the *total* time period. Nevertheless, the results also support previous findings that substantial deficiencies in the representation of certain flow conditions can occur due to the uniformly assumed input data and the lumped model structure. Evaluations of possible discrepancies in the input data, such as a time offset or overestimated values for PE, indicate only negligible effects on model performance for this study.

The effect of different applied spatial interpolation and simulation methods is relatively small. Nevertheless, it can be emphasized that the results obtained with RM simulations perform slightly better in terms of absolute performance level and robustness. The consideration of an external drift, such as a smoothed DEM is also useful when using kriging

methods. The shortcomings of unreliable model predictions and the significant underestimation of *peak* flows during validation generally remain for all approaches. RM simulations with incorporated *greater-equal* constraints are an appropriate tool to account for measurement uncertainties at higher located rain gauge stations. The variability of real precipitation fields is described by many realizations using RM as well as incorporated *greater-equal* constraints. By evaluating model runs with many different precipitation fields, an order of magnitude can be given for the resulting uncertainties of this input variable.

The various approaches to analyze time-dependent model performance focus on time series length, hydrological seasons, and different modeling intervals. Here, the use of time series of different lengths shows that it is not so much the length, but rather the occurrence of unusual hydrological events (e.g., flood events) within the modeling period that affects performance. This suggests that the selection of the study period must include sufficiently varied discharge conditions to activate all model parameters. Modeling using the hydrological year does not yield benefits. The summer half-years outperform the winter half-years in seasonal modeling. Nevertheless, the parameter set of the winter half-year obviously provides more information for validation of the hydrological counterpart. The values of the NSE imply that the respective seasonal modeling can provide slightly better model performance compared to the modeling of the *total* period. This may also indicate that a single parameter set is not able to represent all (seasonal) discharge conditions equally well. The final time-dependent performance analysis also shows that the use of modeling intervals of different lengths or orders does not provide a general determination of the best modeling approach. However, it can be stated that the swapped time series produces the better and more robust results. This is especially observed in the prediction of *peak* flows, where the higher flood events already occur during the corresponding calibration.

The event-based model evaluation using unusual events is used to investigate whether a reduced data series is sufficient to provide good and robust predictions. This approach can help to considerably reduce computational times for data series with very high temporal resolution or length. The results show that this method is not suitable to adequately represent *snowmelt* processes. In addition to discharge, a second temperature-dependent criterion should be considered to determine unusual events. Thus, the model needs only about $20\,\%$ of the data series to derive a robust parameterization for the other discharge conditions.

Finally, the model parameterization is considered. The investigation of different model configurations show that best parameter sets are variable. Thus, the same model performance can be achieved, regardless of the complexity of model structure or the configuration of input data. However, the evaluation of snow parameters also suggests that configurations with unbalanced rain gauge distributions and unstable model results tend to assume physically implausible values or those at the parameter limits. Even the evaluation with the same model complexity and rain gauge configuration, but using many simulated precipitation time series, indicates that identical model performances are possible with unequal optimal parameter sets. Another influencing factor in determining the optimal parameter set could also be the selected time period of the calibration. This problem of equifinality should be taken into account when discussing the transferability of the “optimal” parameter set to other modeling purposes.

7 Model Reliability - Applicability for Flood Prediction

Discussions about climate change, increasingly frequent weather anomalies and extreme flood events around the world intensify the need for reliable predictions. In this context, hydrological models can and have to play an important part in the future. Nevertheless, the use of models to predict extreme flood events is always an extrapolation into the unknown of a hydrological system, since no trustworthy data for the verification of such conditions exist (*Bergström*, 1991; *Seibert*, 2003). Even if a model fits more or less perfect to observations in the calibration and the validation process there is no guarantee that the model predictions to future or till then unknown conditions perform on the same reliable level (*Oreskes et al.*, 1994). Thus, hydrological models are just useful for decision making if they provide reliable predictions for conditions where no observation data are available (*Klemeš*, 1986).

The last chapter attempts to clarify whether and to what extent the results of the investigations in this study allow a statement on the model reliability and applicability for flood prediction. In the previous chapters, numerous approaches and results relating to model and process complexity as well as data and model adequacy are presented and discussed. The flow chart in Figure 7.1 summarizes all the implemented evaluations of this study. The upper part of the diagram refers to the results of Chapter 5 and the lower part to Chapter 6. The boxes and arrows in bold represent the configurations which lead to considerable improvements of the model results in terms of model complexity, performance and robustness during the stepwise investigation. The dashed boxes in the lower part of the diagram separate the three main topics "input data analysis", "effects of data processing" and "model performance analysis". The dashed arrows refer to analyses that are not directly related to the stepwise adaptation of the data and model set-up. The upper part of the flow chart illustrates that in terms of computing time, model performance and robustness, the model configuration $FLEX\ C4_{WLS}$ proves to be the most suitable basic set-up for further investigations on flood predictions. An external spatially distributed modeling of snow processes (ESM) with subsequent spatial aggregation as a *liquid water* time series as model input seems necessary for an accurate consideration of the variable topography. The previous evaluations also show that an hourly resolution is required to represent the discharge dynamics, especially in the case of fast-reacting *peak* flows and in transition between snow and rain. A disaggregation of daily data for the spatial compression of the observation network (III) achieves significant improvements. Considerations regarding the time dependency from the block of the model performance analysis are also taken up again.

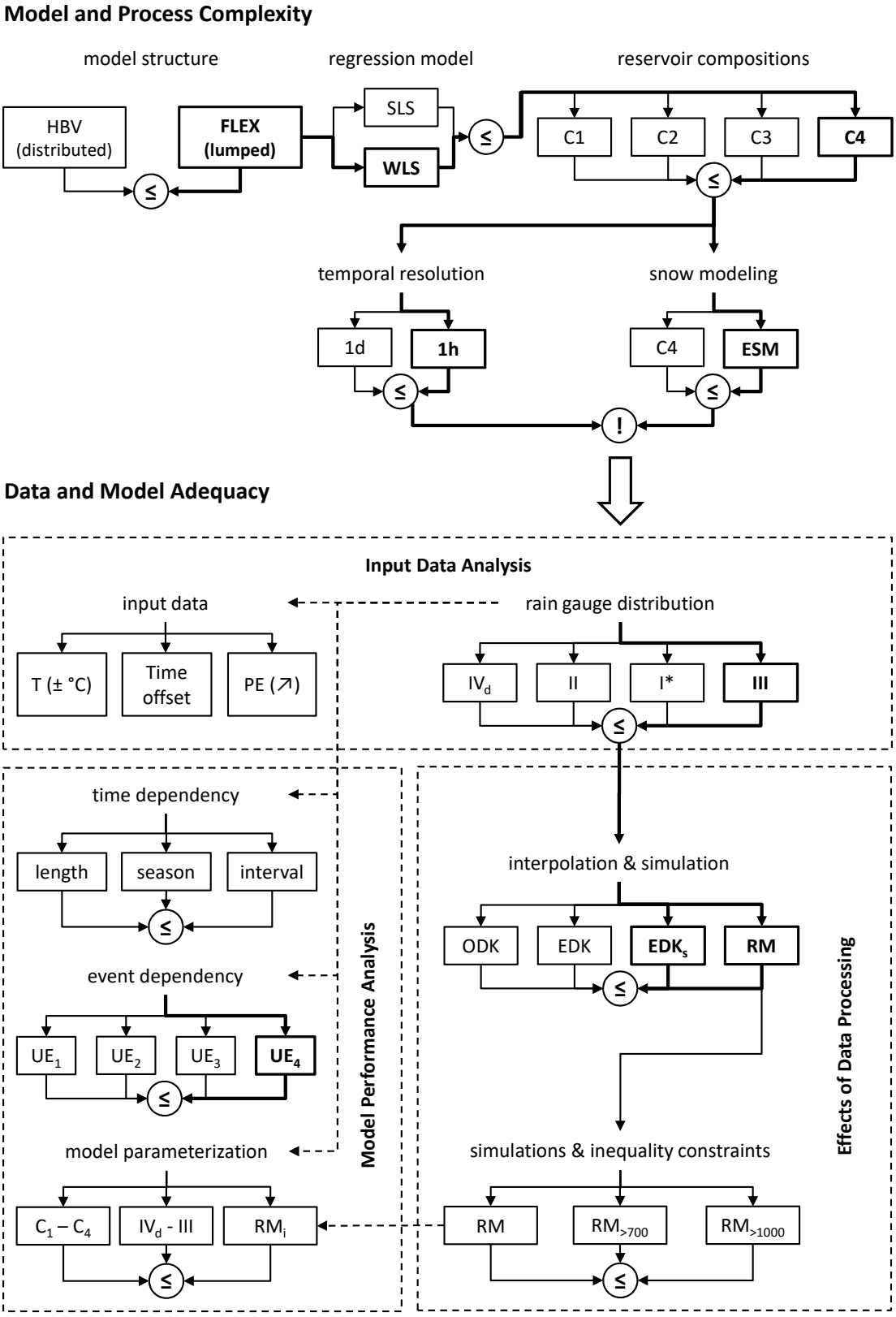
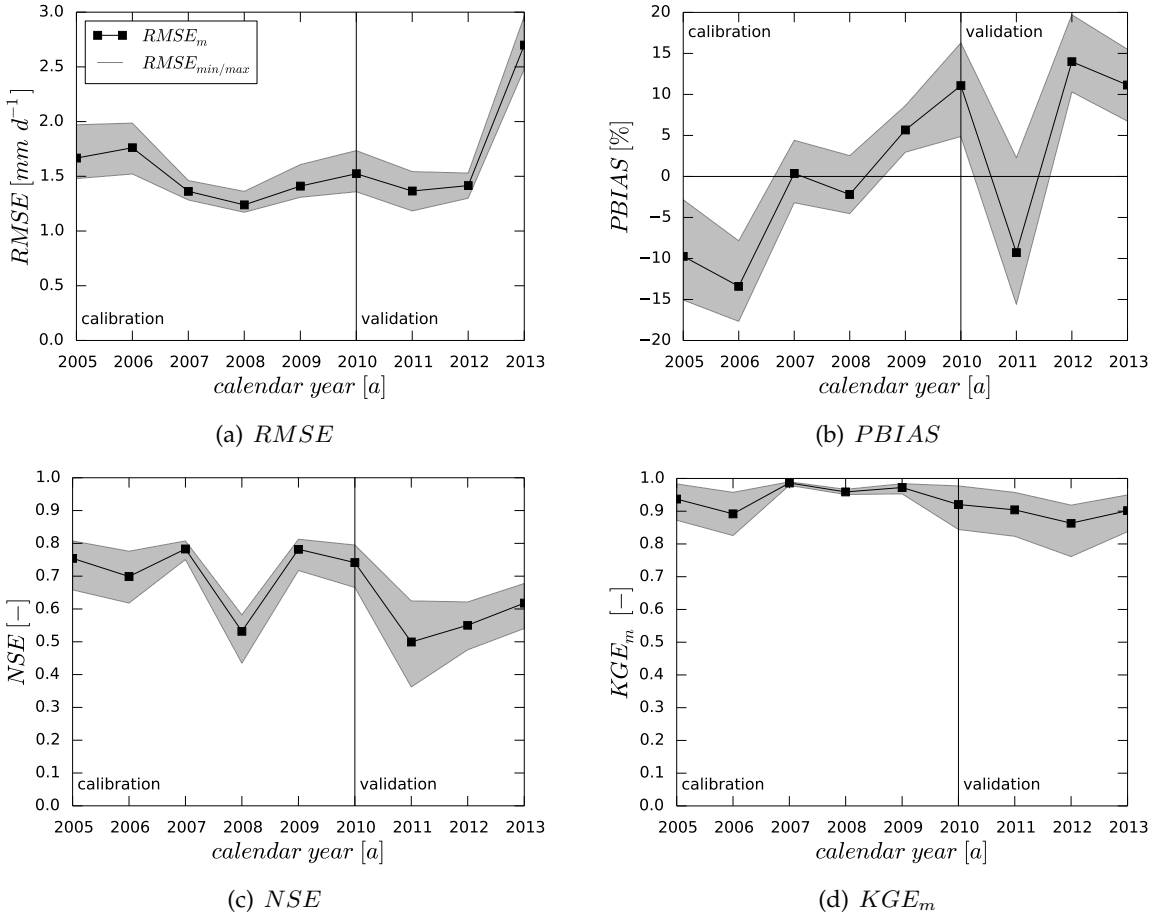


Figure 7.1: Flow chart of the performed model evaluations.

7.1 Annual Evaluation of Model Performance

A review of the results of the investigation on time dependency shows that neither the length of the time series nor a modeling of the hydrological seasons have any significant (improving) influence on the model results. The evaluation of different modeling intervals also demonstrates that longer calibration periods do not necessarily lead to a more robust and better model parameterization. The cases of inter- or extrapolation during calibration are obviously more relevant here. The swapped time series with the two highest *peaks* in calibration tend to show better results. In this context, it should be clarified whether, in addition to the highest *peak* flows, other unusual events or phases also affect the model performance and robustness in the individual years. For this reason, the model performance is first evaluated annually using 100 model runs with simulated precipitation time series (RM). This should help to identify potential weak parts of the time series. In the following, some hydrological characteristics of the time series are first summarized for a better interpretation of the results. The quality control of discharge data (Subsection 4.2.3) provides, among other things, indications of discrepancies at the gauge *Unterjettenberg* for the year 2008. In comparison to the measured values at the upstream and downstream gauges, the observed discharge volume at the gauge *Unterjettenberg* seems to be lower than those actually occurring. In the evaluation of the different modeling intervals (Subsection 6.3.1), the configuration (MI_2) also tends to perform best, where the year 2008 is neither considered in the calibration nor in the validation. The four highest *peak* flows occur during the years 2005, 2006, 2010 and 2013. By far the largest measured event is that in 2013. The wettest and driest years in terms of annual mean daily observed discharge ($Q_{obs,m}$) are 2009 with $4.3\,mm\,d^{-1}$ and 2011 with $2.7\,mm\,d^{-1}$, respectively.

Figure 7.2 presents the annual model performance based on the four measures $RMSE$, $PBIAS$, NSE and KGE_m. The vertical line indicates the start of the validation period. The highest values for $RMSE$ and thus the largest standard deviation of the model prediction error are found in the four years with the highest *peak* flows. Especially for the year 2013, in which the *peak* flow is very clearly underestimated in the predictions, the $RMSE$ shows that it tends to give more weight to higher discharges and the usually associated higher absolute errors. 2008 has the smallest value for the $RMSE$. The $PBIAS$ is in a satisfactory range of $\pm 15\,\%$ on average for the individual years. The $PBIAS$ indicates a model bias towards overestimation for the two years with the highest *peak* flows for calibration. The year 2008 again suggests very good results. A clear model bias towards underestimation is observed for the two years with the flood events and the year 2012 during validation. The year 2011 stands out with an opposite model bias in validation, which may be related to the fact that it is the driest year of the investigation period. The years 2008 and 2011 reach an unsatisfactory level for the NSE with values close to 0.5. 2010 is the only year during validation that achieves an acceptable performance level of 0.74 for the NSE. At a supposedly high performance level, the values for the KGE_m show a significantly lower sensitivity between the different years. The weakest years are 2006 and 2012. Based on the KGE_m, 2008 is also one of the best performing years. The annual evaluation shows that a very different classification of good and weak years can be made depending on the selected performance measure.



(a) $RMSE$

(b) $PBIAS$

(c) NSE

(d) KGE_m

Figure 7.2: Annual analysis of different performance measures based on model results using RM simulations (rain gauge configuration II).

The following Tables 7.1 and 7.2 present the set-up and model performance for different configurations in which the years 2008, 2011 and 2013 are disregarded for calibration or validation. These three years are selected according to the subsequent criteria. The year 2013 is chosen because of the highest flood event ever measured there. In addition, a potentially weak year should always be part of the calibration and one of the validation period. Here, the NSE is the determining performance measure for the selection of the supposedly weak years 2008 and 2011 (Figure 7.2), as this is an objective function in the optimization process of FLEX. The two performance measures $PBIAS$ and KGE_m are calculated subsequently and are therefore not directly part of the model optimization. The modified modeling intervals are marked with an asterisk (MI^*). The model performance of the associated reference configurations are given in Table 6.16.

The mean daily observed discharge ($Q_{obs,m}$) is relatively equal for the calibration and the validation period for all configurations. Only the configurations MI^*_{11}, MI^*_2 and MI^*_4 show considerably lower values during validation. The mean daily modeled discharge ($Q_{mod,m}$) matches very well during calibration. The representation of $Q_{mod,m}$ in validation is also

Table 7.1: Configuration overview and description for the different modeling intervals with disregarded years (MI^*).

ID	description	modeling interval		disregarded	no. of years	
		cal.	val.	years	cal.	val.
MI_1	basic configuration	'05 - '09	'10 - '13		5	4
MI^*_{11}				'13	5	3
MI^*_{12}				'08	4	4
MI^*_{13}				'08, '11	4	3
MI^*_2	swapped calibration	'09 - '13	'04 - '07	'11	4	4
MI^*_3	short calibration	'05 - '08	'09 - '13	'13	4	4
MI^*_4	long calibration	'05 - '10	'11 - '13	'13	6	2

good and usually tends to be marginally underestimated. The configurations MI^*_2 and MI^*_4 overestimate $Q_{mod,m}$ in validation. The $RMSE$ generally indicates a higher model prediction error for validation. In those configurations where the year 2013 is disregarded during validation, the $RMSE$ is lower than in calibration. This shows that apparently a single very unusual flood event that is not well represented by the model is sufficient to noticeably affect the model prediction error of the *total* validation period. The $PBIAS$ suggests good model results for all configurations. The model bias tends to be slightly overestimated during calibration and slightly underestimated during validation. With a model bias toward overestimation, the two configurations MI^*_2 and MI^*_4 again show deviating behavior during validation. According to the NSE, the modified basic configurations MI^*_{11-3} and MI^*_2 are improved in all modeling sequences in which individual weak years are not taken into account. The NSE increases from 0.69 to 0.73 for the configuration MI^*_{11} without 2013 in validation. It should be noted that the NSE even improves to 0.74 for configuration MI^*_{12}, although 2013 is included in the validation. The configurations MI^*_{12}, MI^*_{13} and MI^*_2, which do not take 2008 into account, seem to provide the best model results based on the NSE. The configurations MI^*_3 and MI^*_4 show a small drop in performance for calibration. Configuration MI^*_3 includes 2008 in addition to a shortened calibration phase. There is an additional significant decline in performance during validation of MI^*_4, which only considers the years 2010 and 2011. In particular, 2011 is already identified as a potentially weak year in the performance analysis of the RM simulations due to the NSE and $PBIAS$ (Figure 7.2). The variation of KGE_m is relatively small for all configurations. MI^*_2 has to be emphasized here, since it shows the strongest drop in performance during validation. This is probably due to the overestimated discharge volume, which is also indicated by the values of $Q_{mod,m}$ and the $PBIAS$.

An annual performance evaluation is also carried out for the configurations MI^*_{12} and MI^*_2 based on the previous findings. MI^*_{12} shows the most promising model performance of the modified basic configurations, and for MI^*_2 the possible overestimation of the discharge volume during validation is examined more detailed. MI_1 and MI_2 are the corresponding

Table 7.2: Performance analysis of different modeling intervals with disregarded years (rain gauge configuration III).

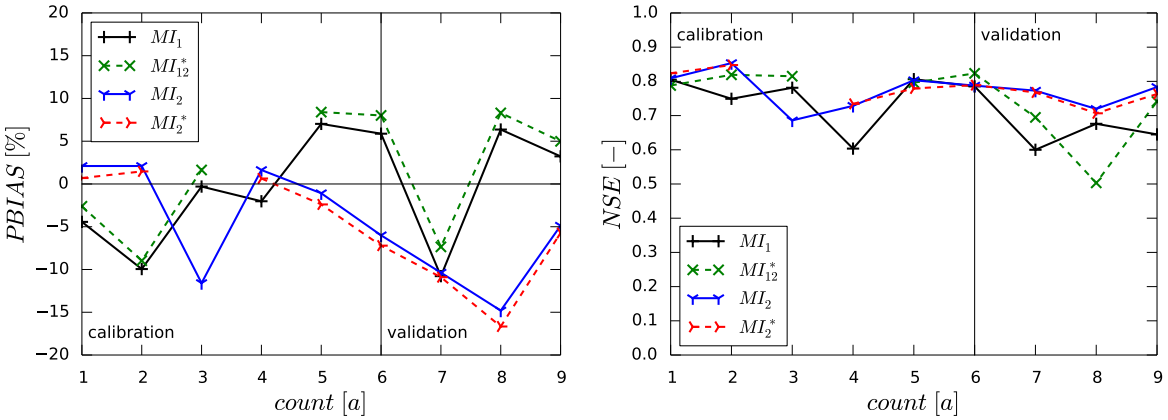
		$Q_{obs,m}$	$Q_{mod,m}$	$RMSE$	$PBIAS$	NSE	KGE_m
MI_1	*cal.*	3.7	3.7	1.4	-1.6	0.78	0.98
	val.	3.6	3.5	1.7	2.1	0.69	0.97
MI^*_{11}	*cal.*	3.7	3.7	1.4	-1.6	0.78	0.98
	val.	3.4	3.3	1.3	1.7	0.73	0.98
MI^*_{12}	*cal.*	3.8	3.8	1.4	0.0	0.81	0.99
	val.	3.6	3.4	1.6	4.4	0.74	0.97
MI^*_{13}	*cal.*	3.8	3.8	1.4	0.0	0.81	0.99
	val.	3.8	3.6	1.7	7.1	0.73	0.96
MI^*_2	*cal.*	4.0	4.0	1.4	0.1	0.80	0.99
	val.	3.5	3.9	1.5	-10.1	0.75	0.93
MI^*_3	*cal.*	3.5	3.5	1.4	-1.1	0.76	0.98
	val.	3.6	3.4	1.3	7.1	0.75	0.96
MI^*_4	*cal.*	3.6	3.8	1.5	-3.5	0.76	0.98
	val.	3.3	3.5	1.3	-3.7	0.63	0.96

reference configurations. Figure 7.3 shows the annual performance values of the $PBIAS$ and the NSE. Table 7.3 above the figure assigns the counters to the corresponding years of the chronological and swapped time series which are presented in the same figure. The values of $PBIAS$ show that those of configuration MI^*_{12} are always greater than those of MI_1. The configurations of MI_2 and MI^*_2 have the reverse behavior. Nevertheless, the deviations are relatively small for both configuration pairs. The configurations MI_1 and MI^*_{12} show an alternating behavior of over- and underestimation in a range of $\pm 10\,\%$ for the individual years during calibration as well as validation. Although configuration MI^*_2 shows values close to zero for all years of the calibration disregarding 2011, the course of values during validation is almost identical to configuration MI_2. Both tend to overestimate the discharge during all years of validation. The performance curve of the NSE is only almost identical for the configurations MI_2 and MI^*_2. In contrast, the NSE differs more clearly for the other configurations MI_1 and MI^*_{12}. Here, configuration MI^*_{12} outperforms MI_1 in almost every year. Only for the year 2012, configuration MI^*_{12} has a clear drop in performance compared to the reference configuration. Despite this clear outlier in configuration MI^*_{12}, the NSE improves by 0.05 for the *total* validation period (Table 7.2).

The annual evaluation of the 100 model runs at the beginning of this section already shows that the classification of good and weak years is difficult depending on the selected perfor-

Table 7.3: Counter assignment to the respective years for Table 7.3.

	calibration					validation			
count	1	2	3	4	5	6	7	8	9
MI_1, MI_{12}^*	2005	2006	2007	2008	2009	2010	2011	2012	2013
MI_2, MI_2^*	2009	2010	2011	2012	2013	2004	2005	2006	2007



(a) $PBIAS$ (b) NSE

Figure 7.3: Annual performance analysis for different modeling intervals with disregarded years (rain gauge configuration III).

mance measure. Nevertheless, individual conspicuous years are selected, which are disregarded for modeling. The evaluation of the model performance related to the *total* calibration or validation period shows that disregarding potential weak years does not necessarily lead to improved model results. It can also be seen that improvements are not equally indicated by all performance measures. The way how years with particularly high discharge events are taken into account also influences the model results. Disregarding the year 2013 usually leads to an improvement. It also has a positive effect on the model results if the highest discharge events occur during calibration (case of interpolation). Here, it has a greater impact than the additional disregard of a single year. MI_2 and MI_2^* have almost the same performance level. Disregarding 2011 rather causes a slightly stronger model bias toward overestimation. This may be related to the fact that 2011 is the driest discharge year of the investigation period and thus lacks crucial information on these runoff conditions. The year 2008 has a negative impact on the model performance. Possible measurement errors have already been identified in the profound data quality analysis in advance. This is confirmed by the annual evaluation and the disregard of this year. Furthermore, the evaluation of the modeled hydrograph shows that the model has massive problems to represent the *snowmelt* period correctly. Some years just show anomalies for certain performance measures, such as $PBIAS$ and KGE_m, where a correct representation of the water balance has a greater weight. The years 2006 and 2012 are to be mentioned here. On the one hand, 2012 is among

the wettest years and 2006 contains one of the highest *peak* flows. On the other hand, both years also show above-average SWE. In particular, 2012 stands out due to a very high discharge volume, especially in the first half of the year, and very many small *peak* flows. The year 2007 is always characterized by very good model performances for all measures. In terms of $Q_{obs,m}$, the year ranks in the middle range of the investigation period. The year has an extremely low snow winter and three mean flood events. It is again seen that especially the snow processes are a problem for the lumped model structure. Years with some higher flood events and a regular alternation of low and high flow conditions also seem to contain more sufficient information to activate all model parameters.

The evaluation shows that individual weak years can have a significant impact on the *total* performance of the model. An annual evaluation can be used to identify such periods. It also helps to analyze the hydrological conditions of single years and the resulting model behavior in more detail. Subsequently, it has to be decided whether these periods can be disregarded for a stabilization of the model parameterization. For example, this may be the case if there are obvious errors in the input data. However, this should be done carefully, so that periods are not left out, whose poor model representation can be traced back to other causes such as unusual hydrological conditions or structural model errors. The model analysis (e.g., MI_1 and MI_{11}^*) shows that individual performance measures, such as the $RMSE$ or NSE, react very sensitive to single extreme events. The interpretation of the model bias using the $PBIAS$ must also be done carefully. The annual evaluation of MI_1 and MI_{12}^* indicates a clearly alternating behavior of under- and overestimation for calibration and validation. The $PBIAS$ of the *total* calibration and validation period suggests another conclusion. Here, the performance would be classified much better because of compensation effects.

7.2 Most Reliable Model Set-up for Flood Prediction

The flow chart at the beginning of Chapter 7 already highlights the most promising approaches made in the previous investigations. These improvements found so far have not yet been combined and finally evaluated in one model set-up. An external snow model (ESM) and the concept of a *liquid water* time series as model input are only applied at daily resolution for rain gauge configuration I. The analyses show that an hourly resolution is required to catch the flow dynamics in such study areas with heterogeneous topography. A sufficiently large quantity of well spatially distributed rain gauge stations also has a significant positive effect on model performance and robustness. Here, the disaggregation of daily precipitation data to an hourly resolution is an appropriate tool to further densify sparse observation networks. Rain gauge configuration III outperforms the other configurations for both daily and hourly resolution. Therefore, configuration III with an hourly *liquid water* time series is analyzed as the most promising model set-up for reliable flood predictions in the first part of this section. Configuration IV with 24 stations in hourly resolution is used as reference. In the second part of this section, the model reliability is evaluated using observed and modeled flow duration curves.

Table 7.4: Hourly performance analysis of different rain gauge and *liquid water* configurations.

		$RMSE$		$PBIAS$		NSE		KGE_m	
		cal.	val.	cal.	val.	cal.	val.	cal.	val.
total	III_h	0.07	0.07	-5.2	-1.0	0.74	0.76	0.97	0.99
	$IV_{h,ESM}$	0.06	0.06	-2.5	14.8	0.79	0.78	0.99	0.88
	$III_{h,ESM}$	0.05	0.05	-2.6	3.1	0.82	0.87	0.99	0.99
snow	III_h	0.09	0.08	8.6	13.3	0.39	0.33	0.91	0.86
	$IV_{h,ESM}$	0.08	0.08	14.9	33.9	0.59	0.28	0.87	0.41
	$III_{h,ESM}$	0.07	0.05	13.7	20.5	0.62	0.65	0.89	0.78
no snow	III_h	0.05	0.06	-13.3	-7.6	0.80	0.81	0.91	0.96
	$IV_{h,ESM}$	0.05	0.06	-12.8	6.0	0.82	0.85	0.91	0.98
	$III_{h,ESM}$	0.05	0.05	-12.3	-5.0	0.85	0.89	0.92	0.99
peak	III_h	0.10	0.16	-21.7	6.5	0.88	0.90	0.76	0.98
	$IV_{h,ESM}$	0.11	0.20	-17.3	21.1	0.86	0.84	0.85	0.78
	$III_{h,ESM}$	0.11	0.10	-24.4	2.9	0.88	0.96	0.70	1.00

Table 7.4 first summarizes the evaluation of performance measures for different hourly model results. The values of $RMSE$ are very similar for the respective configurations during calibration and validation. The evaluation of the *no snow* period results in slightly better values than for the *snowmelt* period. Except for the configuration $III_{h,ESM}$, the $RSME$ is significantly higher for the *peak* flows during validation. The interpretation of the results using the $PBIAS$ is not quite so clear for the different investigation periods as well as configurations. The model tends to slightly overestimate the volume for all configurations during calibration. The $PBIAS$ of the *total* calibration as well as validation period is close to zero for the configurations III_h and $III_{h,ESM}$. A clear tendency for one or the other is not so obvious, since the *total* period is well represented by both approaches. Configuration $IV_{h,ESM}$ provides the worst results here, especially in terms of robustness. It has the most significant model bias toward underestimation in all periods during validation. $PBIAS_{snow}$ presents a volume underestimation by the model for all configurations, both in calibration and validation. This is even more distinct for the two approaches with ESM, although theoretically more precipitation volume is available to the model in this period. Here, the results of the *snowmelt* period would suggest the FLEX configuration with an internal snow model. In contrast, $PBIAS_{no\ snow}$ shows a volume overestimation for almost all modeling sequences. The *no snow* period has better results for configuration $III_{h,ESM}$. The $PBIAS_{peak}$ displays the already frequently observed trend of overestimation during calibration and underestimation during validation. The $PBIAS$ of validated *peak* flows also tends to the configuration with an ESM. The NSE provides robust results at a high performance level for all investigated time periods, except for the *snowmelt* period. Only configuration $III_{h,ESM}$ shows significant

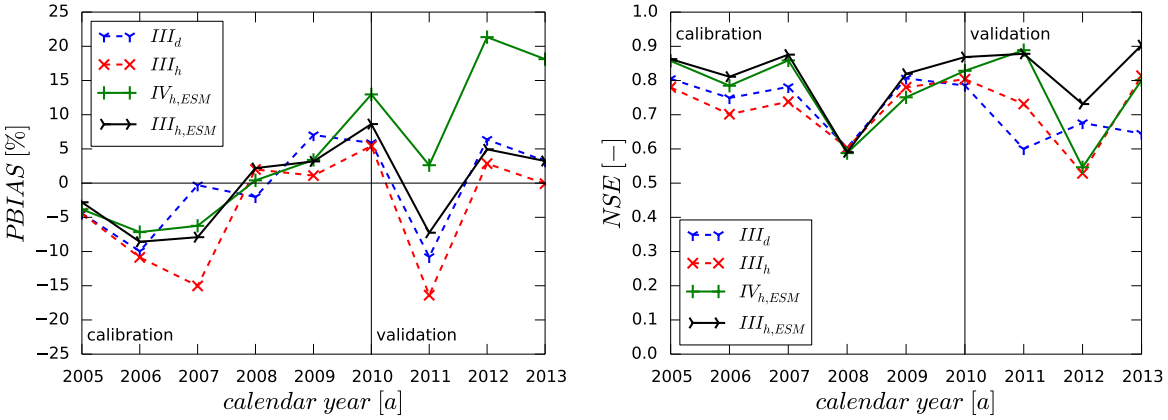
Table 7.5: Mean areal precipitation depth (h_{MAP}) and peak error (EP) for different hourly rain gauge and *liquid water* configurations.

	unit	III_h		$IV_{h,ESM}$		$III_{h,ESM}$	
		cal.	val.	cal.	val.	cal.	val.
$h_{MAP,tot}$	$[mm]$	7947	6012	8150	5675	7959	6026
$h_{MAP,snow}$	$[mm]$	1852	1161	2752	1652	2719	1832
$h_{MAP,no\ snow}$	$[mm]$	6095	4851	5397	4023	5240	4194
$h_{MAP,peak}$	$[mm]$	514	467	498	413	514	470
EP	[-]	0.85	0.85	0.75	0.66	0.89	0.90

improvements and acceptable results for *snowmelt*. The two approaches with an ESM generally produce better results. The rain gauge configuration has a greater influence on model performance when representing *peak* flows. The configurations III outperform the configuration $IV_{h,ESM}$ regardless of the consideration of the snow process. Based on the NSE, configuration $III_{h,ESM}$ performs best. The KGE_m basically reflects a similar trend, where configurations III, especially with an ESM, outperform configuration $IV_{h,ESM}$. Here, the values of KGE_m suggest an almost perfect model fit for the *total* time period. Only for the *peak* flows does configuration $IV_{h,ESM}$ perform better in calibration, whereas configurations III then again show very good model performances in validation.

Table 7.5 lists the mean areal precipitation depth (h_{MAP}) as well as the peak error (EP) for different hourly rain gauge configurations. Configurations III_h and $III_{h,ESM}$ result in about $135\,mm$ to $160\,mm$ more precipitation in the system over the *total* modeling period of nine years. Nevertheless, the configurations III with more rain gauge stations do not necessarily lead to more h_{MAP} for all periods and modeling sequences. Configuration $IV_{h,ESM}$ always assumes higher h_{MAP} in calibration for the model approaches with ESM, except for the *peak* flows. The precipitation depth for the *total* as well as the *snowmelt* period is also the highest during calibration for all model set-ups. In contrast, both configurations with a denser gauging network produce considerably higher h_{MAP} for the *total* validation period. The greatest differences occur in the way snow processes are accounted for. There is a clear shift of h_{MAP} toward the *snowmelt* period for both approaches with an ESM. h_{MAP} is almost identical for configurations III_h and $III_{h,ESM}$ during the *peak* events. These two configurations have significantly more precipitation volume than configuration $IV_{h,ESM}$, especially during validation. The comparison of precipitation depth (h_{MAP}) as input and volume error of the model ($PBIAS$) indicate that not only the absolute available precipitation volume is important, but also its temporal distribution and thus the storage and release of the individual model reservoirs. This is particularly evident during the *snowmelt* period, where despite significantly more h_{MAP} available, there is a larger volume underestimation in the approaches with ESM. Nevertheless, the NSE shows a clear model stabilization during validation for configuration $III_{h,ESM}$. The results of EP clearly argue for both configurations III_h and $III_{h,ESM}$ in terms of performance level and robustness. Configuration $IV_{h,ESM}$ has a significant drop in performance here. $III_{h,ESM}$ shows the best model perfor-

mance at a very good level. The evaluation in Subsection 6.1.1 already provides the findings that denser and spatially well-distributed rain gauge stations improve the model results. This is also confirmed when comparing configurations $III_{h,ESM}$ and $IV_{h,ESM}$ at hourly resolution. The application of an ESM also leads to improvements. However, the station configuration appears to have the slightly greater influence on model performance than the ESM when modeling *peak* flows. In conclusion, configuration $III_{h,ESM}$ suggests the best and most robust model output.

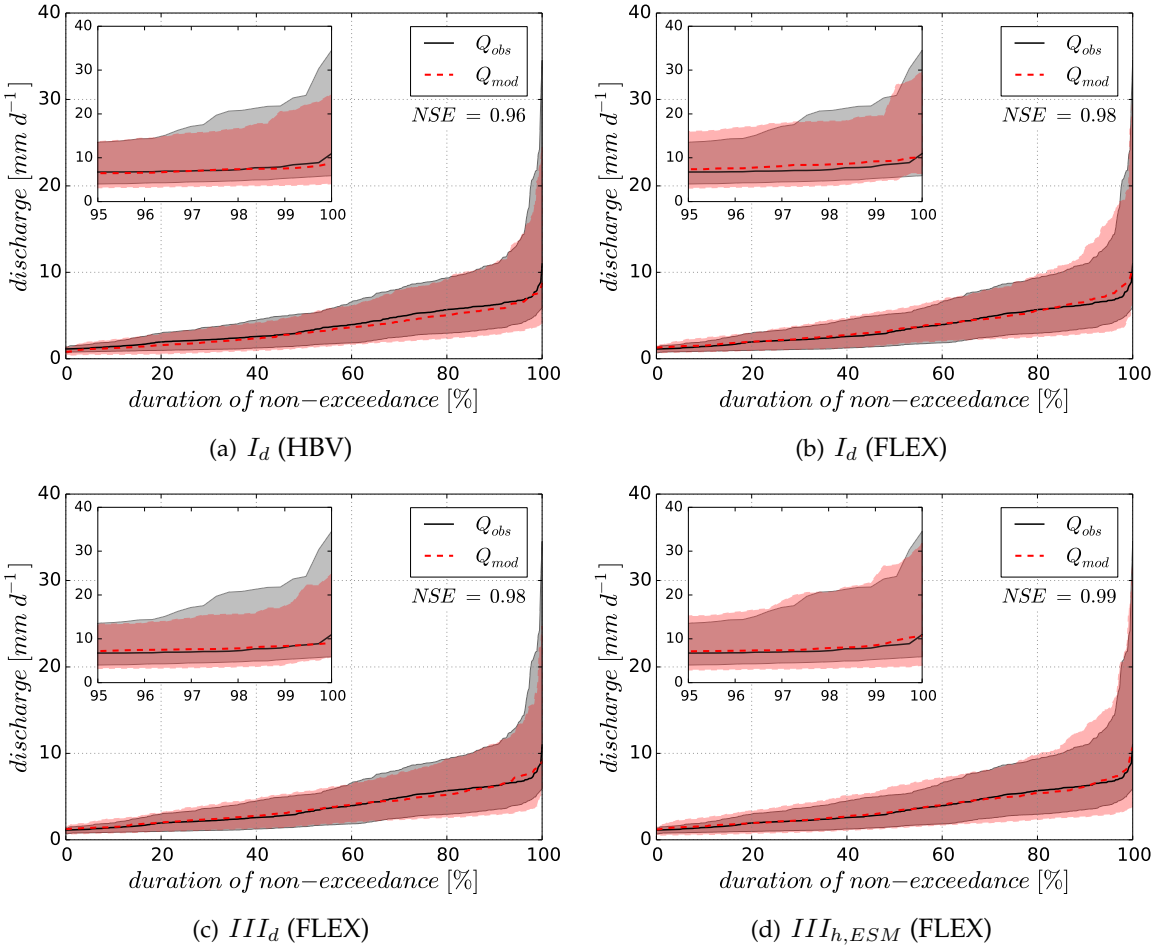


(a) $PBIAS$ (b) NSE

Figure 7.4: Annual performance analysis for different rain gauge and *liquid water* configurations.

The annual evaluation of model performance has already identified individual years with potentially poor influence on the model results in the previous Section 7.1. Figure 7.4 now shows the annual evaluation for different rain gauge and *liquid water* configurations in hourly resolution. The purpose of this presentation ($PBIAS$ and NSE) is to determine if there are any changes in annual model performance based on temporal resolution (see also Figure 7.2 and 7.3). Configuration III_d is used as reference at daily resolution. The comparison of configurations III_d and III_h shows a consistent behavior for the $PBIAS$ during validation and for the NSE during calibration. The respective other modeling sequence does not give such an uniform result. The largest difference is found by analyzing the $PBIAS$ in 2007. Here, III_h has the most significant model error toward volume overestimation for all configurations considered. The annual values of the $PBIAS$ have a similar course for all four configurations during validation. All approaches with rain gauge configuration III overestimate the volume for 2011. All other years show a model error toward underestimation. The result is a mean value that is close to zero over all years. This trend is consistent with the annual evaluations on daily basis. Configuration IV_h underestimates the discharge volume in all validated years, in some cases quite significantly. The evaluation of the NSE demonstrates that the configurations with ESM tend to perform better. $III_{h,ESM}$ outperforms all others in both calibration and validation. The two configurations III_d and III_h achieve roughly the same results. The comparison based on the NSE indicates that the

year 2008 is almost equally poorly represented regardless of temporal resolution and rain gauge configuration. This further confirms the assumption that the discharge data for this year are erroneous. The model performance for 2011 is supposedly weak when modeled on a daily resolution. The performance improves based on NSE for all hourly model approaches in 2011. Particularly significant improvements are seen in both $PBIAS$ and NSE for the two versions with ESM. The year 2011 has the lowest annual mean daily observed discharge ($Q_{obs,m}$) and has below mean SWE in the first half of the year. The annual hydrograph is characterized by several small events. The highest *peak* flow already occurs in mid-January. This may indicate that there are many short-term alternating conditions of snow and rain during the year. A higher temporal resolution and a spatially distributed consideration of snow accumulation and melt seems to have a positive effect on the model output here. The year 2012 shows a decrease in performance for both $PBIAS$ and NSE, especially for configuration IV_h. However, no clear relationship to the rain gauge configuration or the consideration of the snow processes can be established here.



(a) I_d (HBV)

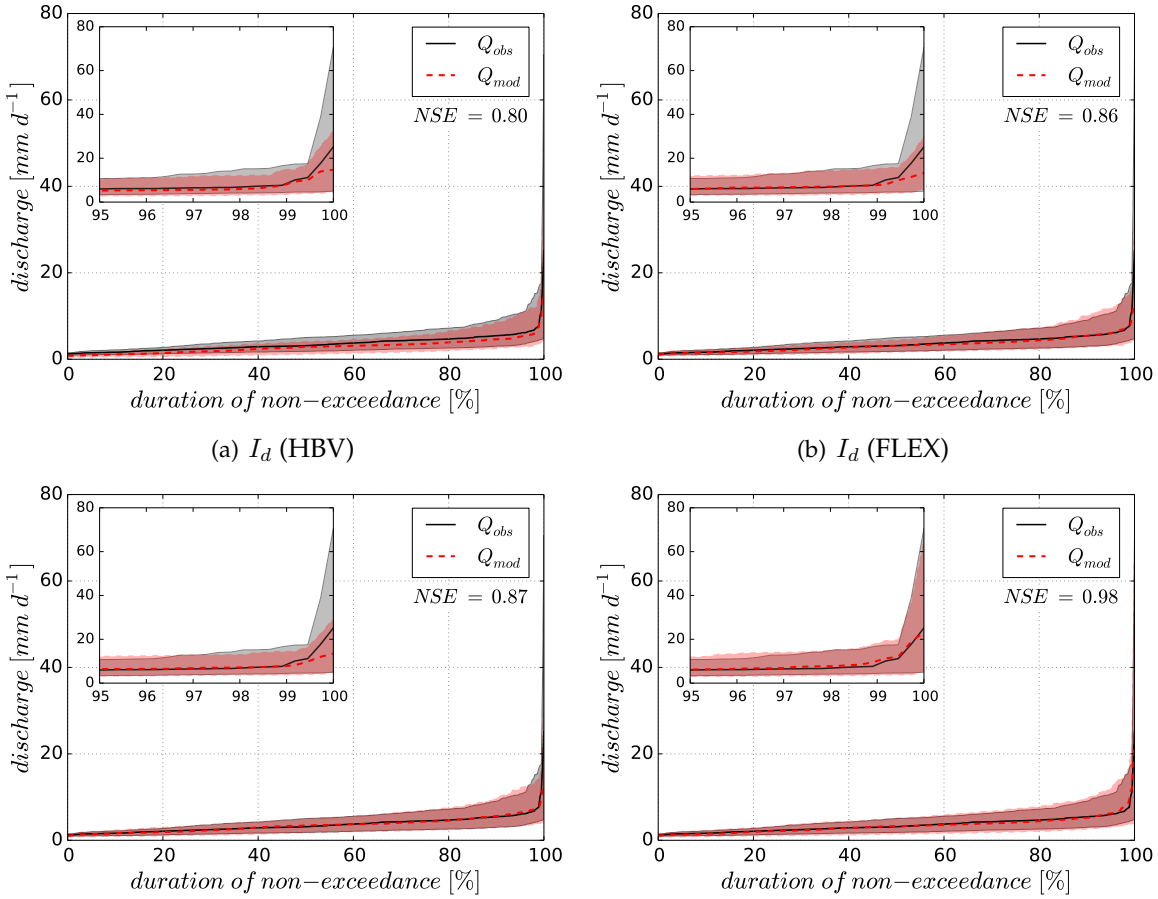
(b) I_d (FLEX)

(c) III_d (FLEX)

(d) $III_{h,ESM}$ (FLEX)

Figure 7.5: Flow duration curve of different modeling approaches for the calibration period 2005 - 2009.

Flow duration curves are already presented in Chapter 2 as a possible tool to display the general match of the discharge characteristics or rather frequencies over a selected time period. The used data pairs are not compared time-dependent but according to their non-exceedance probability. The introduction of the data availability in Chapter 3 already shows that the investigation period is representative for the long-term discharge characteristics at the gauge *Unterjettenberg*. Even higher flood events have been observed within the selected period, which is useful in terms of examining predictive reliability for extreme events. This section concludes with an evaluation of model reliability using observed and modeled flow duration curves. The objective is to verify the extent to which the duration of non-exceedance of observed and modeled discharge match over the calibration or the validation period for different modeling approaches.



(a) I_d (HBV)

(b) I_d (FLEX)

(c) III_d (FLEX)

(d) $III_{h,ESM}$ (FLEX)

Figure 7.6: Flow duration curve of different modeling approaches for the validation period 2010 - 2013.

The results are presented in Figure 7.5 for calibration and Figure 7.6 for validation. The solid black line defines the mean duration of non-exceedance of the observed discharge data for the respective modeling period. The red dashed line corresponds to the modeled

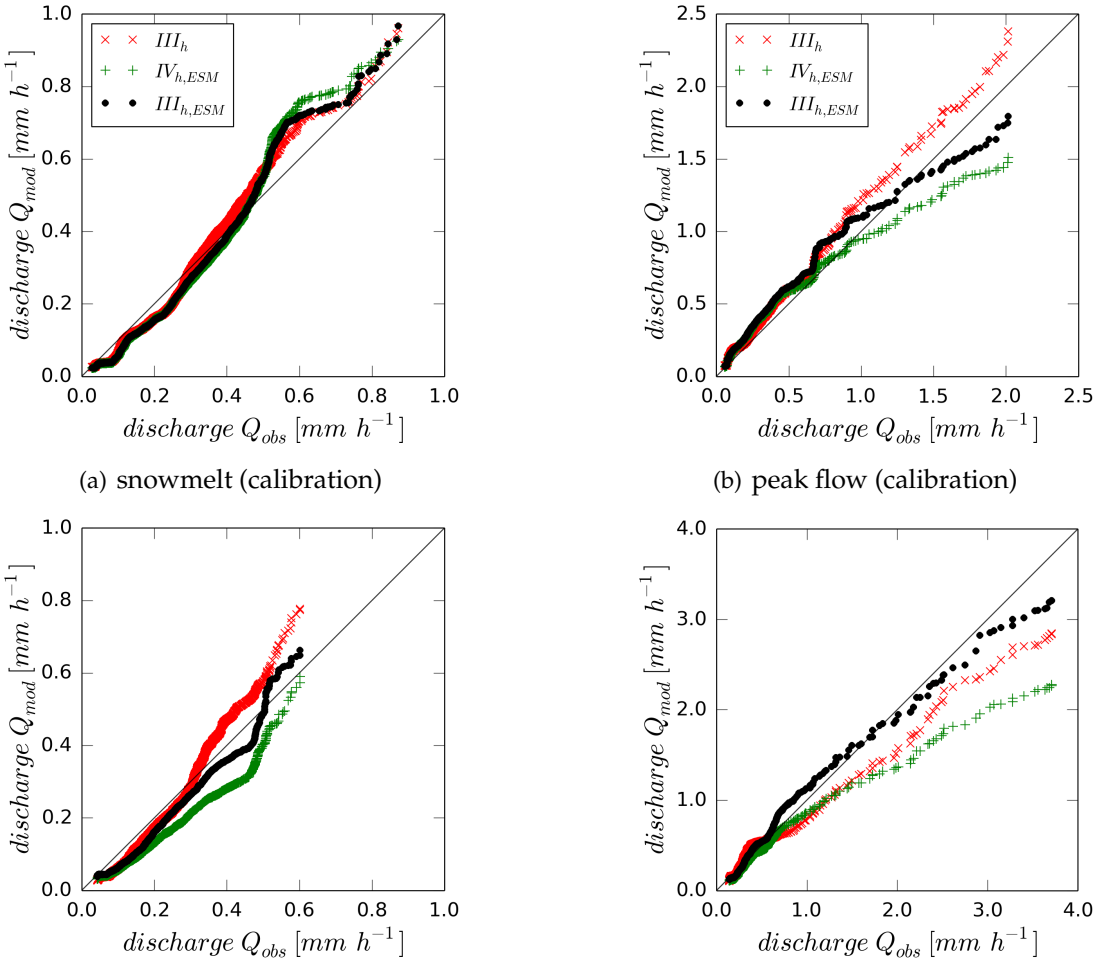
counterpart. The gray and red shaded area represents the related range of lower and upper hull curve. The enlarged section of the figure corresponds to the upper $5\,\%$ duration of non-exceedance. The NSE is given as a quantitative measure for the different model approaches. The order of the selected modeling approaches starts with the comparison of distributed (HBV) and lumped (FLEX) model structure with rain gauge configuration I_d at daily resolution. The third subfigure represents the most promising rain gauge configuration III_d. The results for configuration $III_{h,ESM}$ with ESM in hourly resolution are shown in the fourth subfigure. Both calibration and validation have increasing model performances according to the preceding order of modeling approaches. The value of NSE is on a very good level for all four configurations during calibration. The performance differences are more obvious during validation. There is a clear jump between distributed and lumped model structure as well as configuration III_d and $III_{h,ESM}$.

A closer look at the flow duration curve during calibration suggests that the approach using the HBV model (I_d) is most likely to underestimate the discharge conditions. Configuration I_d (FLEX) tends to overestimate discharge the most, especially at the higher discharge conditions. Configuration III_d appears to have the most promising model results up to about $80\,\%$ probability of non-exceedance, also taking the range of the hull curve into account. Configuration $III_{h,ESM}$ best represents the very high discharges. The statistics of the mean discharge over the calibration period are presented very similarly to III_d. However, it can be seen that configuration $III_{h,ESM}$ is a bit more prone to overestimating discharge in some areas of the upper boundary. The four selected model approaches have basically the same tendency during validation as during calibration. The problem with underestimated discharges becomes even more pronounced when using the HBV model. The first three configurations indicate the same weaknesses toward underestimation for the highest *peak* discharges. Here, configuration $III_{h,ESM}$ clearly outperforms all others, which is also confirmed by the very good value of NSE. The evaluation of the flow duration curve shows that configuration $III_{h,ESM}$ is the only one to provide very good results for both level of model performance and robustness. Based on the comparison of observed and modeled discharge statistics, it can be summarized that $III_{h,ESM}$ is the most reliable modeling approach for the prediction of extreme events.

7.3 Development of Snowmelt Periods and Peak Flows

The last section of this study again addresses the two identified weaknesses of the rainfall-runoff modeling for the selected study area: *snowmelt* periods and *peak* flows. Finally, for the two discharge characteristics, the development is shown that results from different modeling approaches, which are influenced by data availability, quality and processing, as well as model and process complexity.

Figure 7.7 shows the predictive reliability of the two discharge conditions for three modeling approaches with hourly resolution. The three configurations III_h, $IV_{h,ESM}$ and $III_{h,ESM}$ are used for this evaluation. The two left subfigures represent the results for calibration and validation during the *snowmelt* period. For the calibration, the Q-Q plot reproduces a



(a) snowmelt (calibration)

(b) peak flow (calibration)

(c) snowmelt (validation)

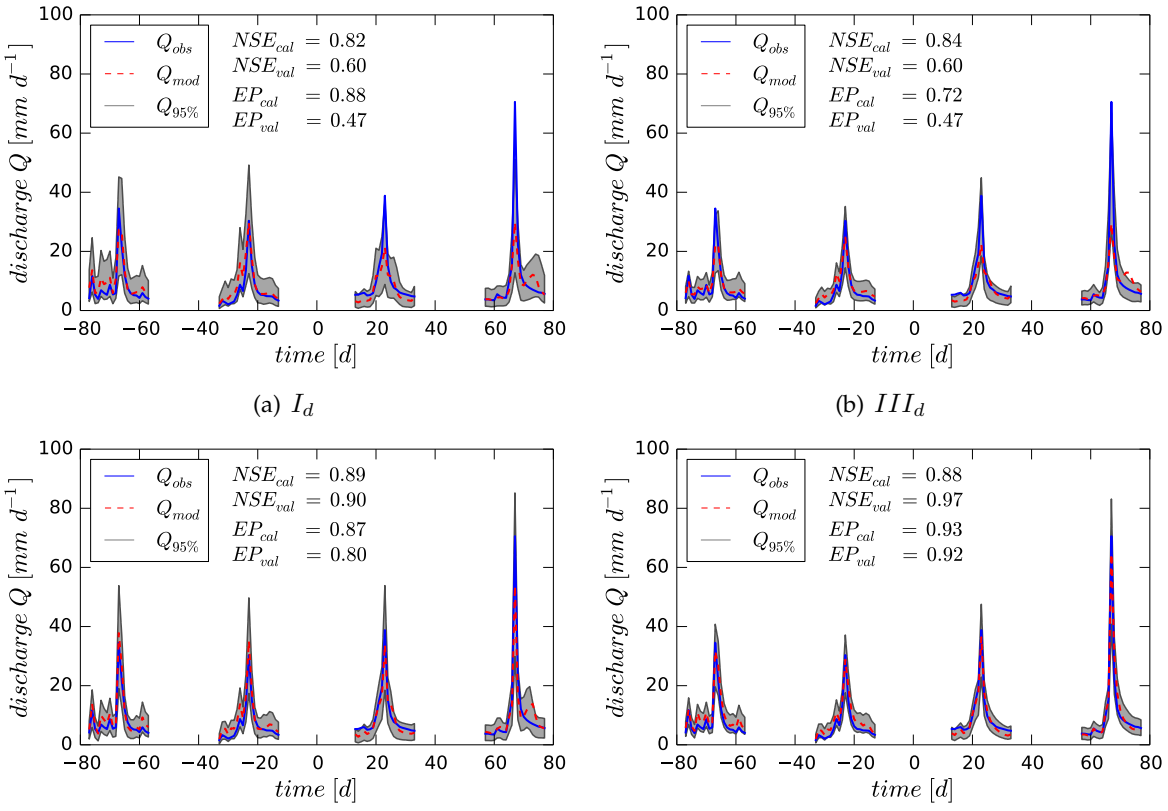
(d) peak flow (validation)

Figure 7.7: Q-Q plots for assessing the predictive reliability of hourly model results with and without ESM.

relatively good and similar curve for all three configurations. All three configurations have in common that the medium-high discharges are more clearly overestimated. The range of low discharges is slightly underestimated by all three approaches. Configuration III_h generally tends to overestimate discharge the most. For the high discharge, it is configuration $IV_{h,ESM}$. The predictive reliability of validation differs significantly more between the three configurations for the *snowmelt* period. It can also be observed that the maximum discharges are considerably lower than during the calibration phase. For configuration III_h, the predicted distribution at the low discharges still agrees very well with the observations. At the higher discharges, however, a continuously increasing overestimation is observed. Configuration $IV_{h,ESM}$ shows an opposing trend toward underestimation. Then, the high discharge values again agree much better with the observed distribution. The distribution of configuration $III_{h,ESM}$ generally is much closer to the 1:1 line. Thus, $III_{h,ESM}$ provides the most reliable results for considering *snowmelt* periods. The general observation that low

discharges tend to be underestimated and high discharges tend to be overestimated may indicate that the release of water from the snow reservoir is initially too delayed and then over-represented by the model as it progresses.

Q-Q plots of the *peak* flows are shown in the two subfigures on the right (Figure 7.7). The predictive reliability of all three configurations agrees well with observations at low discharges, both in calibration and validation. Configuration III_h again continuously overestimates the discharge distribution during calibration. The other configurations underestimate the distribution above medium discharge conditions. This fact is much more pronounced for configuration $IV_{h,ESM}$. Looking at the prediction uncertainty during validation, configuration III_h now has an opposing behavior toward underestimation. Configuration $IV_{h,ESM}$ shows the greatest predictive uncertainty for *peak* flows during validation. Configuration $III_{h,ESM}$ has the best agreement with the observed discharge distribution. The analysis of predictive uncertainty for the two discharge conditions *snowmelt* and *peak* flows concludes that configuration $III_{h,ESM}$ generates the most reliable model predictions. Particularly, when considering *peak* flows, it is apparent that a combination of rain gauge configuration and consideration of snow processes lead to significant improvements. Here, a simple graphical representation by Q-Q plots is a suitable tool to compare the predictive reliability of different modeling approaches.



(a) I_d (b) III_d

(c) III_h (aggregated) (d) $III_{h,ESM}$ (aggregated)

Figure 7.8: Peak flow development using different FLEX model configurations.

To conclude this study, the major developments for modeling *peak* flows are shown in Figure 7.8 and Table 7.6. The three main improving points concerning spatially well-distributed rain gauge stations, application of a spatially distributed external snow model (ESM) and higher temporal resolution are compared here. In addition, the predictive uncertainty of the individual model approaches is evaluated. In contrast to the previous evaluation with Q-Q plots, this interpretation is time-dependent and takes the uncertainty bands into account. In Table 7.6, two additional measures ($COVER$ and $SHARPNESS$) are presented for the evaluation of the predictive uncertainty. An extensive mathematical derivation of these quantities is omitted here. Further information about the methodology and the measures are provided in various publications (*Gneiting et al.*, 2007; *Pinson et al.*, 2007; *Thordarson et al.*, 2012; *Del Giudice et al.*, 2015). The $SHARPNESS$ describes the mean distance between the lower and upper uncertainty bands. The measure $COVER$ indicates the percentage of validation data included within the $95\,\%$ uncertainty bands.

Table 7.6: Evaluation of predictive reliability with additional performance measures for the peak flows during validation.

	unit	I_d	III_d	III_h	$III_{h,ESM}$
$SHARPNESS$	$[mm\,d^{-1}]$	11.0	8.6	10.0	7.8
$COVER$	[%]	93	88	95	98

Figure 7.8 compares four different $FLEX$ model configurations. Hourly model results are aggregated to daily resolution. In addition to the graphical comparison, the two performance measures NSE and EP are also given for the individual model runs. Configurations I_d and III_d are characterized by different rain gauge configurations as daily input data for the model. Although Configuration III_d shows significant improvements for the *total* period and other discharge conditions in the evaluations of this study so far, this is not as evident for the *peak* flows on daily resolution. In particular, *peak* flows during calibration are more underestimated by III_d. Even though the two main *peaks* of III_d are represented worse, the NSE indicates slightly better results. The better modeled conditions prior to the flood event are one reason for this. The results of the validation do not show clear differences, which is also confirmed by the NSE and EP. The uncertainty bands of III_d are narrower than I_d, but do not cover all observations. This is also verified for validation by the two performance measures in Table 7.6. Configuration III_h has clearly better model results. The NSE indicates a robust performance on a very good level. Although the *peak* flows are still underestimated during the validation, they are already much better represented, which can also be seen for the value of EP. The $SHARPNESS$ is slightly worse compared to the approach at daily resolution. On the other Hand, the performance measure $COVER$ improves. In all three model approaches considered so far, it is noticeable that after the main *peak* of the last validated event (2013), a second smaller *peak* is always assumed by the model. A considerable part of the discharge volume is released with a delay during this flood event. Obviously, this is due to incorrect assumptions of snow accumulation and melt processes, since this second *peak* disappears in the last model approach with ESM ($III_{h,ESM}$). The spatial and temporal resolution seems to be of great influence, as stable snow packs do not

usually form over the entire elevation profile of the catchment at this time of the year. Here, the limitations of a lumped model structure are that it can only assume snow or rain at one time step, which is solved by applying an ESM. In addition, both *peak* flows during validation are very well represented. Regarding the EP, a further improvement is achieved, which is also reflected in robust model outputs. The two performance measures $SHARPNESS$ and $COVER$ show the best results in terms of predictive reliability.

The successive adjustment of the model set-up develops the representation of the *peak* flows very successfully. Both daily configurations (I_d and III_d) do not suggest robust and reliable model outputs at the beginning. The underestimated *peaks* during validation are absolutely unacceptable for the prediction of flood events. The model calculations in hourly resolution then result in significant improvements. *Del Giudice et al.* (2015) considers predictions to be reliable if $COVER$ is greater-equal $95\,\%$. This is valid for configurations III_h and $III_{h,ESM}$ within this evaluation. With respect to all applied evaluation criteria, configuration $III_{h,ESM}$ provides the most reliable predictions.

7.4 Summary

The objective of the last chapter of this study is to make a statement about the predictive reliability of different modeling approaches, especially for flood events. Figure 7.1 provides an overview of all relevant data processing and model choice approaches performed in this study. Based on the most promising findings so far, evaluations of model performance, model reliability, and the development of inadequate represented discharge conditions have now been performed.

An annual analysis of model performance helps identifying unusual phases of the investigation time period that may be caused by extreme hydrological conditions or data inadequacies. For this purpose, simulation runs generated by RM are used, since a single model run can be erroneous due to other reasons. One of the main findings of this evaluation is that the identification of supposedly weak years can vary considerably. Depending on the performance measure, an assessment can even be completely contrary. Using this procedure, conspicuous years are identified and disregarded during calibration and validation in various model approaches. The evaluation shows that disregarded individual years does not necessarily improve model performance. A major influence is whether the highest *peak* flows are considered during calibration (interpolation) or validation (extrapolation). Thus, disregarding individual years must be done very carefully in order not to lose important unusual hydrological information. A poorer model performance can also reveal structural model deficiencies of reproducing certain hydrological conditions. Nevertheless, this approach is also able to identify years whose poor model representation is apparently due to erroneous input data. Furthermore, the analysis shows that some measures (e.g., $PBIAS$) suggest a good model performance over the *total* period. However, this is only due to compensating effects of alternating underestimated and overestimated years, which would be considered as insufficient in a year-by-year evaluation.

In a next step, the improvements found are combined in one model configuration. The objective is to determine the most reliable model set-up for flood prediction. Here, a model approach with hourly resolution and a spatially distributed external calculation of snow processes is appropriate. Furthermore, the spatial densification of the hourly observation network with daily rain gauge stations yields a significant benefit. These modifications of the model set-up provide a considerably increase in model performance and finally the required robustness for reliable predictions. The annual evaluation of model performance, also realized on an hourly basis, using the $PBIAS$ and the NSE show the basically same potential deficiencies within the selected investigation period. The poor performance of 2008 is also confirmed by the NSE for all hourly configurations. For 2011, clear improvements can be achieved with the application of an ESM in hourly resolution. Based on the NSE, it can be summarized that the approaches with ESM have better performances. Especially for the representation of *snowmelt* periods and *peak* flows, acceptable to very good results are achieved with configuration $III_{h,ESM}$. This indicates that a high spatial and temporal resolution of input data and model set-up is required to represent certain hydrological processes that are rapidly changing both spatially and temporally. Finally, the general agreement of the discharge characteristics is evaluated using time-independent flow duration curves. The selected order of modeling approaches has an increasing model performance for both calibration and validation. Configuration $III_{h,ESM}$ is the only one to provide very good results in terms of performance level as well as robustness. Due to the very good agreement of the non-exceedance probability, it is consequently the most reliable modeling approach for the prediction of extreme events.

The previous evaluations show that the model calculations on a daily time step can lead to an inadequate representation of the catchment dynamics. Daily values account for a uniform distribution of input data (e.g., precipitation and temperature) over the day which can be problematic if the hydrological response of the catchment takes considerably less than a day. Such an influence of temporal resolution or consideration of snow processes is also a source of uncertainty. In addition to a required high temporal resolution, the combination of sufficiently dense rain gauge networks and the consideration of spatially distributed snow processes is beneficial. Thus, an increase in predictive reliability can be achieved. The reasons for the mostly underestimated *peak* flows are complex. One reason is certainly erroneous observations for the events with high precipitation intensities, since precipitation measurements can be very uncertain due to spatial variations. This can cause a lack of precipitation volume in the input data. In some cases, there is enough modeled discharge volume available, which is simply distributed incorrectly in time. Insufficient model structures or process parameterizations can be responsible for this. The end of this chapter shows the extent to which the interaction of uncertainty sources and measures to reduce them can yield robust and reliable approaches for modeling flood events. The successive adjustment of the model set-up shows very clearly how the evaluation of the *peak* flows has developed from absolutely unacceptable to very good and robust.

8 Conclusions

Efficient flood protection and the corresponding flood risk management are becoming increasingly important due to more frequent and severe flood events. For this purpose, powerful and reliable hydrological models are required, which are adapted to the on-site situation and serve as a prediction tool for planners and decision-makers. Previous studies show that hydrological models are often not yet at a satisfactory and reliable level for operational use.

The motivation for this work arose from a project dealing with emergency planning and evacuation measures for urban areas and transport infrastructure during extreme flood events. The selected study area has a very heterogeneous and alpine topography. The task of hydrology was to find a robust and reliable modeling approach to generate previously unobserved hydrometeorological extreme events. Recent developments in hydrological modeling show that the trend is towards increasingly complex model structures with higher data and parameter demand. The accompanying, frequently one-sided, focus on improving the performance of calibration results and often inadequate model validation, can pretend to provide the necessary conditions for reliable predictions. This thesis focuses not only on improving absolute model performance by applying various model and process modifications. Rather, the objective was to identify and understand where and why certain modeling approaches fail in order to subsequently find solutions for improvements. The common evaluation of model performance is based on the modeled output data. This often involves an attempt to correct the discrepancies between observed and modeled data by adjusting the model structure and parameters. This approach is based on the widespread assumption that the error variance is basically caused by the modeled data and that the measured values are assumed to be error-free. For this reason, this thesis pays special attention to the input data and its processing before it is passed to the used models.

At the beginning of the investigations, two major shortcomings of the applied model configurations could be identified. On the one hand, all configurations, regardless of their complexity, have considerable deficits in generating robust model results. This is noticeable by a significant gap in model performance between calibration and validation. On the other hand, clear problems in the correct reproduction of certain discharge conditions are evident. The water balance of alpine catchments is strongly influenced by the storage and release of water in the form of snow and ice. From a hydrological perspective, *snowmelt* is also of particular interest, as it has a high flood potential in spring with overlying intense and long-lasting precipitation. Initial modeling attempts yielded unacceptable results for the defined *snowmelt* period. Based on the task and related objectives, the representation of the highest *peak* flows is crucial. Many modeling approaches have shown that the *peak* flows are significantly underestimated during calibration and especially during validation. The *peak* flows

are often not even captured by the uncertainty bands during validation. Such model results are not a basis for reliable predictions of previously unobserved flood events.

In order to systematically identify the potential causes for the aforementioned deficits, a comprehensive quality control of the input data is first performed. Geostatistical interpolation methods of different complexity are compared. A new conditional simulation method (*random mixing* (RM)) for the spatial interpolation of precipitation fields is applied. This stochastic approach allows for a stronger consideration of measurement uncertainties by incorporating linear inequality (*greater-equal*) constraints. The data analysis shows that regardless of the used rain gauge configuration and the applied interpolation method, there is a significant deficit in the water balance. However, using the configuration with the most and most spatially uniformly distributed rain gauge stations already leads to a reduction of the water deficit, especially in the winter half-years. The consideration of topography as additional information in the kriging methods has an additional positive effect on the water balance. Another aspect is that accumulated spatial precipitation amounts are usually not affected by the small-scale variability of topography. Therefore, a smoothed digital elevation model (DEM) is used to correct for the implausible replication of the underlying small-scale terrain structure in the precipitation fields. Precipitation fields simulated using RM reduce the water deficit by more than half in this study. The water balance can be almost balanced if *greater-equal* constraints are considered for the simulations at the highest located rain gauge stations. Thus, RM incorporating *greater-equal* constraints is suitable to better describe measurement uncertainties, especially under difficult topographic boundary conditions.

The application of the two different model structures demonstrates that a comparatively simple lumped model approach (FLEX) is not fundamentally inferior to the more complex spatially distributed structure of the HBV model. A major advantage of the FLEX approach is the fairly straightforward adaptation of the model structure and application to different temporal resolutions. The relatively fast computation times of the model are another benefit. Thus, model runs at daily resolution with hundreds of simulated precipitation time series can be performed with a reasonable expenditure of time. Nevertheless, regardless of the model complexity, robust results could not be obtained in many modeling attempts. This is an indication that this problem is not primarily caused by the model structure or parameterization. The same is valid for the insufficiently represented discharge conditions. For the modeling of *snowmelt* periods, the spatially distributed consideration of the input data precipitation and temperature in the HBV model proves to be more advantageous. From this finding, the concept of *liquid water* time series is developed. For this purpose, the flexibility of the FLEX structure is used by disabling the snow process in the model. Instead, an external snow model (ESM) is developed using the degree-day method, which calculates the accumulation and melting processes in a spatially distributed grid. This allows simultaneous consideration of different aggregate states of precipitation and snow processes within the catchment. The spatially distributed water availability from different sources is then aggregated to a spatially lumped value per time step. The resulting *liquid water* time series can then be passed to the FLEX model. This method yields promising results on a daily basis, although the model performance for *snowmelt* is better but still unsatisfactory. It can be concluded that for a very heterogeneous and alpine topography, a lumped representation

of snow processes alone is not sufficient. The combination of a lumped basic framework for modeling the main runoff components and the individual outsourcing of complex processes proves to be a suitable tool here.

Due to its relatively fast computation times, the FLEX model is very well suited for modeling with higher temporal resolution. Therefore, the influence of the temporal resolution is also investigated in detail within the scope of this work. The existing rain gauge stations with hourly data are very sparse, which is why the observation network is additionally densified with disaggregated data from daily stations. The daily data are disaggregated using the *nearest-neighbor* method, which allows to determine the hourly precipitation distribution at each daily station. The evaluated hourly model runs indicate more robust results for all applied rain gauge configurations. This suggests that in smaller catchments a high temporal resolution is required to represent the fast changing discharge conditions sufficiently well. The additional spatial densification of the observation network by disaggregated daily data thus has a positive effect on model performance and robustness.

Another aspect of this thesis is a comprehensive, often neglected, assessment of data and model adequacy to gain a better understanding of their implications for hydrological modeling. A larger number of performance measures than usual are first applied and evaluated. On the one hand, it should not be ignored that performance measures can be very sensitive to single, inadequately modeled (flood) events. On the other hand, it may be that systematic overestimation or underestimation of discharge conditions by the model has hardly any negative effect on the performance level of the individual quality measures. It has been shown that evaluations based only on one performance measure (e.g., NSE) can lead to wrong decisions concerning the quality and reliability of model output. Other measures, such as the modified Kling-Gupta efficiency (KGE_m), have rather low sensitivity between different model approaches in this study. This can also be difficult to interpret in the evaluation and lead to wrong decisions. Measures such as $PBIAS$, which describe model bias in terms of volume overestimation or underestimation, also require some caution. An apparently very good volume balance over the *total* modeling period may be caused only by compensation effects of individual clearly diverging and relatively poor results, as shown by annual evaluations of model performance.

The time-dependent model evaluation concludes that the length of the time series, hydrological seasons, or different modeling intervals are only of minor relevance. Here, sufficiently frequent and as varying as possible hydrological conditions, such as severe flood events and extended low water phases, are rather decisive. These should occur as equally as possible during the selected calibration and validation period. Seasonal modeling using hydrological summer and winter half-years shows that the parameter sets obtained provide different information for the hydrological counterpart. This is an example of the difficulty in determining an unique parameter set describing different flow conditions equally well. Thus, one approach could be to optimize different parameter sets separately based on different discharge conditions. The different parameter sets can then be applied according to the prevailing conditions. Such an approach is demonstrated using time series with reduced data input determined by unusual hydrometeorological events. Except for snow processes, robust parameterization for specific discharge conditions can be achieved with a significant

reduction in data use and computation time. The problem of equifinality is also addressed in the analysis of model parameterization. Here, the precipitation time series simulated using RM prove to be suitable for obtaining a sufficient number of model runs with identical model performance in both calibration and validation. Thus, it can be shown that the optimal parameter sets can vary. This finding can impede a clear decision on the best modeling approach and must also be considered with respect to the transferability of parameters to other modeling purposes.

Finally, the most promising modeling approaches are assembled from all these individual findings. Based on this stepwise adjustment of the modeling approaches, a conclusion on the predictive reliability is drawn. The three most important modifications that contribute to a substantially improved development in both *snowmelt* and *peak* flows are: Model calculations in hourly resolution, a spatial densification of the observation network with disaggregated daily data, and an external spatially distributed snow model (ESM). The main sources of uncertainty can be attributed to these three influencing factors in this study. The importance of high-quality spatial and temporal data collection as a basis for reliable modeling approaches becomes very apparent. It is a misleading assumption that increasingly complex model structures can replace the use of real data. The most accurate model is useless if it is not supplied and verified with correct data. The time-independent evaluation with flow duration curves confirms that the configuration which includes the aforementioned improving modifications best reproduces the statistical discharge characteristics. The evaluation of the highest *peak* flows, taking into account the uncertainties, also shows that only in this way a reliable modeling approach is achieved, which appears to be applicable for flood predictions. Nevertheless, predictions of previously unobserved extreme flood events must be interpreted cautiously and applied in decision-making, as they represent an extrapolation into the unknown with increasing uncertainties.

This work shows the adequacy of different approaches to data processing and model choice to progress from initially unacceptable to very good and robust model results. Based on many different evaluations, it can be seen which factors have to be considered in hydrological modeling and which seem to play a rather minor role. The aim of this thesis is to investigate the catchment characteristics, the required data base, the process and model behavior very comprehensively. Therefore, all investigations were performed on a single catchment. The results of this study should support a more targeted approach to data processing and model choice in similar applications. For further uses, the proposed solutions and their validity are to be investigated on a larger number of catchments with different characteristics as well as climatic conditions.

Bibliography

Aarts, E. H. L., and J. Korst (1990), *Simulated annealing and Boltzmann machines: A stochastic approach to combinatorial optimization and neural computing*, reprinted ed., XII, 272 pp., Wiley, Chichester [et. al.].

Allen, R. G., L. S. Pereira, D. Raes, and M. Smith (1998), Crop evapotranspiration – Guidelines for computing crop water requirements, *FAO Irrigation and drainage paper*, *56*, ISBN 92-5-104219-5.

Anderson, E. A. (1973), National Weather Service River Forecast System - Snow Accumulation and Ablation Model, *NOAA Technical Memorandum NWS HYDRO-17*, 217pp.

Andréassian, V., C. Perrin, and C. Michel (2004), Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models, *Journal of Hydrology*, *286*(1–4), 19–35, doi:10.1016/j.jhydrol.2003.09.030.

Andréassian, V., C. Perrin, L. Berthet, N. Le Moine, J. Lerat, C. Loumagne, L. Oudin, T. Mathevet, M.-H. Ramos, and A. Valéry (2009), HESS Opinions "Crash tests for a standardized evaluation of hydrological models", *Hydrology and Earth System Sciences Discussions*, *6*(3), 3669–3685, doi:10.5194/hessd-6-3669-2009.

Bárdossy, A., and T. Das (2008), Influence of rainfall observation network on model calibration and application, *Hydrology and Earth System Sciences*, *12*, 77–89, doi:10.5194/hess-12-77-2008.

Bárdossy, A., and S. Hörning (2016a), Gaussian and non-Gaussian inverse modeling of groundwater flow using copulas and random mixing, *Water Resources Research*, *52*, 4504–4526, doi:10.1002/2014WR016820.

Bárdossy, A., and S. Hörning (2016b), Random Mixing: An Approach to Inverse Modeling for Groundwater Flow and Transport Problems, *Transport in Porous Media*, *114*(2), 241–259, doi:10.1007/s11242-015-0608-4.

Bárdossy, A., and G. Pegram (2013), Interpolation of precipitation under topographic influence at different time scales, *Water Resources Research*, *49*, 4545–4565, doi:10.1002/wrcr.20307.

Bárdossy, A., and S. K. Singh (2008), Robust estimation of hydrological model parameters, *Hydrology and Earth System Sciences*, *12*, 1273–1283.

Bárdossy, A., Y. Huang, and T. Wagener (2016), Simultaneous calibration of hydrological models in geographical space, *Hydrology and Earth System Sciences*, *20*, 2913–2928, doi:10.5194/hess-20-2913-2016.

Bergström, S. (1991), Principles and confidence in hydrological modelling, *Nordic Hydrology*, *22*, 123–136.

Bergström, S. (1992), THE HBV MODEL – its structure and applications, *Reports Hydrology No. 4*, Swedish Meteorological and Hydrological Institute (SMHI), ISSN 0283-1104.

Beven, K. (1993), Prophecy, reality and uncertainty in distributed hydrological modelling, *Advances in Water Resources*, *16*, 41–51, doi:10.1016/0309-1708(93)90028-E.

Beven, K. (2006), A manifesto for the equifinality thesis, *Journal of Hydrology*, *320*(1–2), 18–36, doi:10.1016/j.jhydrol.2005.07.007.

Beven, K. (2007), Towards integrated environmental models of everywhere: uncertainty, data and modelling as a learning process, *Hydrology and Earth System Sciences*, *11*(1), 460–467, doi:10.5194/hess-11-460-2007.

Beven, K. (2012), *Rainfall-Runoff Modelling: The Primer*, 2. ed., Wiley-Blackwell, Chichester, doi:10.1002/9781119951001.

Blöschl, G., and R. Kirnbauer (1991), Point snowmelt models with different degrees of complexity – internal processes, *Journal of Hydrology*, *129*(1–4), 127–147, doi:10.1016/0022-1694(91)90048-M.

BMI (2013), Bericht zur Flutkatastrophe 2013: Katastrophenhilfe, Entschädigung, Wiederaufbau, `https://www.bmi.bund.de/DE/startseite/startseite-node.html`, last access via search: 2022-02-18.

BMU (2003), *Hydrologischer Atlas von Deutschland (HAD)*, Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit, Postfach 12 06 29, 53048 Bonn, ISBN 3-00-005624-6.

Brath, A., A. Montanari, and E. Toth (2004), Analysis of the effects of different scenarios of historical data availability on the calibration of a spatially-distributed hydrological model, *Journal of Hydrology*, *291*(3–4), 232–253, doi:10.1016/j.jhydrol.2003.12.044.

Braun, L. N., and H. Lang (1986), Simulation of snowmelt runoff in lowland and lower Alpine regions of Switzerland, *IAHS Publications*, *155*, 125–140.

Chvíla, B., B. Sevruk, and M. Ondrás (2005), The wind-induced loss of thunderstorm precipitation measurements, *Atmospheric Research*, *77*(1–4), 29–38, doi:10.1016/j.atmosres.2004.11.032.

Clarke, R. T. (1973), A review of some mathematical models used in hydrology, with observations on their calibration and use, *Journal of Hydrology*, *19*(1), 1–20, doi:10.1016/0022-1694(73)90089-9.

Das, T., A. Bárdossy, E. Zehe, and Y. He (2008), Comparison of conceptual model performance using different representations of spatial variability, *Journal of Hydrology*, *356*(1–2), 106–118, doi:10.1016/j.jhydrol.2008.04.008.

Del Giudice, D., P. Reichert, V. Bareă, C. Albert, and J. Rieckermann (2015), Model bias and complexity - Understanding the effects of structural deficits and input errors on runoff predictions, *Environmental Modelling & Software*, *64*, 205–214, doi:10.1016/j.envsoft.2014.11.006.

Doherty, J., and D. Welter (2010), A short exploration of structural noise, *Water Resources Research*, *46*(5), doi:10.1029/2009WR008377, W05525.

Einfalt, T., and S. Michaelides (2008), Quality control of precipitation data, in *Precipitation: Advances in Measurement, Estimation and Prediction*, edited by S. Michaelides, pp. 101–126, Springer Berlin Heidelberg, doi:10.1007/978-3-540-77655-0_5.

Einfalt, T., N. Gerlach, C. Podlasly, and N. Demuth (2008), Rainfall and Climate Data Quality Control, in *11th International Conference on Urban Drainage, Edinburgh, Scotland, UK*.

Estévez, J., P. Gavilán, and J. Giráldez (2011), Guidelines on validation procedures for meteorological data from automatic weather stations, *Journal of Hydrology*, *402*(1–2), 144–154, doi:10.1016/j.jhydrol.2011.02.031.

Fenicia, F., H. H. G. Savenije, P. Matgen, and L. Pfister (2008), Understanding catchment behavior through stepwise model concept improvement, *Water Resources Research*, *44*(1), doi:10.1029/2006WR005563, W01402.

Fenicia, F., D. Kavetski, and H. H. G. Savenije (2011), Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, *Water Resources Research*, *47*(11), doi:10.1029/2010WR010174, W11510.

Franz, K. J., P. Butcher, and N. K. Ajami (2010), Addressing snow model uncertainty for hydrologic prediction, *Advances in Water Resources*, *33*(8), 820–832, doi:10.1016/j.advwatres.2010.05.004.

Gan, T. Y., and G. F. Biftu (1996), Automatic Calibration of Conceptual Rainfall-Runoff Models: Optimization Algorithms, Catchment Conditions, and Model Structure, *Water Resources Research*, *32*(12), 3513–3524, doi:10.1029/95WR02195.

Gan, T. Y., E. M. Dlamini, and G. F. Biftu (1997), Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling, *Journal of Hydrology*, *192*(1–4), 81–103, doi:10.1016/S0022-1694(96)03114-9.

Garrick, M., C. Cunnane, and J. E. Nash (1978), A criterion of efficiency for rainfall-runoff models, *Journal of Hydrology*, *36*(3–4), 375–381, doi:10.1016/0022-1694(78)90155-5.

Gayathri, K. D., B. P. Ganasri, and G. S. Dwarakish (2015), A Review on Hydrological Models, *Aquatic Procedia*, *4*, 1001–1007, doi:10.1016/j.aqpro.2015.02.126.

Gharari, S., M. Hrachowitz, F. Fenicia, H. Gao, and H. H. G. Savenije (2014), Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration, *Hydrology and Earth System Sciences*, *18*(12), 4839–4859, doi:10.5194/hess-18-4839-2014.

Gneiting, T., F. Balabdaoui, and A. E. Raftery (2007), Probabilistic forecasts, calibration and sharpness, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *69*(2), 243–268, doi:10.1111/j.1467-9868.2007.00587.x.

Götzinger, J., and A. Bárdossy (2007), Comparison of four regionalisation methods for a distributed hydrological model, *Journal of Hydrology*, *333*(2–4), 374–384, doi:10.1016/j.jhydrol.2006.09.008.

Götzinger, J., and A. Bárdossy (2008), Generic error model for calibration and uncertainty estimation of hydrological models, *Water Resources Research*, *44*(12), doi:10.1029/2007WR006691, W00B07.

Gupta, H. V., S. Sorooshian, and P. Yapo (1999), Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration, *Journal of Hydrologic Engineering*, *4*(2), 135–143, doi:10.1061/(ASCE)1084-0699(1999)4:2(135).

Gupta, H. V., T. Wagener, and Y. Liu (2008), Reconciling theory with observations: elements of a diagnostic approach to model evaluation, *Hydrological Processes*, *22*(18), 3802–3813, doi:10.1002/hyp.6989.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *Journal of Hydrology*, *377*(1–2), 80–91, doi:10.1016/j.jhydrol.2009.08.003.

Gupta, H. V., M. P. Clark, J. A. Vrugt, G. Abramowitz, and M. Ye (2012), Towards a comprehensive assessment of model structural adequacy, *Water Resources Research*, *48*, doi:10.1029/2011WR011044, W08301.

Gupta, V. K., and S. Sorooshian (1985), The relationship between data and the precision of parameter estimates of hydrologic models, *Journal of Hydrology*, *81*(1–2), 57–77, doi:10.1016/0022-1694(85)90167-2.

Häckel, H. (2008), *Meteorologie*, 6th ed., 447 pp., Ulmer, Stuttgart.

Hargreaves, G. H., and Z. A. Samani (1985), Reference crop evapotranspiration from temperature, *Applied Engineering in Agriculture*, *1*(2), 96–99, doi:10.13031/2013.26773.

Harmel, R. D., R. J. Cooper, R. M. Slade, R. L. Haney, and J. G. Arnold (2006), Cumulative Uncertainty in Measured Streamflow and Water Quality Data for Small Watersheds, *Transactions of the ASABE*, *49*(3), 689–701, doi:10.13031/2013.20488.

Hawkins, D. M. (2004), The Problem of Overfitting, *Journal of Chemical Information and Computer Sciences*, *44*(1), 1–12, doi:10.1021/ci0342472, PMID: 14741005.

Hock, R. (2003), Temperature index melt modelling in mountain areas, *Journal of Hydrology*, *282*(1–4), 104–115, doi:10.1016/S0022-1694(03)00257-9, Mountain Hydrology and Water Resources .

Hörning, S., J. Sreekanth, and A. Bárdossy (2019), Computational efficient inverse groundwater modeling using Random Mixing and Whittaker–Shannon interpolation, *Advances in Water Resources*, *123*, 109–119, doi:10.1016/j.advwatres.2018.11.012.

Hrachowitz, M., O. Fovet, L. Ruiz, T. Euser, S. Gharari, R. Nijzink, J. Freer, H. H. G. Savenije, and C. Gascuel-Odoux (2014), Process consistency in models: The importance of system signatures, expert knowledge, and process complexity, *Water Resources Research*, *50*(9), 7445–7469, doi:10.1002/2014WR015484.

Hundecha, Y., and A. Bárdossy (2004), Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, *Journal of Hydrology*, *292*(1–4), 281–295, doi:10.1016/j.jhydrol.2004.01.002.

Iqbal, M. (1983), *An introduction to solar radiation*, XVIII, 390 pp., Academic Pr., Toronto [et al.].

Jain, S. K., and K. P. Sudheer (2008), Fitting of Hydrologic Models: A Close Look at the Nash – Sutcliffe Index, *Journal of Hydrologic Engineering*, *13*(10), 981–986, doi:10.1061/(ASCE) 1084-0699(2008)13:10(981).

Jakeman, A. J., and G. M. Hornberger (1993), How much complexity is warranted in a rainfall-runoff model?, *Water Resources Research*, *29*(8), 2637–2649, doi:10.1029/93WR00877.

Jónsson, P., A. Petersen-Øverleir, E. Nilsson, M. Edström, H. L. Iversen, and H. Sirvio (2002), Methodological and personal uncertainties in the establishment of rating curves, in *Nordic Hydrological Conference 2002. Volume 1. Nordic Hydrological Programme, NHP Report No. 47*, edited by Å. Killingtveit, pp. 35–44, XXII Nordic Hydrological Conference, Røros, Nordic Association for Hydrology.

Kavetski, D., and M. P. Clark (2010), Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction, *Water Resources Research*, *46*(10), doi:10.1029/2009WR008896, W10511.

Kavetski, D., and F. Fenicia (2011), Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights, *Water Resources Research*, *47*(11), doi:10.1029/2011WR010748, W11511.

Kavetski, D., S. W. Franks, and G. Kuczera (2002), Confronting Input Uncertainty in Environmental Modelling, in *Calibration of watershed models*, vol. 6, edited by Q. Duan et al., pp. 49–68, American Geophysical Union, Washington, DC.

Kavetski, D., G. Kuczera, and S. W. Franks (2006a), Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts, *Journal of Hydrology*, *320*(1–2), 173–186, doi:10.1016/j.jhydrol.2005.07.012.

Kavetski, D., G. Kuczera, and S. W. Franks (2006b), Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resources Research*, *42*(3), doi:10.1029/2005WR004368, W03407.

Kavetski, D., G. Kuczera, and S. W. Franks (2006c), Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, *Water Resources Research*, *42*(3), doi:10.1029/2005WR004376, W03408.

Kirnbauer, R., G. Blöschl, and D. Gutknecht (1994), Entering the Era of Distributed Snow Models, *Nordic Hydrology*, *25*, 1–24.

Kitanidis, P. K. (1997), *Introduction to Geostatistics: Applications to Hydrogeology*, 1st ed., XX, 249 pp., Cambridge Univ. Press, Cambridge.

Klemeš, V. (1986), Operational testing of hydrological simulation models, *Hydrological Sciences Journal*, *31(1)*, 13–24, doi:10.1080/02626668609491024.

Krause, P., D. P. Boyle, and F. Bäse (2005), Comparison of different efficiency criteria for hydrological model assessment, *Advances in Geosciences*, *5*, 89–97, doi:10.5194/adgeo-5-89-2005.

Kult, J., W. Choi, and J. Choi (2014), Sensitivity of the Snowmelt Runoff Model to snow covered area and temperature inputs , *Applied Geography*, *55*, 30–38, doi:10.1016/j.apgeog.2014.08.011.

Lecher, K., H.-P. Lühr, and U. C. E. Zanke (Eds.) (2015), *Taschenbuch der Wasserwirtschaft: Grundlagen, Planungen, Maßnahmen*, 9th ed., XLI, 1305 pp., Springer Vieweg, Wiesbaden, ISBN 978-3-528-12580-6.

Legates, D. R., and G. J. McCabe (1999), Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation, *Water Resources Research*, *35*, 233–241, doi:10.1029/1998WR900018.

Lindström, G., B. Johansson, M. Persson, M. Gardelin, and S. Bergström (1997), Development and test of the distributed HBV-96 hydrological model, *Journal of Hydrology*, *201*(1–4), 272–288, doi:10.1016/S0022-1694(97)00041-3.

Linsley, R. K., M. A. Kohler, and J. L. Paulhus (1988), *Hydrology for Engineers*, 3rd ed., SI metric ed., 1. pr. ed., XX, 492 pp., McGraw-Hill, London.

Maniak, U. (1997), *Hydrologie und Wasserwirtschaft: Eine Einführung für Ingenieure*, 4th ed., XVI, 650 pp., Springer, Berlin.

McCuen, R. H., Z. Knight, and A. Cutter (2006), Evaluation of the Nash-Sutcliffe Efficiency Index, *Journal of Hydrologic Engineering*, *11*(6), 597–602, doi:10.1061/(ASCE)1084-0699(2006)11:6(597).

McMillan, H., B. Jackson, M. Clark, D. Kavetski, and R. Woods (2011), Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models, *Journal of Hydrology*, *400*(1–2), 83–94, doi:10.1016/j.jhydrol.2011.01.026.

Merz, R., and G. Blöschl (2003), A process typology of regional floods, *Water Resources Research*, *39*(12), doi:10.1029/2002WR001952, 1340.

Michelson, D. B. (2004), Systematic correction of precipitation gauge observations using analyzed meteorological variables, *Journal of Hydrology*, *290*(3–4), 161–177, doi:10.1016/j.jhydrol.2003.10.005.

Molini, A., P. L. Barbera, L. Lanza, and L. Stagi (2001), Rainfall intermittency and the sampling error of tipping-bucket rain gauges, *Physics and Chemistry of the Earth, Part C: Solar, Terrestrial & Planetary Science*, *26*(10–12), 737–742, doi:10.1016/S1464-1917(01)95018-4.

Molini, A., L. G. Lanza, and P. La Barbera (2005), The impact of tipping-bucket raingauge measurement errors on design rainfall for urban-scale applications, *Hydrological Processes*, *19*(5), 1073–1088, doi:10.1002/hyp.5646.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith (2007), Model Evaluation Guidelines For Systematic Quantification of Accuracy in Watershed Simulations, *American Society of Agricultural and Biological Engineers*, *50*(3), 885–900, ISSN 0001–2351.

Morton, A. (1993), Mathematical Models: Questions of Trustworthiness, *The British Journal for the Philosophy of Science*, *44*(4), 659–674.

Moulin, L., E. Gaume, and C. Obled (2009), Uncertainties on mean areal precipitation: assessment and impact on streamflow simulations, *Hydrology and Earth System Sciences*, *13*(2), 99–114, doi:10.5194/hess-13-99-2009.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models. Part 1: A discussion of principles, *Journal of Hydrology*, *10*, 282–290.

Neumann, M. B., and W. Gujer (2008), Underestimation of Uncertainty in Statistical Regression of Environmental Models: Influence of Model Structure Uncertainty, *Environ. Sci. Technol.*, *42*(11), 4037–4043, doi:10.1021/es702397q.

Nocedal, J., and S. J. Wright (2006), *Numerical optimization*, 2. ed., XXII, 664 pp., Springer, New York, NY, ISBN 978-0-387-30303-1.

Oreskes, N., K. Shrader-Frechette, and K. Belitz (1994), Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences, *Science, New Series*, *263*(5147), 641–646, Published by: American Association for the Advancement of Science.

Oudin, L., C. Michel, and F. Anctil (2005a), Which potential evapotranspiration input for a lumped rainfall-runoff model? – Part 1: Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs?, *Journal of Hydrology*, *303*(1–4), 275–289, doi:10.1016/j.jhydrol.2004.08.025.

Oudin, L., F. Hervieu, C. Michel, C. Perrin, V. Andréassian, F. Anctil, and C. Loumagne (2005b), Which potential evapotranspiration input for a lumped rainfall-runoff model? – Part 2: Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling, *Journal of Hydrology*, *303*(1–4), 290–306, doi:10.1016/j.jhydrol.2004.08.026.

Oudin, L., C. Perrin, T. Mathevet, V. Andréassian, and C. Michel (2006), Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models, *Journal of Hydrology*, *320*(1–2), 62–83, doi:10.1016/j.jhydrol.2005.07.016.

Perrin, C., C. Michel, and V. Andréassian (2001), Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments, *Journal of Hydrology*, *242*(3–4), 275–301, doi:10.1016/S0022-1694(00)00393-0.

Perrin, C., C. Michel, and V. Andréassian (2003), Improvement of a parsimonious model for streamflow simulation, *Journal of Hydrology*, *279*(1–4), 275–289, doi:10.1016/S0022-1694(03)00225-7.

Pinson, P., H. Nielsen, J. Møller, H. Madsen, and G. Kariniotakis (2007), Non-parametric probabilistic forecasts of wind power: Required properties and evaluation, *Wind Energy*, *10*, 497 – 516, doi:10.1002/we.230.

Pushpalatha, R., C. Perrin, N. Le Moine, and V. Andréassian (2012), A review of efficiency criteria suitable for evaluating low-flow simulations, *Journal of Hydrology*, *420–421(2012)*, 171–182, doi:10.1016/j.jhydrol.2011.11.055.

Refsgaard, J. C., and H. J. Henriksen (2004), Modelling guidelines – terminology and guiding principles, *Advances in Water Resources*, *27*(1), 71–82, doi:10.1016/j.advwatres.2003.08.006.

Refsgaard, J. C., and J. Knudsen (1996), Operational Validation and Intercomparison of Different Types of Hydrological Models, *Water Resources Research*, *32*(7), 2189–2202, doi:10.1029/96WR00896.

Refsgaard, J. C., B. Storm, and A. Refsgaard (1995), Validation and applicability of distributed hydrological models, *Modelling and Management of Sustainable Basin-scale Water Resource Systems. IAHS Publ.*, *231*, 387–397.

Ritter, A., and R. Muñoz-Carpena (2013), Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments, *Journal of Hydrology*, *480*, 33 – 45, doi:10.1016/j.jhydrol.2012.12.004.

Samani, Z. A. (2000), Estimating Solar Radiation and Evapotranspiration Using Minimum Climatological Data, *Journal of Irrigation and Drainage Engineering*, *126*(4), 265–267, doi:10.1061/(ASCE)0733-9437(2000)126:4(265).

Savenije, H. H. G. (2009), HESS Opinions "The art of hydrology", *Hydrology and Earth System Sciences*, *13*, 157–161, doi:10.5194/hess-13-157-2009.

Savora, U., and B. J. Sackl (2011), Gefahrenzonen Saalach, *Technischer Bericht*, Hydroconsult GmbH, Graz, geprüft durch Bundeswasserbauverwaltung Salzburg.

Schaefli, B., and H. V. Gupta (2007), Do Nash values have value?, *Hydrological Processes*, *21*(15), 2075–2080, doi:10.1002/hyp.6825.

Schaipp, B. (2002), Saalach-Studie: Zwischenbericht der Arbeiten des Landesamtes für Wasserwirtschaft, der Regierung von Oberbayern und des Wasserwirtschaftsamtes Traunstein, *Berichte der ANL 26, 119–130*, Bayerische Akademie für Naturschutz und Landschaftspflege (ANL), ISSN 0344-6042, ISBN 3-931175-68-5.

Schlesinger, S., R. E. Crosbie, R. E. Gagné, G. S. Innis, C. S. Lalwani, J. Loch, R. J. Sylvester, R. D. Wright, N. Kheir, and D. Bartos (1979), Terminology for Model Credibility, *SIMULATION*, *32*(3), 103–104, doi:10.1177/003754977903200304.

Schoups, G., N. C. van de Giesen, and H. H. G. Savenije (2008), Model complexity control for hydrologic prediction, *Water Resources Research*, *44*(12), doi:10.1029/2008WR006836, W00B03.

Seibert, J. (1997), Estimation of Parameter Uncertainty in the HBV Model, *Nordic Hydrology*, *28*(4), 247–262.

Seibert, J. (2003), Reliability of Model Prediction Outside Calibration Conditions, *Nordic Hydrology*, *34 (5)*, 477–492.

Sevruk, B. (1983), Correction of Measured Precipitation in the Alps Using the Water Equivalent of New Snow, *Nordic Hydrology*, *14*(2), 49–58.

Sevruk, B. (1996), Adjustment of tipping-bucket precipitation gauge measurements, *Atmospheric Research*, *42*(1–4), 237–246, doi:10.1016/0169-8095(95)00066-6.

Sevruk, B., and B. Chvíla (2005), Error sources of precipitation measurements using electronic weight systems, *Atmospheric Research*, *77*(1–4), 39–47, doi:10.1016/j.atmosres.2004.10.026.

Sevruk, B., M. Ondrás, and B. Chvíla (2009), The WMO precipitation measurement intercomparisons, *Atmospheric Research*, *92*, 376–380, doi:10.1016/j.atmosres.2009.01.016.

Sikorska, A. E., A. Scheidegger, K. Banasik, and J. Rieckermann (2012), Bayesian uncertainty assessment of flood predictions in ungauged urban basins for conceptual rainfall-runoff models, *Hydrology and Earth System Sciences*, *16*(4), 1221–1236, doi:10.5194/hess-16-1221-2012.

Silberstein, R. P. (2006), Hydrological models are so good, do we still need data?, *Environmental Modelling & Software*, *21*, 1340 – 1352, doi:10.1016/j.envsoft.2005.04.019.

Singh, J., H. Vernon Knapp, J. Arnold, and M. Demissie (2005), Hydrological Modeling of the Iroquois River Watershed Using HSPF and SWAT, *Journal of the American Water Resources Association (JAWRA)*, *41*(2), 343–360, doi:10.1111/j.1752-1688.2005.tb03740.x.

Singh, S. K., and A. Bárdossy (2012), Calibration of hydrological models on hydrologically unusual events, *Advances in Water Resources*, *38*, 81–91, doi:10.1016/j.advwatres.2011.12.006.

Sorooshian, S. (Ed.) (2008), *Hydrological Modelling and the Water Cycle: Coupling the Atmospheric and Hydrological Models*, 291 pp., Springer, Berlin, ISBN 978-3-540-77842-4.

Thordarson, F. Ö., A. Breinholt, J. K. Møller, P. S. Mikkelsen, M. Grum, and H. Madsen (2012), Evaluation of probabilistic flow predictions in sewer systems using grey box models and a skill score criterion, *Stochastic Environmental Research and Risk Assessment*, *26*(8), 1151–1162, doi:10.1007/s00477-012-0563-3.

Tomkins, K. M. (2014), Uncertainty in streamflow rating curves: methods, controls and consequences, *Hydrological Processes*, *28*(3), 464–481, doi:10.1002/hyp.9567.

Tukey, J. W. (1974), Mathematics and the Picturing of Data, in *Proceedings of the International Congress of Mathematicians*, Vancouver.

Villarini, G., P. V. Mandapaka, W. F. Krajewski, and R. J. Moore (2008), Rainfall and sampling uncertainties: A rain gauge perspective, *Journal of Geophysical Research: Atmospheres*, *113*(D11), doi:10.1029/2007JD009214.

Wackernagel, H. (1998), *Multivariate Geostatistics: An Introduction with Applications*, 2nd ed., XIV, 291 pp., Springer, Berlin, ISBN 3-540-64721-X.

Wagener, T. (2003), Evaluation of catchment models, *Hydrological Processes*, *17*(16), 3375–3378, doi:10.1002/hyp.5158.

Willmott, C. J. (1981), On the validation of models, *Physical Geography*, *2*(2), 184–194, doi:10.1080/02723646.1981.10642213.

Willmott, C. J., S. G. Ackleson, R. E. Davis, J. J. Feddema, K. M. Klink, D. R. Legates, J. O'Donnell, and C. M. Rowe (1985), Statistics for the evaluation and comparison of models, *Journal of Geophysical Research: Oceans*, *90*(C5), 8995–9005, doi:10.1029/JC090iC05p08995.

Xu, C.-Y. (1999), Operational testing of a water balance model for predicting climate change impacts, *Agricultural and Forest Meteorology*, *98–99*, 295–304, doi:10.1016/S0168-1923(99)00106-9.

Yu, B., C. A. A. Ciesiolka, C. W. Rose, and K. J. Coughlan (1997), A note on sampling errors in the rainfall and runoff data collected using tipping bucket technology, *Transactions of the ASAE*, *40*, 1305–1309, doi:10.13031/2013.21388.

Institut für Wasser- und Umweltsystemmodellierung
Universität Stuttgart

Pfaffenwaldring 61
70569 Stuttgart (Vaihingen)
Telefon (0711) 685 - 60156
Telefax (0711) 685 - 51073
E-Mail: iws@iws.uni-stuttgart.de
http://www.iws.uni-stuttgart.de

Direktoren
Prof. Dr. rer. nat. Dr.-Ing. András Bárdossy
Prof. Dr.-Ing. Rainer Helmig
Prof. Dr.-Ing. Wolfgang Nowak
Prof. Dr.-Ing. Silke Wieprecht

Emeriti
Prof. Dr.-Ing. habil. Dr.-Ing. E.h. Jürgen Giesecke
Prof. Dr.h.c. Dr.-Ing. E.h. Helmut Kobus, PhD

Lehrstuhl für Wasserbau und Wassermengenwirtschaft
Leiterin: Prof. Dr.-Ing. Silke Wieprecht
Stellv.: Dr.-Ing. Kristina Terheiden
Versuchsanstalt für Wasserbau
Leiter: Stefan Haun, PhD

Lehrstuhl für Hydromechanik und Hydrosystemmodellierung
Leiter: Prof. Dr.-Ing. Rainer Helmig
Stellv.: apl. Prof. Dr.-Ing. Holger Class

Lehrstuhl für Hydrologie und Geohydrologie
Leiter: Prof. Dr. rer. nat. Dr.-Ing. András Bárdossy
Stellv.: Dr. rer. nat. Jochen Seidel
Hydrogeophysik der Vadosen Zone
(mit Forschungszentrum Jülich)
Leiter: Prof. Dr. J.A. Sander Huisman

Lehrstuhl für Stochastische Simulation und Sicherheitsforschung für Hydrosysteme
Leiter: Prof. Dr.-Ing. Wolfgang Nowak
Stellv.: apl. Prof. Dr.-Ing. Sergey Oladyshkin

***VEGAS,* Versuchseinrichtung zur Grundwasser- und Altlastensanierung**
Leiter: Dr.-Ing. Simon Kleinknecht
PD Dr.-Ing. Claus Haslauer

Verzeichnis der Mitteilungshefte

5 Plate, Erich: *Beitrag zur Bestimmung der Windgeschwindigkeitsverteilung in der durch eine Wand gestörten bodennahen Luftschicht*, und Röhnisch, Arthur; Marotz, Günter: *Neue Baustoffe und Bauausführungen für den Schutz der Böschungen und der Sohle von Kanälen, Flüssen und Häfen; Gestehungskosten und jeweilige Vorteile*, sowie Unny, T.E.: *Schwingungsuntersuchungen am Kegelstrahlschieber*, 1967

6 Seiler, Erich: *Die Ermittlung des Anlagenwertes der bundeseigenen Binnenschiffahrtsstraßen und Talsperren und des Anteils der Binnenschiffahrt an diesem Wert*, 1967

7 *Sonderheft anläßlich des 65. Geburtstages von Prof. Arthur Röhnisch mit Beiträgen von* Benk, Dieter; Breitling, J.; Gurr, Siegfried; Haberhauer, Robert; Honekamp, Hermann; Kuz, Klaus Dieter; Marotz, Günter; Mayer-Vorfelder, Hans-Jörg; Miller, Rudolf; Plate, Erich J.; Radomski, Helge; Schwarz, Helmut; Vollmer, Ernst; Wildenhahn, Eberhard; 1967

8 Jumikis, Alfred: *Beitrag zur experimentellen Untersuchung des Wassernachschubs in einem gefrierenden Boden und die Beurteilung der Ergebnisse*, 1968

9 Marotz, Günter: *Technische Grundlagen einer Wasserspeicherung im natürlichen Untergrund*, 1968

10 Radomski, Helge: *Untersuchungen über den Einfluß der Querschnittsform wellenförmiger Spundwände auf die statischen und rammtechnischen Eigenschaften*, 1968

11 Schwarz, Helmut: *Die Grenztragfähigkeit des Baugrundes bei Einwirkung vertikal gezogener Ankerplatten als zweidimensionales Bruchproblem*, 1969

12 Erbel, Klaus: *Ein Beitrag zur Untersuchung der Metamorphose von Mittelgebirgsschneedecken unter besonderer Berücksichtigung eines Verfahrens zur Bestimmung der thermischen Schneequalität*, 1969

13 Westhaus, Karl-Heinz: *Der Strukturwandel in der Binnenschiffahrt und sein Einfluß auf den Ausbau der Binnenschiffskanäle*, 1969

14 Mayer-Vorfelder, Hans-Jörg: *Ein Beitrag zur Berechnung des Erdwiderstandes unter Ansatz der logarithmischen Spirale als Gleitflächenfunktion*, 1970

15 Schulz, Manfred: *Berechnung des räumlichen Erddruckes auf die Wandung kreiszylindrischer Körper*, 1970

16 Mobasseri, Manoutschehr: *Die Rippenstützmauer. Konstruktion und Grenzen ihrer Standsicherheit*, 1970

17 Benk, Dieter: *Ein Beitrag zum Betrieb und zur Bemessung von Hochwasserrückhaltebecken*, 1970

18 Gàl, Attila: *Bestimmung der mitschwingenden Wassermasse bei überströmten Fischbauchklappen mit kreiszylindrischem Staublech*, 1971, *vergriffen*

19 Kuz, Klaus Dieter: *Ein Beitrag zur Frage des Einsetzens von Kavitationserscheinungen in einer Düsenströmung bei Berücksichtigung der im Wasser gelösten Gase*, 1971, *vergriffen*

20 Schaak, Hartmut: *Verteilleitungen von Wasserkraftanlagen*, 1971

21 *Sonderheft zur Eröffnung der neuen Versuchsanstalt des Instituts für Wasserbau der Universität Stuttgart mit Beiträgen von* Brombach, Hansjörg; Dirksen, Wolfram; Gàl, Attila; Gerlach, Reinhard; Giesecke, Jürgen; Holthoff, Franz-Josef; Kuz, Klaus Dieter; Marotz, Günter; Minor, Hans-Erwin; Petrikat, Kurt; Röhnisch, Arthur; Rueff, Helge; Schwarz, Helmut; Vollmer, Ernst; Wildenhahn, Eberhard; 1972

22 Wang, Chung-su: *Ein Beitrag zur Berechnung der Schwingungen an Kegelstrahlschiebern*, 1972

23 Mayer-Vorfelder, Hans-Jörg: *Erdwiderstandsbeiwerte nach dem Ohde-Variationsverfahren*, 1972

24 Minor, Hans-Erwin: *Beitrag zur Bestimmung der Schwingungsanfachungsfunktionen überströmter Stauklappen*, 1972, *vergriffen*

25 Brombach, Hansjörg: *Untersuchung strömungsmechanischer Elemente (Fluidik) und die Möglichkeit der Anwendung von Wirbelkammerelementen im Wasserbau*, 1972, *vergriffen*

26 Wildenhahn, Eberhard: *Beitrag zur Berechnung von Horizontalfilterbrunnen*, 1972

27 Steinlein, Helmut: *Die Eliminierung der Schwebstoffe aus Flußwasser zum Zweck der unterirdischen Wasserspeicherung, gezeigt am Beispiel der Iller*, 1972

28 Holthoff, Franz Josef: *Die Überwindung großer Hubhöhen in der Binnenschiffahrt durch Schwimmerhebewerke*, 1973

29 Röder, Karl: *Einwirkungen aus Baugrundbewegungen auf trog- und kastenförmige Konstruktionen des Wasser- und Tunnelbaues*, 1973

30 Kretschmer, Heinz: *Die Bemessung von Bogenstaumauern in Abhängigkeit von der Talform*, 1973

31 Honekamp, Hermann: *Beitrag zur Berechnung der Montage von Unterwasserpipelines*, 1973

32 Giesecke, Jürgen: *Die Wirbelkammertriode als neuartiges Steuerorgan im Wasserbau*, und Brombach, Hansjörg: *Entwicklung, Bauformen, Wirkungsweise und Steuereigenschaften von Wirbelkammerverstärkern*, 1974

33 Rueff, Helge: *Untersuchung der schwingungserregenden Kräfte an zwei hintereinander angeordneten Tiefschützen unter besonderer Berücksichtigung von Kavitation*, 1974

34 Röhnisch, Arthur: *Einpreßversuche mit Zementmörtel für Spannbeton - Vergleich der Ergebnisse von Modellversuchen mit Ausführungen in Hüllwellrohren*, 1975

35 *Sonderheft anläßlich des 65. Geburtstages von Prof. Dr.-Ing. Kurt Petrikat mit Beiträgen von*: Brombach, Hansjörg; Erbel, Klaus; Flinspach, Dieter; Fischer jr., Richard; Gàl, Attila; Gerlach, Reinhard; Giesecke, Jürgen; Haberhauer, Robert; Hafner Edzard; Hausenblas, Bernhard; Horlacher, Hans-Burkhard; Hutarew, Andreas; Knoll, Manfred; Krummet, Ralph; Marotz, Günter; Merkle, Theodor; Miller, Christoph; Minor, Hans-Erwin; Neumayer, Hans; Rao, Syamala; Rath, Paul; Rueff, Helge; Ruppert, Jürgen; Schwarz, Wolfgang; Topal-Gökceli, Mehmet; Vollmer, Ernst; Wang, Chung-su; Weber, Hans-Georg; 1975

36 Berger, Jochum: *Beitrag zur Berechnung des Spannungszustandes in rotationssymmetrisch belasteten Kugelschalen veränderlicher Wandstärke unter Gas- und Flüssigkeitsdruck durch Integration schwach singulärer Differentialgleichungen*, 1975

37 Dirksen, Wolfram: *Berechnung instationärer Abflußvorgänge in gestauten Gerinnen mittels Differenzenverfahren und die Anwendung auf Hochwasserrückhaltebecken*, 1976

38 Horlacher, Hans-Burkhard: *Berechnung instationärer Temperatur- und Wärmespannungsfelder in langen mehrschichtigen Hohlzylindern*, 1976

39 Hafner, Edzard: *Untersuchung der hydrodynamischen Kräfte auf Baukörper im Tiefwasserbereich des Meeres*, 1977, ISBN 3-921694-39-6

40 Ruppert, Jürgen: *Über den Axialwirbelkammerverstärker für den Einsatz im Wasserbau*, 1977, ISBN 3-921694-40-X

41 Hutarew, Andreas: *Beitrag zur Beeinflußbarkeit des Sauerstoffgehalts in Fließgewässern an Abstürzen und Wehren*, 1977, ISBN 3-921694-41-8, *vergriffen*

42 Miller, Christoph: *Ein Beitrag zur Bestimmung der schwingungserregenden Kräfte an unterströmten Wehren*, 1977, ISBN 3-921694-42-6

43 Schwarz, Wolfgang: *Druckstoßberechnung unter Berücksichtigung der Radial- und Längsverschiebungen der Rohrwandung*, 1978, ISBN 3-921694-43-4

44 Kinzelbach, Wolfgang: *Numerische Untersuchungen über den optimalen Einsatz variabler Kühlsysteme einer Kraftwerkskette am Beispiel Oberrhein*, 1978, ISBN 3-921694-44-2

45 Barczewski, Baldur: *Neue Meßmethoden für Wasser-Luftgemische und deren Anwendung auf zweiphasige Auftriebsstrahlen*, 1979, ISBN 3-921694-45-0

46 Neumayer, Hans: *Untersuchung der Strömungsvorgänge in radialen Wirbelkammerverstärkern*, 1979, ISBN 3-921694-46-9

47 Elalfy, Youssef-Elhassan: *Untersuchung der Strömungsvorgänge in Wirbelkammerdioden und -drosseln*, 1979, ISBN 3-921694-47-7

48 Brombach, Hansjörg: *Automatisierung der Bewirtschaftung von Wasserspeichern*, 1981, ISBN 3-921694-48-5

49 Geldner, Peter: *Deterministische und stochastische Methoden zur Bestimmung der Selbstdichtung von Gewässern*, 1981, ISBN 3-921694-49-3, *vergriffen*

50 Mehlhorn, Hans: *Temperaturveränderungen im Grundwasser durch Brauchwassereinleitungen*, 1982, ISBN 3-921694-50-7, *vergriffen*

51 Hafner, Edzard: *Rohrleitungen und Behälter im Meer*, 1983, ISBN 3-921694-51-5

52 Rinnert, Bernd: *Hydrodynamische Dispersion in porösen Medien: Einfluß von Dichteunterschieden auf die Vertikalvermischung in horizontaler Strömung*, 1983, ISBN 3-921694-52-3, *vergriffen*

53 Lindner, Wulf: *Steuerung von Grundwasserentnahmen unter Einhaltung ökologischer Kriterien*, 1983, ISBN 3-921694-53-1, *vergriffen*

54 Herr, Michael; Herzer, Jörg; Kinzelbach, Wolfgang; Kobus, Helmut; Rinnert, Bernd: *Methoden zur rechnerischen Erfassung und hydraulischen Sanierung von Grundwasserkontaminationen*, 1983, ISBN 3-921694-54-X

55 Schmitt, Paul: *Wege zur Automatisierung der Niederschlagsermittlung*, 1984, ISBN 3-921694-55-8, *vergriffen*

56 Müller, Peter: *Transport und selektive Sedimentation von Schwebstoffen bei gestautem Abfluß*, 1985, ISBN 3-921694-56-6

57 El-Qawasmeh, Fuad: *Möglichkeiten und Grenzen der Tropfbewässerung unter besonderer Berücksichtigung der Verstopfungsanfälligkeit der Tropfelemente*, 1985, ISBN 3-921694-57-4, *vergriffen*

58 Kirchenbaur, Klaus: *Mikroprozessorgesteuerte Erfassung instationärer Druckfelder am Beispiel seegangsbelasteter Baukörper*, 1985, ISBN 3-921694-58-2

59 Kobus, Helmut (Hrsg.): *Modellierung des großräumigen Wärme- und Schadstofftransports im Grundwasser*, Tätigkeitsbericht 1984/85 (DFG-Forschergruppe an den Universitäten Hohenheim, Karlsruhe und Stuttgart), 1985, ISBN 3-921694-59-0, *vergriffen*

60 Spitz, Karlheinz: *Dispersion in porösen Medien: Einfluß von Inhomogenitäten und Dichteunterschieden*, 1985, ISBN 3-921694-60-4, *vergriffen*

61 Kobus, Helmut: *An Introduction to Air-Water Flows in Hydraulics*, 1985, ISBN 3-921694-61-2

62 Kaleris, Vassilios: *Erfassung des Austausches von Oberflächen- und Grundwasser in horizontalebenen Grundwassermodellen*, 1986, ISBN 3-921694-62-0

63 Herr, Michael: *Grundlagen der hydraulischen Sanierung verunreinigter Porengrundwasserleiter*, 1987, ISBN 3-921694-63-9

64 Marx, Walter: *Berechnung von Temperatur und Spannung in Massenbeton infolge Hydratation*, 1987, ISBN 3-921694-64-7

65 Koschitzky, Hans-Peter: *Dimensionierungskonzept für Sohlbelüfter in Schußrinnen zur Vermeidung von Kavitationsschäden*, 1987, ISBN 3-921694-65-5

66 Kobus, Helmut (Hrsg.): *Modellierung des großräumigen Wärme- und Schadstofftransports im Grundwasser*, Tätigkeitsbericht 1986/87 (DFG-Forschergruppe an den Universitäten Hohenheim, Karlsruhe und Stuttgart) 1987, ISBN 3-921694-66-3

67 Söll, Thomas: *Berechnungsverfahren zur Abschätzung anthropogener Temperaturanomalien im Grundwasser*, 1988, ISBN 3-921694-67-1

68 Dittrich, Andreas; Westrich, Bernd: *Bodenseeufererosion, Bestandsaufnahme und Bewertung*, 1988, ISBN 3-921694-68-X, *vergriffen*

69 Huwe, Bernd; van der Ploeg, Rienk R.: *Modelle zur Simulation des Stickstoffhaushaltes von Standorten mit unterschiedlicher landwirtschaftlicher Nutzung*, 1988, ISBN 3-921694-69-8, *vergriffen*

70 Stephan, Karl: *Integration elliptischer Funktionen*, 1988, ISBN 3-921694-70-1

71 Kobus, Helmut; Zilliox, Lothaire (Hrsg.): *Nitratbelastung des Grundwassers, Auswirkungen der Landwirtschaft auf die Grundwasser- und Rohwasserbeschaffenheit und Maßnahmen zum Schutz des Grundwassers*. Vorträge des deutsch-französischen Kolloquiums am 6. Oktober 1988, Universitäten Stuttgart und Louis Pasteur Strasbourg (Vorträge in deutsch oder französisch, Kurzfassungen zweisprachig), 1988, ISBN 3-921694-71-X

72 Soyeaux, Renald: *Unterströmung von Stauanlagen auf klüftigem Untergrund unter Berücksichtigung laminarer und turbulenter Fließzustände*,1991, ISBN 3-921694-72-8

73 Kohane, Roberto: *Berechnungsmethoden für Hochwasserabfluß in Fließgewässern mit überströmten Vorländern*, 1991, ISBN 3-921694-73-6

74 Hassinger, Reinhard: *Beitrag zur Hydraulik und Bemessung von Blocksteinrampen in flexibler Bauweise*, 1991, ISBN 3-921694-74-4, *vergriffen*

75 Schäfer, Gerhard: *Einfluß von Schichtenstrukturen und lokalen Einlagerungen auf die Längsdispersion in Porengrundwasserleitern*, 1991, ISBN 3-921694-75-2

76 Giesecke, Jürgen: *Vorträge, Wasserwirtschaft in stark besiedelten Regionen; Umweltforschung mit Schwerpunkt Wasserwirtschaft*, 1991, ISBN 3-921694-76-0

77 Huwe, Bernd: *Deterministische und stochastische Ansätze zur Modellierung des Stickstoffhaushalts landwirtschaftlich genutzter Flächen auf unterschiedlichem Skalenniveau*, 1992, ISBN 3-921694-77-9, *vergriffen*

78 Rommel, Michael: *Verwendung von Kluftdaten zur realitätsnahen Generierung von Kluftnetzen mit anschließender laminar-turbulenter Strömungsberechnung*, 1993, ISBN 3-92 1694-78-7

79 Marschall, Paul: *Die Ermittlung lokaler Stofffrachten im Grundwasser mit Hilfe von Einbohrloch-Meßverfahren*, 1993, ISBN 3-921694-79-5, *vergriffen*

80 Ptak, Thomas: *Stofftransport in heterogenen Porenaquiferen: Felduntersuchungen und stochastische Modellierung*, 1993, ISBN 3-921694-80-9, *vergriffen*

81 Haakh, Frieder: *Transientes Strömungsverhalten in Wirbelkammern,* 1993, ISBN 3-921694-81-7

82 Kobus, Helmut; Cirpka, Olaf; Barczewski, Baldur; Koschitzky, Hans-Peter: *Versuchseinrichtung zur Grundwasser- und Altlastensanierung VEGAS, Konzeption und Programmrahmen*, 1993, ISBN 3-921694-82-5

83 Zang, Weidong: *Optimaler Echtzeit-Betrieb eines Speichers mit aktueller Abflußregenerierung*, 1994, ISBN 3-921694-83-3, *vergriffen*

84 Franke, Hans-Jörg: *Stochastische Modellierung eines flächenhaften Stoffeintrages und Transports in Grundwasser am Beispiel der Pflanzenschutzmittelproblematik*, 1995, ISBN 3-921694-84-1

85 Lang, Ulrich: *Simulation regionaler Strömungs- und Transportvorgänge in Karstaquiferen mit Hilfe des Doppelkontinuum-Ansatzes: Methodenentwicklung und Parameteridenti-fikation*, 1995, ISBN 3-921694-85-X, *vergriffen*

86 Helmig, Rainer: *Einführung in die Numerischen Methoden der Hydromechanik*, 1996, ISBN 3-921694-86-8, *vergriffen*

87 Cirpka, Olaf: *CONTRACT: A Numerical Tool for Contaminant Transport and Chemical Transformations - Theory and Program Documentation* -, 1996, ISBN 3-921694-87-6

88 Haberlandt, Uwe: *Stochastische Synthese und Regionalisierung des Niederschlages für Schmutzfrachtberechnungen*, 1996, ISBN 3-921694-88-4

89 Croisé, Jean: *Extraktion von flüchtigen Chemikalien aus natürlichen Lockergesteinen mittels erzwungener Luftströmung*, 1996, ISBN 3-921694-89-2, *vergriffen*

90 Jorde, Klaus: *Ökologisch begründete, dynamische Mindestwasserregelungen bei Ausleitungskraftwerken*, 1997, ISBN 3-921694-90-6, *vergriffen*

91 Helmig, Rainer: *Gekoppelte Strömungs- und Transportprozesse im Untergrund - Ein Beitrag zur Hydrosystemmodellierung-*, 1998, ISBN 3-921694-91-4, *vergriffen*

92 Emmert, Martin: *Numerische Modellierung nichtisothermer Gas-Wasser Systeme in porösen Medien,* 1997, ISBN 3-921694-92-2

93 Kern, Ulrich: *Transport von Schweb- und Schadstoffen in staugeregelten Fließgewässern am Beispiel des Neckars*, 1997, ISBN 3-921694-93-0, *vergriffen*

94 Förster, Georg: *Druckstoßdämpfung durch große Luftblasen in Hochpunkten von Rohrleitungen 1997*, ISBN 3-921694-94-9

95 Cirpka, Olaf: *Numerische Methoden zur Simulation des reaktiven Mehrkomponententransports im Grundwasser*, 1997, ISBN 3-921694-95-7, *vergriffen*

96 Färber, Arne: *Wärmetransport in der ungesättigten Bodenzone: Entwicklung einer thermischen In-situ-Sanierungstechnologie*, 1997, ISBN 3-921694-96-5

97 Betz, Christoph: *Wasserdampfdestillation von Schadstoffen im porösen Medium: Entwicklung einer thermischen In-situ-Sanierungstechnologie*, 1998, SBN 3-921694-97-3

98 Xu, Yichun: *Numerical Modeling of Suspended Sediment Transport in Rivers*, 1998, ISBN 3-921694-98-1, *vergriffen*

99 Wüst, Wolfgang: *Geochemische Untersuchungen zur Sanierung CKW-kontaminierter Aquifere mit Fe(0)-Reaktionswänden,* 2000, ISBN 3-933761-02-2

100 Sheta, Hussam: *Simulation von Mehrphasenvorgängen in porösen Medien unter Einbeziehung von Hysterese-Effekten*, 2000, ISBN 3-933761-03-4

101 Ayros, Edwin: *Regionalisierung extremer Abflüsse auf der Grundlage statistischer Verfahren*, 2000, ISBN 3-933761-04-2, *vergriffen*

102 Huber, Ralf: *Compositional Multiphase Flow and Transport in Heterogeneous Porous Media*, 2000, ISBN 3-933761-05-0

103 Braun, Christopherus: *Ein Upscaling-Verfahren für Mehrphasenströmungen in porösen Medien*, 2000, ISBN 3-933761-06-9

104 Hofmann, Bernd: *Entwicklung eines rechnergestützten Managementsystems zur Beurteilung von Grundwasserschadensfällen*, 2000, ISBN 3-933761-07-7

105 Class, Holger: *Theorie und numerische Modellierung nichtisothermer Mehrphasenprozesse in NAPL-kontaminierten porösen Medien,* 2001, ISBN 3-933761-08-5

106 Schmidt, Reinhard: *Wasserdampf- und Heißluftinjektion zur thermischen Sanierung kontaminierter Standorte*, 2001, ISBN 3-933761-09-3

107 Josef, Reinhold: *Schadstoffextraktion mit hydraulischen Sanierungsverfahren unter Anwendung von grenzflächenaktiven Stoffen*, 2001, ISBN 3-933761-10-7

108 Schneider, Matthias: *Habitat- und Abflussmodellierung für Fließgewässer mit unscharfen Berechnungsansätzen*, 2001, ISBN 3-933761-11-5

109 Rathgeb, Andreas: *Hydrodynamische Bemessungsgrundlagen für Lockerdeckwerke an überströmbaren Erddämmen*, 2001, ISBN 3-933761-12-3

110 Lang, Stefan: *Parallele numerische Simulation instätionärer Probleme mit adaptiven Methoden auf unstrukturierten Gittern*, 2001, ISBN 3-933761-13-1

111 Appt, Jochen; Stumpp Simone: *Die Bodensee-Messkampagne 2001, IWS/CWR Lake Constance Measurement Program 2001*, 2002, ISBN 3-933761-14-X

112 Heimerl, Stephan: *Systematische Beurteilung von Wasserkraftprojekten,* 2002, ISBN 3-933761-15-8, *vergriffen*

113 Iqbal, Amin: *On the Management and Salinity Control of Drip Irrigation,* 2002, ISBN 3-933761-16-6

114 Silberhorn-Hemminger, Annette: *Modellierung von Kluftaquifersystemen: Geostatistische Analyse und deterministisch-stochastische Kluftgenerierung,* 2002, ISBN 3-933761-17-4

115 Winkler, Angela: *Prozesse des Wärme- und Stofftransports bei der In-situ-Sanierung mit festen Wärmequellen,* 2003, ISBN 3-933761-18-2

116 Marx, Walter: *Wasserkraft, Bewässerung, Umwelt - Planungs- und Bewertungsschwerpunkte der Wasserbewirtschaftung,* 2003, ISBN 3-933761-19-0

117 Hinkelmann, Reinhard: *Efficient Numerical Methods and Information-Processing Techniques in Environment Water,* 2003, ISBN 3-933761-20-4

118 Samaniego-Eguiguren, Luis Eduardo: *Hydrological Consequences of Land Use / Land Cover and Climatic Changes in Mesoscale Catchments*, 2003, ISBN 3-933761-21-2

119 Neunhäuserer, Lina: *Diskretisierungsansätze zur Modellierung von Strömungs- und Transportprozessen in geklüftet-porösen Medien*, 2003, ISBN 3-933761-22-0

120 Paul, Maren: *Simulation of Two-Phase Flow in Heterogeneous Poros Media with Adaptive Methods*, 2003, ISBN 3-933761-23-9

121 Ehret, Uwe: *Rainfall and Flood Nowcasting in Small Catchments using Weather Radar*, 2003, ISBN 3-933761-24-7

122 Haag, Ingo: *Der Sauerstoffhaushalt staugeregelter Flüsse am Beispiel des Neckars - Analysen, Experimente, Simulationen -,* 2003, ISBN 3-933761-25-5

123 Appt, Jochen: *Analysis of Basin-Scale Internal Waves in Upper Lake Constance,* 2003, ISBN 3-933761-26-3

124 Hrsg.: Schrenk, Volker; Batereau, Katrin; Barczewski, Baldur; Weber, Karolin und Koschitzky, Hans-Peter: *Symposium Ressource Fläche und VEGAS - Statuskolloquium 2003, 30. September und 1. Oktober 2003,* 2003, ISBN 3-933761-27-1

125 Omar Khalil Ouda: *Optimisation of Agricultural Water Use: A Decision Support System for the Gaza Strip,* 2003, ISBN 3-933761-28-0

126 Batereau, Katrin: *Sensorbasierte Bodenluftmessung zur Vor-Ort-Erkundung von Schadensherden im Untergrund*, 2004, ISBN 3-933761-29-8

127 Witt, Oliver: *Erosionsstabilität von Gewässersedimenten mit Auswirkung auf den Stofftransport bei Hochwasser am Beispiel ausgewählter Stauhaltungen des Oberrheins*, 2004, ISBN 3-933761-30-1

128 Jakobs, Hartmut: *Simulation nicht-isothermer Gas-Wasser-Prozesse in komplexen Kluft-Matrix-Systemen*, 2004, ISBN 3-933761-31-X

129 Li, Chen-Chien: *Deterministisch-stochastisches Berechnungskonzept zur Beurteilung der Auswirkungen erosiver Hochwasserereignisse in Flussstauhaltungen,* 2004, ISBN 3-933761-32-8

130 Reichenberger, Volker; Helmig, Rainer; Jakobs, Hartmut; Bastian, Peter; Niessner, Jennifer: *Complex Gas-Water Processes in Discrete Fracture-Matrix Systems: Up-scaling, Mass-Conservative Discretization and Efficient Multilevel Solution,* 2004, ISBN 3-933761-33-6

131 Hrsg.: Barczewski, Baldur; Koschitzky, Hans-Peter; Weber, Karolin; Wege, Ralf: *VEGAS - Statuskolloquium 2004,* Tagungsband zur Veranstaltung am 05. Oktober 2004 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2004, ISBN 3-933761-34-4

132 Asie, Kemal Jabir: *Finite Volume Models for Multiphase Multicomponent Flow through Porous Media*. 2005, ISBN 3-933761-35-2

133 Jacoub, George: *Development of a 2-D Numerical Module for Particulate Contaminant Transport in Flood Retention Reservoirs and Impounded Rivers,* 2004, ISBN 3-933761-36-0

134 Nowak, Wolfgang: *Geostatistical Methods for the Identification of Flow and Transport Parameters in the Subsurface,* 2005, ISBN 3-933761-37-9

135 Süß, Mia: *Analysis of the influence of structures and boundaries on flow and transport processes in fractured porous media*, 2005, ISBN 3-933761-38-7

136 Jose, Surabhin Chackiath: *Experimental Investigations on Longitudinal Dispersive Mixing in Heterogeneous Aquifers,* 2005, ISBN: 3-933761-39-5

137 Filiz, Fulya: *Linking Large-Scale Meteorological Conditions to Floods in Mesoscale Catchments,* 2005, ISBN 3-933761-40-9

138 Qin, Minghao: *Wirklichkeitsnahe und recheneffiziente Ermittlung von Temperatur und Spannungen bei großen RCC-Staumauern*, 2005, ISBN 3-933761-41-7

139 Kobayashi, Kenichiro: *Optimization Methods for Multiphase Systems in the Subsurface - Application to Methane Migration in Coal Mining Areas*, 2005, ISBN 3-933761-42-5

140 Rahman, Md. Arifur: *Experimental Investigations on Transverse Dispersive Mixing in Heterogeneous Porous Media,* 2005, ISBN 3-933761-43-3

141 Schrenk, Volker: *Ökobilanzen zur Bewertung von Altlastensanierungsmaßnahmen,* 2005, ISBN 3-933761-44-1

142 Hundecha, Hirpa Yeshewatesfa: *Regionalization of Parameters of a Conceptual Rainfall-Runoff Model*, 2005, ISBN: 3-933761-45-X

143 Wege, Ralf: *Untersuchungs- und Überwachungsmethoden für die Beurteilung natürlicher Selbstreinigungsprozesse im Grundwasser,* 2005, ISBN 3-933761-46-8

144 Breiting, Thomas: *Techniken und Methoden der Hydroinformatik - Modellierung von komplexen Hydrosystemen im Untergrund*, 2006, ISBN 3-933761-47-6

145 Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Müller, Martin: *Ressource Untergrund: 10 Jahre VEGAS: Forschung und Technologieentwicklung zum Schutz von Grundwasser und Boden,* Tagungsband zur Veranstaltung am 28. und 29. September 2005 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2005, ISBN 3-933761-48-4

146 Rojanschi, Vlad: *Abflusskonzentration in mesoskaligen Einzugsgebieten unter Berücksichtigung des Sickerraumes,* 2006, ISBN 3-933761-49-2

147 Winkler, Nina Simone: *Optimierung der Steuerung von Hochwasserrückhaltebeckensystemen,* 2006, ISBN 3-933761-50-6

148 Wolf, Jens: *Räumlich differenzierte Modellierung der Grundwasserströmung alluvialer Aquifere für mesoskalige Einzugsgebiete*, 2006, ISBN: 3-933761-51-4

149 Kohler, Beate: *Externe Effekte der Laufwasserkraftnutzung*, 2006, ISBN 3-933761-52-2

150 Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Stuhrmann, Matthias: *VEGAS-Statuskolloquium 2006,* Tagungsband zur Veranstaltung am 28. September 2006 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2006, ISBN 3-933761-53-0

151 Niessner, Jennifer: *Multi-Scale Modeling of Multi-Phase - Multi-Component Processes in Heterogeneous Porous Media*, 2006, ISBN 3-933761-54-9

152 Fischer, Markus: *Beanspruchung eingeerdeter Rohrleitungen infolge Austrocknung bindiger Böden,* 2006, ISBN 3-933761-55-7

153 Schneck, Alexander: *Optimierung der Grundwasserbewirtschaftung unter Berücksichtigung der Belange der Wasserversorgung, der Landwirtschaft und des Naturschutzes*, 2006, ISBN 3-933761-56-5

154 Das, Tapash: *The Impact of Spatial Variability of Precipitation on the Predictive Uncertainty of Hydrological Models,* 2006, ISBN 3-33761-57-3

155 Bielinski, Andreas: *Numerical Simulation of CO_2 sequestration in geological formations*, 2007, ISBN 3-933761-58-1

156 Mödinger, Jens: *Entwicklung eines Bewertungs- und Entscheidungsunterstützungssystems für eine nachhaltige regionale Grundwasserbewirtschaftung*, 2006, ISBN 3-933761-60-3

157 Manthey, Sabine: *Two-phase flow processes with dynamic effects in porous media - parameter estimation and simulation,* 2007, ISBN 3-933761-61-1

158 Pozos Estrada, Oscar: *Investigation on the Effects of Entrained Air in Pipelines,* 2007, ISBN 3-933761-62-X

159 Ochs, Steffen Oliver: *Steam injection into saturated porous media – process analysis including experimental and numerical investigations*, 2007, ISBN 3-933761-63-8

160 Marx, Andreas: *Einsatz gekoppelter Modelle und Wetterradar zur Abschätzung von Niederschlagsintensitäten und zur Abflussvorhersage, 2007,* ISBN 3-933761-64-6

161 Hartmann, Gabriele Maria: *Investigation of Evapotranspiration Concepts in Hydrological Modelling for Climate Change Impact Assessment*, 2007, ISBN 3-933761-65-4

162 Kebede Gurmessa, Tesfaye: *Numerical Investigation on Flow and Transport Characteristics to Improve Long-Term Simulation of Reservoir Sedimentation*, 2007, ISBN 3-933761-66-2

163 Trifković, Aleksandar: *Multi-objective and Risk-based Modelling Methodology for Planning, Design and Operation of Water Supply Systems*, 2007, ISBN 3-933761-67-0

164 Götzinger, Jens: *Distributed Conceptual Hydrological Modelling - Simulation of Climate, Land Use Change Impact and Uncertainty Analysis*, 2007, ISBN 3-933761-68-9

165 Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Stuhrmann, Matthias: *VEGAS – Kolloquium 2007,* Tagungsband zur Veranstaltung am 26. September 2007 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2007, ISBN 3-933761-69-7

166 Freeman, Beau: *Modernization Criteria Assessment for Water Resources Planning; Klamath Irrigation Project, U.S.*, 2008, ISBN 3-933761-70-0

167 Dreher, Thomas: *Selektive Sedimentation von Feinstschwebstoffen in Wechselwirkung mit wandnahen turbulenten Strömungsbedingungen*, 2008, ISBN 3-933761-71-9

168 Yang, Wei: *Discrete-Continuous Downscaling Model for Generating Daily Precipitation Time Series*, 2008, ISBN 3-933761-72-7

169 Kopecki, Ianina: *Calculational Approach to FST-Hemispheres for Multiparametrical Benthos Habitat Modelling,* 2008*,* ISBN 3-933761-73-5

170 Brommundt, Jürgen: *Stochastische Generierung räumlich zusammenhängender Niederschlagszeitreihen*, 2008, ISBN 3-933761-74-3

171 Papafotiou, *Alexandros: Numerical Investigations of the Role of Hysteresis in Heterogeneous Two-Phase Flow Systems*, 2008, ISBN 3-933761-75-1

172 He, Yi: *Application of a Non-Parametric Classification Scheme to Catchment Hydrology*, 2008, ISBN 978-3-933761-76-7

173 Wagner, Sven: *Water Balance in a Poorly Gauged Basin in West Africa Using Atmospheric Modelling and Remote Sensing Information*, 2008, ISBN 978-3-933761-77-4

174 Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Stuhrmann, Matthias; Schrenk, Volker: *VEGAS-Kolloquium 2008 Ressource Fläche III*, Tagungsband zur Veranstaltung am 01. Oktober 2008 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2008, ISBN 978-3-933761-78-1

175 Patil, Sachin: *Regionalization of an Event Based Nash Cascade Model for Flood Predictions in Ungauged Basins*, 2008, ISBN 978-3-933761-79-8

176 Assteerawatt, Anongnart: *Flow and Transport Modelling of Fractured Aquifers based on a Geostatistical Approach,* 2008, ISBN 978-3-933761-80-4

177 Karnahl, Joachim Alexander: *2D numerische Modellierung von multifraktionalem Schwebstoff- und Schadstofftransport in Flüssen*, 2008, ISBN 978-3-933761-81-1

178 Hiester, Uwe: *Technologieentwicklung zur In-situ-Sanierung der ungesättigten Bodenzone mit festen Wärmequellen,* 2009, ISBN 978-3-933761-82-8

179 Laux, Patrick: *Statistical Modeling of Precipitation for Agricultural Planning in the Volta Basin of West Africa,* 2009, ISBN 978-3-933761-83-5

180 Ehsan, Saqib: *Evaluation of Life Safety Risks Related to Severe Flooding,* 2009, ISBN 978-3-933761-84-2

181 Prohaska, Sandra: *Development and Application of a 1D Multi-Strip Fine Sediment Transport Model for Regulated Rivers,* 2009, ISBN 978-3-933761-85-9

182 Kopp, Andreas: *Evaluation of CO_2 Injection Processes in Geological Formations for Site Screening,* 2009, ISBN 978-3-933761-86-6

183 Ebigbo, Anozie: *Modelling of biofilm growth and its influence on CO_2 and water (two-phase) flow in porous media,* 2009, ISBN 978-3-933761-87-3

184 Freiboth, Sandra: *A phenomenological model for the numerical simulation of multiphase multicomponent processes considering structural alterations of porous media*, 2009, ISBN 978-3-933761-88-0

185 Zöllner, Frank: *Implementierung und Anwendung netzfreier Methoden im Konstruktiven Wasserbau und in der Hydromechanik*, 2009, ISBN 978-3-933761-89-7

186 Vasin, Milos: *Influence of the soil structure and property contrast on flow and transport in the unsaturated zone*, 2010, ISBN 978-3-933761-90-3

187 Li, Jing: *Application of Copulas as a New Geostatistical Tool*, 2010, ISBN 978-3-933761-91-0

188 AghaKouchak, Amir: *Simulation of Remotely Sensed Rainfall Fields Using Copulas*, 2010, ISBN 978-3-933761-92-7

189 Thapa, Pawan Kumar: *Physically-based spatially distributed rainfall runoff modelling for soil erosion estimation*, 2010, ISBN 978-3-933761-93-4

190 Wurms, Sven: *Numerische Modellierung der Sedimentationsprozesse in Retentionsanlagen zur Steuerung von Stoffströmen bei extremen Hochwasserabflussereignissen*, 2011, ISBN 978-3-933761-94-1

191 Merkel, Uwe: *Unsicherheitsanalyse hydraulischer Einwirkungen auf Hochwasserschutzdeiche und Steigerung der Leistungsfähigkeit durch adaptive Strömungsmodellierung*, 2011, ISBN 978-3-933761-95-8

192 Fritz, Jochen: *A Decoupled Model for Compositional Non-Isothermal Multiphase Flow in Porous Media and Multiphysics Approaches for Two-Phase Flow*, 2010, ISBN 978-3-933761-96-5

193 Weber, Karolin (Hrsg.): *12. Treffen junger WissenschaftlerInnen an Wasserbauinstituten*, 2010, ISBN 978-3-933761-97-2

194 Bliefernicht, Jan-Geert: *Probability Forecasts of Daily Areal Precipitation for Small River Basins,* 2011, ISBN 978-3-933761-98-9

195 Hrsg.: Koschitzky, Hans-Peter; Braun, Jürgen: *VEGAS-Kolloquium 2010 In-situ-Sanierung - Stand und Entwicklung Nano und ISCO -,* Tagungsband zur Veranstaltung am 07. Oktober 2010 an der Universität Stuttgart, Campus Stuttgart-Vaihingen*,* 2010*,* ISBN 978-3-933761-99-6

196 Gafurov, Abror: *Water Balance Modeling Using Remote Sensing Information - Focus on Central Asia,* 2010, ISBN 978-3-942036-00-9

197 Mackenberg, Sylvia: *Die Quellstärke in der Sickerwasserprognose: Möglichkeiten und Grenzen von Labor- und Freilanduntersuchungen,* 2010, ISBN 978-3-942036-01-6

198 Singh, Shailesh Kumar: *Robust Parameter Estimation in Gauged and Ungauged Basins,* 2010, ISBN 978-3-942036-02-3

199 Doğan, Mehmet Onur: *Coupling of porous media flow with pipe flow,* 2011, ISBN 978-3-942036-03-0

200 Liu, Min: *Study of Topographic Effects on Hydrological Patterns and the Implication on Hydrological Modeling and Data Interpolation,* 2011, ISBN 978-3-942036-04-7

201 Geleta, Habtamu Itefa: *Watershed Sediment Yield Modeling for Data Scarce Areas,* 2011, ISBN 978-3-942036-05-4

202 Franke, Jörg: *Einfluss der Überwachung auf die Versagenswahrscheinlichkeit von Staustufen,* 2011, ISBN 978-3-942036-06-1

203 Bakimchandra, Oinam: *Integrated Fuzzy-GIS approach for assessing regional soil erosion risks,* 2011, ISBN 978-3-942036-07-8

204 Alam, Muhammad Mahboob: *Statistical Downscaling of Extremes of Precipitation in Mesoscale Catchments from Different RCMs and Their Effects on Local Hydrology,* 2011, ISBN 978-3-942036-08-5

205 Hrsg.: Koschitzky, Hans-Peter; Braun, Jürgen: *VEGAS-Kolloquium 2011 Flache Geothermie - Perspektiven und Risiken,* Tagungsband zur Veranstaltung am 06. Oktober 2011 an der Universität Stuttgart, Campus Stuttgart-Vaihingen*,* 2011*,* ISBN 978-3-933761-09-2

206 Haslauer, Claus: *Analysis of Real-World Spatial Dependence of Subsurface Hydraulic Properties Using Copulas with a Focus on Solute Transport Behaviour,* 2011, ISBN 978-3-942036-10-8

207 Dung, Nguyen Viet: *Multi-objective automatic calibration of hydrodynamic models – development of the concept and an application in the Mekong Delta,* 2011, ISBN 978-3-942036-11-5

208 Hung, Nguyen Nghia: *Sediment dynamics in the floodplain of the Mekong Delta, Vietnam,* 2011, ISBN 978-3-942036-12-2

209 Kuhlmann, Anna: *Influence of soil structure and root water uptake on flow in the unsaturated zone,* 2012, ISBN 978-3-942036-13-9

210 Tuhtan, Jeffrey Andrew: *Including the Second Law Inequality in Aquatic Ecodynamics: A Modeling Approach for Alpine Rivers Impacted by Hydropeaking,* 2012, ISBN 978-3-942036-14-6

211 Tolossa, Habtamu: *Sediment Transport Computation Using a Data-Driven Adaptive Neuro-Fuzzy Modelling Approach,* 2012, ISBN 978-3-942036-15-3

212 Tatomir, Alexandru-Bodgan: *From Discrete to Continuum Concepts of Flow in Fractured Porous Media,* 2012, ISBN 978-3-942036-16-0

213 Erbertseder, Karin: *A Multi-Scale Model for Describing Cancer-Therapeutic Transport in the Human Lung,* 2012, ISBN 978-3-942036-17-7

214 Noack, Markus: *Modelling Approach for Interstitial Sediment Dynamics and Reproduction of Gravel Spawning Fish,* 2012, ISBN 978-3-942036-18-4

215 De Boer, Cjestmir Volkert: *Transport of Nano Sized Zero Valent Iron Colloids during Injection into the Subsurface*, 2012, ISBN 978-3-942036-19-1

216 Pfaff, Thomas: *Processing and Analysis of Weather Radar Data for Use in Hydrology*, 2013, ISBN 978-3-942036-20-7

217 Lebrenz, Hans-Henning: *Addressing the Input Uncertainty for Hydrological Modeling by a New Geostatistical Method*, 2013, ISBN 978-3-942036-21-4

218 Darcis, Melanie Yvonne: *Coupling Models of Different Complexity for the Simulation of CO_2 Storage in Deep Saline Aquifers*, 2013, ISBN 978-3-942036-22-1

219 Beck, Ferdinand: *Generation of Spatially Correlated Synthetic Rainfall Time Series in High Temporal Resolution - A Data Driven Approach*, 2013, ISBN 978-3-942036-23-8

220 Guthke, Philipp: *Non-multi-Gaussian spatial structures: Process-driven natural genesis, manifestation, modeling approaches, and influences on dependent processes*, 2013, ISBN 978-3-942036-24-5

221 Walter, Lena: *Uncertainty studies and risk assessment for CO_2 storage in geological formations*, 2013, ISBN 978-3-942036-25-2

222 Wolff, Markus: *Multi-scale modeling of two-phase flow in porous media including capillary pressure effects,* 2013, ISBN 978-3-942036-26-9

223 Mosthaf, Klaus Roland: *Modeling and analysis of coupled porous-medium and free flow with application to evaporation processes,* 2014, ISBN 978-3-942036-27-6

224 Leube, Philipp Christoph: *Methods for Physically-Based Model Reduction in Time: Analysis, Comparison of Methods and Application*, 2013, ISBN 978-3-942036-28-3

225 Rodríguez Fernández, Jhan Ignacio: *High Order Interactions among environmental variables: Diagnostics and initial steps towards modeling*, 2013, ISBN 978-3-942036-29-0

226 Eder, Maria Magdalena: *Climate Sensitivity of a Large Lake*, 2013, ISBN 978-3-942036-30-6

227 Greiner, Philipp: *Alkoholinjektion zur In-situ-Sanierung von CKW Schadensherden in Grundwasserleitern: Charakterisierung der relevanten Prozesse auf unterschiedlichen Skalen*, 2014, ISBN 978-3-942036-31-3

228 Lauser, Andreas: *Theory and Numerical Applications of Compositional Multi-Phase Flow in Porous Media*, 2014, ISBN 978-3-942036-32-0

229 Enzenhöfer, Rainer: *Risk Quantification and Management in Water Production and Supply Systems*, 2014, ISBN 978-3-942036-33-7

230 Faigle, Benjamin: *Adaptive modelling of compositional multi-phase flow with capillary pressure*, 2014, ISBN 978-3-942036-34-4

231 Oladyshkin, Sergey: *Efficient modeling of environmental systems in the face of complexity and uncertainty*, 2014, ISBN 978-3-942036-35-1

232 Sugimoto, Takayuki: *Copula based Stochastic Analysis of Discharge Time Series*, 2014, ISBN 978-3-942036-36-8

233 Koch, Jonas: *Simulation, Identification and Characterization of Contaminant Source Architectures in the Subsurface*, 2014, ISBN 978-3-942036-37-5

234 Zhang, Jin: *Investigations on Urban River Regulation and Ecological Rehabilitation Measures, Case of Shenzhen in China*, 2014, ISBN 978-3-942036-38-2

235 Siebel, Rüdiger: *Experimentelle Untersuchungen zur hydrodynamischen Belastung und Standsicherheit von Deckwerken an überströmbaren Erddämmen,* 2014, ISBN 978-3-942036-39-9

236 Baber, Katherina: *Coupling free flow and flow in porous media in biological and technical applications: From a simple to a complex interface description,* 2014, ISBN 978-3-942036-40-5

237 Nuske, Klaus Philipp: *Beyond Local Equilibrium — Relaxing local equilibrium assumptions in multiphase flow in porous media,* 2014, ISBN 978-3-942036-41-2

238 Geiges, Andreas: *Efficient concepts for optimal experimental design in nonlinear environmental systems*, 2014, ISBN 978-3-942036-42-9

239 Schwenck, Nicolas: *An XFEM-Based Model for Fluid Flow in Fractured Porous Media*, 2014, ISBN 978-3-942036-43-6

240 Chamorro Chávez, Alejandro: *Stochastic and hydrological modelling for climate change prediction in the Lima region, Peru*, 2015, ISBN 978-3-942036-44-3

241 Yulizar: *Investigation of Changes in Hydro-Meteorological Time Series Using a Depth-Based Approach*, 2015, ISBN 978-3-942036-45-0

242 Kretschmer, Nicole: *Impacts of the existing water allocation scheme on the Limarí watershed – Chile, an integrative approach*, 2015, ISBN 978-3-942036-46-7

243 Kramer, Matthias: *Luftbedarf von Freistrahlturbinen im Gegendruckbetrieb*, 2015, ISBN 978-3-942036-47-4

244 Hommel, Johannes: *Modeling biogeochemical and mass transport processes in the subsurface: Investigation of microbially induced calcite precipitation*, 2016, ISBN 978-3-942036-48-1

245 Germer, Kai: *Wasserinfiltration in die ungesättigte Zone eines makroporösen Hanges und deren Einfluss auf die Hangstabilität*, 2016, ISBN 978-3-942036-49-8

246 Hörning, Sebastian: *Process-oriented modeling of spatial random fields using copulas*, 2016, ISBN 978-3-942036-50-4

247 Jambhekar, Vishal: *Numerical modeling and analysis of evaporative salinization in a coupled free-flow porous-media system,* 2016, ISBN 978-3-942036-51-1

248 Huang, Yingchun: *Study on the spatial and temporal transferability of conceptual hydrological models*, 2016, ISBN 978-3-942036-52-8

249 Kleinknecht, Simon Matthias: *Migration and retention of a heavy NAPL vapor and remediation of the unsaturated zone*, 2016, ISBN 978-3-942036-53-5

250 Kwakye, Stephen Oppong: *Study on the effects of climate change on the hydrology of the West African sub-region*, 2016, ISBN 978-3-942036-54-2

251 Kissinger, Alexander: *Basin-Scale Site Screening and Investigation of Possible Impacts of CO_2 Storage on Subsurface Hydrosystems*, 2016, ISBN 978-3-942036-55-9

252 Müller, Thomas: *Generation of a Realistic Temporal Structure of Synthetic Precipitation Time Series for Sewer Applications*, 2017, ISBN 978-3-942036-56-6

253 Grüninger, Christoph: *Numerical Coupling of Navier-Stokes and Darcy Flow for Soil-Water Evaporation*, 2017, ISBN 978-3-942036-57-3

254 Suroso: *Asymmetric Dependence Based Spatial Copula Models: Empirical Investigations and Consequences on Precipitation Fields*, 2017, ISBN 978-3-942036-58-0

255 Müller, Thomas; Mosthaf, Tobias; Gunzenhauser, Sarah; Seidel, Jochen; Bárdossy, András: *Grundlagenbericht Niederschlags-Simulator (NiedSim3)*, 2017, ISBN 978-3-942036-59-7

256 Mosthaf, Tobias: *New Concepts for Regionalizing Temporal Distributions of Precipitation and for its Application in Spatial Rainfall Simulation,* 2017, ISBN 978-3-942036-60-3

257 Fenrich, Eva Katrin: *Entwicklung eines ökologisch-ökonomischen Vernetzungsmodells für Wasserkraftanlagen und Mehrzweckspeicher,* 2018, ISBN 978-3-942036-61-0

258 Schmidt, Holger: *Microbial stabilization of lotic fine sediments*, 2018, ISBN 978-3-942036-62-7

259 Fetzer, Thomas: *Coupled Free and Porous-Medium Flow Processes Affected by Turbulence and Roughness – Models, Concepts and Analysis*, 2018, ISBN 978-3-942036-63-4

260 Schröder, Hans Christoph: *Large-scale High Head Pico Hydropower Potential Assessment*, 2018, ISBN 978-3-942036-64-1

261 Bode, Felix: *Early-Warning Monitoring Systems for Improved Drinking Water Resource Protection*, 2018, ISBN 978-3-942036-65-8

262 Gebler, Tobias: *Statistische Auswertung von simulierten Talsperrenüberwachungsdaten zur Identifikation von Schadensprozessen an Gewichtsstaumauern*, 2018, ISBN 978-3-942036-66-5

263 Harten, Matthias von: *Analyse des Zuppinger-Wasserrades – Hydraulische Optimierungen unter Berücksichtigung ökologischer Aspekte*, 2018, ISBN 978-3-942036-67-2

264 Yan, Jieru: *Nonlinear estimation of short time precipitation using weather radar and surface observations*, 2018, ISBN 978-3-942036-68-9

265 Beck, Martin: *Conceptual approaches for the analysis of coupled hydraulic and geomechanical processes*, 2019, ISBN 978-3-942036-69-6

266 Haas, Jannik: *Optimal planning of hydropower and energy storage technologies for fully renewable power systems*, 2019, ISBN 978-3-942036-70-2

267 Schneider, Martin: *Nonlinear Finite Volume Schemes for Complex Flow Processes and Challenging Grids*, 2019, ISBN 978-3-942036-71-9

268 Most, Sebastian Christopher: *Analysis and Simulation of Anomalous Transport in Porous Media*, 2019, ISBN 978-3-942036-72-6

269 Buchta, Rocco: *Entwicklung eines Ziel- und Bewertungssystems zur Schaffung nachhaltiger naturnaher Strukturen in großen sandgeprägten Flüssen des norddeutschen Tieflandes*, 2019, ISBN 978-3-942036-73-3

270 Thom, Moritz: *Towards a Better Understanding of the Biostabilization Mechanisms of Sediment Beds*, 2019, ISBN 978-3-942036-74-0

271 Stolz, Daniel: *Die Nullspannungstemperatur in Gewichtsstaumauern unter Berücksichtigung der Festigkeitsentwicklung des Betons*, 2019, ISBN 978-3-942036-75-7

272 Rodriguez Pretelin, Abelardo: *Integrating transient flow conditions into groundwater well protection*, 2020, ISBN: 978-3-942036-76-4

273 Weishaupt, Kilian: *Model Concepts for Coupling Free Flow with Porous Medium Flow at the Pore-Network Scale: From Single-Phase Flow to Compositional Non-Isothermal Two-Phase Flow*, 2020, ISBN: 978-3-942036-77-1

274 Koch, Timo: *Mixed-dimension models for flow and transport processes in porous media with embedded tubular network systems*, 2020, ISBN: 978-3-942036-78-8

275 Gläser, Dennis: *Discrete fracture modeling of multi-phase flow and deformation in fractured poroelastic media*, 2020, ISBN: 978-3-942036-79-5

276 Seitz, Lydia: *Development of new methods to apply a multi-parameter approach – A first step towards the determination of colmation*, 2020, ISBN: 978-3-942036-80-1

277 Ebrahim Bakhshipour, Amin: *Optimizing hybrid decentralized systems for sustainable ur-ban drainage infrastructures planning*, 2021, ISBN: 978-3-942036-81-8

278 Seitz, Gabriele: *Modeling Fixed-Bed Reactors for Thermochemical Heat Storage with the Reaction System CaO/Ca(OH)$_2$*, 2021, ISBN: 978-3-942036-82-5

279 Emmert, Simon: *Developing and Calibrating a Numerical Model for Microbially Enhanced Coal-Bed Methane Production*, 2021, ISBN: 978-3-942036-83-2

280 Heck, Katharina Klara: *Modelling and analysis of multicomponent transport at the interface between free- and porous-medium flow - influenced by radiation and roughness*, 2021, ISBN: 978-3-942036-84-9

281 Ackermann, Sina: *A multi-scale approach for drop/porous-medium interaction*, 2021, ISBN: 978-3-942036-85-6

282 Beckers, Felix: *Investigations on Functional Relationships between Cohesive Sediment Erosion and Sediment Characteristics*, 2021, ISBN: 978-3-942036-86-3

283 Schlabing, Dirk: *Generating Weather for Climate Impact Assessment on Lakes*, 2021, ISBN: 978-3-942036-87-0

284 Becker, Beatrix: *Efficient multiscale multiphysics models accounting for reversible flow at various subsurface energy storage sites*, 2021, ISBN: 978-3-942036-88-7

285 Reuschen, Sebastian: *Bayesian Inversion and Model Selection of Heterogeneities in Geo-statistical Subsurface Modeling*, 2021, ISBN: 978-3-942036-89-4

286 Michalkowski, Cynthia: *Modeling water transport at the interface between porous GDL and gas distributor of a PEM fuel cell cathode*, 2022, ISBN: 978-3-942036-90-0

287 Koca, Kaan: *Advanced experimental methods for investigating flow-biofilm-sediment interactions*, 2022, ISBN: 978-3-942036-91-7

288 Modiri, Ehsan: *Clustering simultaneous occurrences of extreme floods in the Neckar catchment*, 2022, ISBN: 978-3-942036-92-4

289 Mayar, Mohammad Assem: *High-resolution spatio-temporal measurements of the colmation phenomenon under laboratory conditions*, 2022, ISBN: 978-3-942036-93-1

290 Schäfer Rodrigues Silva, Aline: *Quantifying and Visualizing Model Similarities for Multi-Model Methods*, 2022, ISBN: 978-3-942036-94-8

291 Moreno Leiva, Simón: *Optimal planning of water and renewable energy systems for copper production processes with sector coupling and demand flexibility*, 2022, ISBN 978-3-942036-95-5

292 Schönau, Steffen: *Modellierung von Bodenerosion und Sedimentaustrag bei Hochwasserereignissen am Beispiel des Einzugsgebiets der Rems*, 2022, ISBN 978-3-942036-96-2

293 Glatz, Kumiko: *Upscaling of Nanoparticle Transport in Porous Media,* 2022, ISBN 978-3-942036-97-9

294 Pavía Santolamazza, Daniela: *Event-based flood estimation using a random forest algorithm for the regionalization in small catchments*, 2022, ISBN 978-3-942036-98-6

295 Haun, Stefan: *Advanced Methods for a Sustainable Sediment Management of Reservoirs*, 2022, ISBN 978-3-942036-99-3

296 Herma, Felix: *Data Processing and Model Choice for Flood Prediction*, 2022, ISBN 978-3-910293-00-7

Die Mitteilungshefte ab der Nr. 134 (Jg. 2005) stehen als pdf-Datei über die Homepage des Instituts: www.iws.uni-stuttgart.de zur Verfügung.