Dynamics of Exceptional Peak Discharges of Bavarian Rivers in a Changing Climate

Assessment through the Introduction of a Single Model initial Condition Large Ensemble of Climate Data to the Hydrological Impact Modelling Chain

Dissertation zur Erlangung des Doktorgrades
an der
Fakultät für Geowissenschaften
der
Ludwig-Maximilians-Universität München

vorgelegt von

Florian Willkofer

eingereicht am 04.09.2023, München

1st Supervisor: Prof. Dr. Ralf Ludwig, Department of Geography,
Ludwig-Maximilians-Universität, Munich

2nd Supervisor: Prof. Jean-Luc Martel (Ph.D.), Department of Civil Engineering,
École de Technologie Supérieure, Montreal, Québec, Canada

Date of disputation: 22.03.2024

Acknowledgments

During my years of study and employment at the Department of Geography at the Ludiwg-Maximilians Universität I met many new friends and colleagues who accompanied me on my path to my PhD and whom I would like to thank.

First of all, my gratitude belongs to Prof. Dr. Ralf Ludwig. I am extremely grateful for his support and dedication to offer me many positions at his working group for several projects (QBIC^3, BI-KLIM, ClimEx, and HIOS), which allowed me to fulfill a long-desired wish to work with High Performance Computers in the course of my journey to the PhD. Furthermore, I would like to thank him for his everlasting patience as well as his willingness and commitment to support me during this long endeavor.

A special thanks belongs to Raul Wood, with whom I worked many long days and nights to fulfill the many tasks required to accomplish the goals of the ClimEx project which culminated in the data and results displayed in this thesis. I'm especially grateful for his mental support during many stressful stages of my PhD.

To my friends and colleagues Ben Müller, Fabian von Trentini, David Gampe, Seppo Schmid, Thomas Weiß, Thomas Ramsauer, Magdalena Mittermeier, Benjamin Poschlod, Andrea Reimuth, Veronika Locherer, Andrea Böhnisch, Verena Huber-Garcia, and Jean-Luc Martel for their mental support, scientific counseling, and also many memorable moments at the "Feierabendhoibe" or festivities. Especially, Ben Müller, David Gampe, Andrea Reimuth, and Thomas Weiß with whom I had fruitful discussions on figures and posters or provided guidance and suggestions for coding and - most importantly - encouraged me to further pursue my PhD in times of doubt and kept me motivated along the way.

Furthermore, I would like to thank my colleagues at the Leibniz Supercomputing Centre (LRZ), Jens Weismüller, Gilbert Brietzke, André Kurzmann, Anton Frank, and Wolfgang Kurz who provided technical guidance and tools to enable and facilitate the application of the hydrological model on the LRZ's supercomputing systems.

Moreover, my gratitude belongs to my Québec colleagues, Martin Leduc, Richard Turcotte, Anne Frigon, Simon Ricard, and François Brissette for their view and input on the topic of climate and hydrological modelling during the fruitful ClimEx project meetings and webinars.

I would also like to thank the colleagues at the Bavarian Environmental Agency (LfU) and the Bavarian State Ministry for Customer Protection and the Environment (StMUV), Holger Komische, Frank Bäse, and Michael Joneck who supported and promoted the research proposals which lead to the above mentioned projects and consequently to the opportunity for this thesis.

I also thank Jörg Schulla for his support for the hydrological model used in this thesis.

At last, I'm very grateful for the enduring support of my family and friends, especially from my mother and Alexander Schaub who always pushed me to complete this endeavor.

Summary

Over the last 30 years heavy precipitation events, which were either locally bound or spacious in extent have caused several extreme floods in Bavaria, such as the Pentecost flood in 1999 affecting the entire Danube region, followed by floods in 2002, 2005, and 2013. All these floods have been described as events which statistically should only occur once in a hundred years (100-year flood) or longer. The time span between these high flow events, separated by only a few years, indicates that these severe events have become more frequent. In populated areas these events cause severe damage and often involve human casualties leading to an increased attention from the general public and science. According to their definition, extreme events such as the 100-year flood occur rarely; thus, they are only sparsely covered in discharge observations. The 100-year flood is frequently used in Bavaria and elsewhere as a design criterion for the development and construction of flood protection measures or hydro-power facilities. Hence, a reliable estimation of its current value and future dynamics due to a changing climate is important. These critical thresholds are generally derived from the existing discharge time series by applying various methods of extreme value statistics. However, since most available discharge time series are too short for an empirical estimation, the statistical methods need to extrapolate beyond the observed record to estimate the 100-year flood magnitude. For a more reliable quantification of the 100-year flood magnitude and in order to account for any changes in response to climate change, new approaches employ hydrological models to create long time series of discharge data based on meteorological inputs provided by weather generators or large climate model ensembles.

This dissertation investigates the impact of climate change on extreme flood events for all major Bavarian catchments. A hydro-meteorological model chain for the assessment of climate change impacts on the hydrology (scenario - global climate model (GCM) - regional climate model (RCM) - hydrological model) is employed. The model chain uses a *Single-Model Initial Condition Large Ensemble* (*SMILE*) of the *Canadian Regional Climate Model*, version 5 (*CRCM5-LE*) forced by the *Representative Concentration Pathway 8.5* (*RCP8.5*) emission scenario, to drive the hydrological model *WaSiM* (*Water Balance Simulation Model*, formerly WaSiM-ETH). Three scientific publications address different aspects of this model chain regarding its application to simulate high flow events and

their change in dynamics (i.e., frequency and intensity) in response to a changing climate.

The first publication addresses the development of the hydrological model itself. The process-based and fully-distributed model *WaSiM* was set up for 98 catchments of the Bavarian Danube and Main, as well as their tributaries (such as the Inn). Since some of these catchments extend beyond the political borders of Bavaria, the entirety of them all is further referred to as the Hydrological Bavaria. To account for the spatial and temporal dynamics of hydrological extreme events, the model was set up in a high spatio-temporal resolution. Furthermore, regionalized model parameters were determined using a semi-global and semi-automatized approach, focusing on the representation of high flows. To determine the model's performance to simulate these events a confidence value (*Level of Trust, LOT*) was introduced which shows the deviation of discharge values of selected return periods (1 in 5-, 10-, 20-years) between model data and observations at the respective gauge. The discharge values were estimated using an extreme value distribution (*Generalized Pareto Distribution* with *Peak over Threshold* sampling and *L-Moments* for parameter estimation). The results show that the model performs sufficiently well with values of the Nash & Sutcliffe Efficiency and Kling-Gupta Efficiency above 0.6 for most of the gauges. The results regarding the LOT, which represents the capability of the model to reproduce high return period events, depict moderate (between 20% and 30% deviation) to very high (less than 10% deviation) confidence for the majority gauges. However, the number of gauges yielding trustworthy results (above moderate LOT) reduces with an increasing return period.

RCM data often exhibit systematic deviations from long term mean observations (bias) which should be removed for climate change impact studies. Hence, the second publication investigated which of the selected methods for bias correction (BC; *linear scaling*, *local intensity scaling*, *quantile-mapping*, *qm*; yearly and monthly correction factors) is best suited for the adaptation of raw RCM data (by means of their impact on different hydrological indicators) and how these methods affect the climate change signal of different hydrological indicators for a selection of catchments within the Hydrological Bavaria. Although a BC is inevitable in many cases due to a strong bias in precipitation (amounts, seasonal course) and/or temperature, its application is often critically discussed as most approaches result in incoherence between variables and may alter the original climate change signal. As shown in the second publication, the qm approach with monthly correction factors is recommended for the adjustment of RCM outputs for the catchments of the Hydrological Bavaria as it yields either the best adjustment to simulations using observations or performs similarly well than other methods. Further, the presented results of this study illustrate that the employed BC methods affect the change signal of the presented hydrological indicators. Change signals for extreme event indicators are more affected by different BC methods, with more 100% difference in absolute change values in extreme cases, than those of long term mean flow indicators, with differences in relative CCS between 0 and 15 percent points.

The third publication focuses on the main scope of this dissertation: the impact of climate change on the dynamics (i.e., frequency and intensity) of extreme flood events for the 98 catchments of the Hydrological Bavaria focusing on the 100-year flood. For this purpose, the model introduced in the first publication is driven by the CRCM5-LE climate simulations which have been corrected using the qm approach as recommended in the second publication but adapted for daily correction factors. The resulting *hydrological SMILE* (*hydro-SMILE*) provides a large database of 1,500 model years per 30-year period (50 members x 30 years) for the analysis of extreme events. A comparison between values for the 100-year flood obtained by a *Generalized Extreme Value distribution* (*GEV*) and the empirical probability of exceedance is made to illustrate the benefit of the hydro-SMILE for a robust estimation of extreme flood events. The robust estimation of 100-year flood events further allows for the assessment of possible changes in the frequency and intensity by direct comparison between present and future values. The presented results show the benefit of the hydro-SMILE for the robust estimation of extreme high flow events using empirical probabilities compared to statistical estimates using an extreme value distribution. Furthermore, the results show that for catchments exhibiting a nival (snow) component in their runoff regime a considerable to severe increase in frequency (between 7 and 12 times as frequent) and intensity (between 36% and 104%) of 100-year flood events is expected until the end of the century. In catchments exhibiting a more pluvial influence in their flow regime (especially north of the Alps) these dynamics are less pronounced (at least 10% to 25% increase in intensity for more than 50% of the gauges, up to a maximum between 20% and 44%; at least 1.5 times as frequent for more than 50% of the gauges, up to 3 times as frequent at the maximum) or in individual cases even show a decline in frequency and intensity. Other studies also show this behavior in the dynamics of extreme floods for the upper Danube. However, the methods employed in this dissertation allow for a better quantification of a dynamically changing hydrological system under a transient changing climate.

This dissertation illustrates the results of a state-of-the-art modelling chain employing a single RCM large ensemble driving a single hydrological model under a strong emission scenario to study the changes in dynamics of high flow events in the Hydrological Bavaria.

Zusammenfassung

In den letzten 30 Jahren kam es in bayerischen Flusseinzugsgebieten vermehrt zu extremen, teils lokal begrenzten oder weiträumigen Niederschlagsreignissen die zu extremen Hochwässern in den betroffenen Regionen führten, etwa das Pfingsthochwasser 1999 im gesamten Donaugebiet, gefolgt von weiteren Hochwässern in den Jahren 2002, 2005 und 2013. Der zeitliche Abstand dieser Hochwasserereignisse, die nur wenige Jahre trennen, deutet auf eine Zunahme der Häufigkeit dieser schwerwiegenden Ereignisse hin. Da in besiedelten Räumen diese Ereignisse schwere Schäden verursachen und oft auch Menschenleben fordern, erlangen sie erhöhte Aufmerksamkeit in der Öffentlichkeit und Wissenschaft. Extremereignisse wie das 100-jährliche Hochwasser treten ihrer Definition gemäß nur sehr selten auf und sind somit auch in Pegelzeitreihen nur selten zu beobachten. Das 100-jährliche Hochwasser dient allerdings meist als Kriterium für die Entwicklung und Errichtung von Hochwasserschutzmaßnahmen oder Wasserkraftanlagen. Somit ist eine verlässliche Ermittlung, sowie die mögliche Entwicklung dieses Wertes in Zeiten des Klimawandels von großer Bedeutung. Diese kritischen Grenzwerte werden allgemein anhand von existierenden Abflusszeitreihen und verschiedener Methoden der Extremwertstatistik abgeleitet. Da die meisten verfügbaren Zeitreihen für eine robuste empirische Ableitung zu kurz sind, wird die Magnitude des 100-jährlichen Hochwassers durch Extrapolation der statistischen Methoden über die Beobachtungszeitreihe hinaus geschätzt. Um eine verlässlichere Quantifizierung des 100-jährlichen Hochwassers zu ermöglichen und durch den Klimawandel hervorgerufene Änderungen zu berücksichtigen, bedienen sich neue Ansätze der Verwendung hydrologischer Modelle zur Erzeugung langer Abflusszeitreihen basierend auf meteorologischen Daten aus Wettergeneratoren oder großer Klimamodellensembles.

Die vorliegende Dissertation befasst sich mit den Auswirkungen des Klimawandels auf extreme Hochwasserereignisse in Bayerischen Flusseinzugsgebieten. Hierfür wird die hydro-meteorologische Modellkette zur Ermittlung der Auswirkungen des Klimawandels auf die Hydrologie (Szenario - globales Klimamodell (GCM) - regionales Klimamodell (RCM) - hydrologisches Modell) herangezogen. Die Modellkette verwendet ein *Single-Model Initial Condition Large Ensemble* (*SMILE*) eines RCM, das *Canadian Regional Climate Model*, Version 5 Large Ensemble (*CRCM5-LE*) unter Verwendung des *Repre-*

sentative Concentration Pathway 8.5 (*RCP8.5*) Emissionsszenario als Antrieb für das hydrologische Modell *WaSiM* (*Water Balance Simulation Model*, ehemals WaSiM-ETH). In drei wissenschaftlichen Publikationen werden verschiedene Aspekte dieser Modellkette im Hinblick auf ihre Anwendung zur Simulation von Hochwasserereignissen und der durch den Klimawandel hervorgerufenen Dynamik (Intensität und Häufigkeit) dieser Ereignisse untersucht.

Die erste Publikation befasst sich mit der Erstellung des hydrologischen Modells. Hierfür wurde das prozessbasierte, flächenhaft differenziert arbeitende Modell *WaSiM* für 98 Einzugsgebiete der bayerischen Donau und des Main, sowie deren Zuflüsse (z.B. Inn) verwendet. Da einige dieser Einzugsgebiete über das politische Bayern hinaus reichen, wird ihre Gesamtheit im Folgenden als Hydrologisches Bayern bezeichnet. Um der räumlich-zeitlichen Dynamik hydrologischer Extremereignisse gerecht zu werden, wurde für das Modell eine hohe räumliche und zeitliche Auflösung gewählt. Zudem wurden regional einheiltiche Modellparameter in einem semi-globalen und semi-automatisierten Verfahren ermittelt, mit dem Fokus auf die Abbildung von Hochwasserereignissen. Zur Bewertung der Performanz hinsichtlich der Simulation dieser Ereignisse wurde ein Vertrauenswert (*Level of Trust, LOT*) eingeführt, der die Abweichung für Abflusswerte ausgewählter Hochwasserjährlichkeiten (5-, 10-, 20-jährlich) aus Modell und Beobachtungen am jeweiligen Pegel angibt. Die Abflusswerte wurden dabei durch eine Extremwertverteilung (*Generalized Pareto Distribution*; *Peak over Threshold* sampling; *L-Moments* zur Parameterschätzung) ermittelt. Die Ergebnisse zeigen, dass die Performanz des Modells, die an den meisten betrachteten Pegeln eine Nash & Sutcliffe Effizienz sowie eine Kling-Gupta Effizients über 0.6 erreicht, zufriedenstellend ist. Die Ergebnisse zum LOT, welche die Fähigkeit des Modells zur Abbildung von Hochwässern hoher Jährlichkeiten widerspiegelt, zeigen für die Mehrheit der Pegel einen moderaten (zwischen 20% und 30 % Abweichung) bis sehr hohen (unter 10% Abweichung) Vertrauenswert. Die Anzahl vertrauenswürdiger Pegel reduziert sich allerdings mit einer ansteigenden Jährlichkeit.

RCM Daten weisen häufig systematische Abweichungen vom langjährigen Mittel der Beobachtungen auf (Bias), die für Klimwandelfolgestudien entfernt werden müssen. Daher wurde in der zweiten Publikation für ausgewählte Einzugsgebiete des hydrologischen Bayerns untersucht, welche der ausgewählten Methoden zur Bias-Korrektur (BC; *linear scaling*; *local intensity scaling*; *quantile-mapping*, *qm*; monatliche und jährliche Korrekturfaktoren) am besten für eine Anpassung roher RCM Daten geeignet ist (anhand ihres Einflusses auf unterschiedliche hydrologische Indikatoren) und welche Auswirkungen diese Methoden auf das Änderungssignal verschiedener hydrologischer Indikatoren haben. Obwohl eine Bias-Korrektur in vielen Fällen aufgrund starker Abweichungen im Niederschlag (Menge, saisonaler Verlauf) und/oder in der Temperatur unumgänglich ist, ist sie aufgrund möglicher Inkohärenz zwischen den Variablen und möglichen Auswirkungen auf das ursprüngliche Klimawandelsignal umstritten. Wie in der zweiten Publikation gezeigt, wird der qm Ansatz mit monatlichen Korrekturfaktoren für eine Anpassung der RCM Ergebnisse für die Einzugsgebiete des Hydrologischen Bayerns empfohlen, da diese Methode entweder zur besten Anpassung an die Simulationen angetrieben durch beobachtete Werte führt, oder verglichen mit anderen Methoden vergleichbare Resultate erzielt. Weiter zeigen die Ergebnisse der Studie, dass die verwendeten BC Methoden das Änderungssignal der gezeigten hydrologischen Indikatoren beeinflusst. Dabei sind Signale von Extremereignissen mit Abweichungen im absoluten Wert von teils mehr als 100% in extremen Fällen stärker betroffen als die Signale langjähriger Mittel, die Änderungen im Signal zwischen 0 und 15 Prozentpunkten ausweisen.

Die dritte Publikation befasst sich mit Kernthema der Dissertation: den Auswirkungen des Klimawandels auf die Dynamik (Intensität und Häufigkeit) extremer Hochwasserereignisse in den 98 Einzugsgebieten des Hydrologischen Bayerns mit Fokus auf das 100-jährliche Hochwasser. Für diese Untersuchung wird das in der ersten Publikation beschriebene Modell durch Daten des CRCM5-LE angetrieben, die mittels der in der zweiten Publikation empfohlenen qm Methode korrigiert wurden, allerdings angepasst für tägliche Korrekturfaktoren. Das dadurch entstehende *hydrologische SMILE* (*hydro-SMILE*) bietet eine umfassende Datengrundlage von 1.500 Modelljahren pro 30-Jahres Zeitraum (50 Member x 30 Jahre) für die Analyse von Extremereignissen. Ein Vergleich zwischen den Werten für das 100-jährliche Hochwasser, die anhand der *Generalized Extreme Value distribution* (*GEV*) und der empirischen Überschreitungswahrscheinlichkeit ermittelt wurden, wurde durchgeführt, um den Vorteil eines hydro-SMILE für eine robuste Schätzung dieser Extremereignisse zu veranschaulichen. Anhand dieser robusten Werte für das 100-jährliche Hochwasser kann eine mögliche Änderung in Häufigkeit und Intensität dieser Ereignisse durch einen direkten Vergleich vergangener und zukünftiger Werte ermittelt werden. Die Ergebnisse zeigen den Vorteil des hydro-SMILE für eine robuste Abschützung extremer Hochwasserereignisse unter Verwendung empirischer Wahrscheinlichkeiten im Vergleich zur statistischen Schätzung durch eine Extremwertverteilung. Weiterhin zeigen die Ergebnisse, dass in Einzugsgebieten mit nival beeinflusstem Abflussregime eine deutliche bis starke Zunahme von Häufigkeit (zwischen 7 bis 12 mal häufiger) und Intensität (zwischen 36% und 104%) des 100-jährlichen Ereignisses gegen Ende des Jahrhunderts zu erwarten ist. In Einzugsgebieten mit zunehmend pluvial beeinflusstem Regime (vor allem nördlich der Alpen) ist diese Entwicklung weniger stark ausgeprägt (eine Zunahme der Intensität von mindestens 10% bis 25% für mehr als 50% der Pegel; bis zu einem Maximum zwischen 20% und 44%; mindestens 1,5 mal bis zu maximal 3 mal häufiger) oder zeigt in Einzelfällen sogar eine Abnahme der Häufigkeit und Intensität. Vergleichbare Studien zeigen ebenfalls dieses Verhalten für das Gebiet der oberen Donau. Allerdings erlauben die in dieser Dissertation verwendeten Methoden eine bessere Quantifikation eines sich dynamisch verändernden hydrologischen Systems unter transienten Klimawandelbedingungen.

Diese Dissertation zeigt die Resultate einer aktuellen Modellkette, die ein einzelnes RCM Large Ensemble mit einem starken Emissionsszenario als Antrieb für ein hydrologisches Modell verwendet, um Änderungen in der Dynamik von Hochwasserereignissen innerhalb des Hydrologischen Bayerns zu untersuchen.

Contents

List of Figures

List of Tables

List of Acronyms

AM	Annual maximum
AR	Assessment Report
BC	bias correction
CC	Climate Change
CCI	Climate Change Impact
CCS	Climate Change Signal
CMIP	Coupled Model Intercomparison Project
CRCM5	Canadian Regional Climate Model, version 5
ECDF	Empirical cumulative distribution function
ESM	Earth System Model
EVA	Extreme value analysis
EVD	Extreme value distribution
FFA	Flood frequency analysis
GCM	Global Circulation (or Climate) Model
GEV	Generalized Extreme Value distribution
GHG	green house gas
GPD	Generalized Pareto Distribution
HF	High flow
HFx	High flow of return period x
HPC	High Performance Computing
IPCC	Intergovernmental Panel on Climate Change
KGE	Kling-Gupta Efficiency
LE	Large Ensemble
LOT	Level of Trust
NSE, logNSE	Nash & Sutcliffe Efficiency and its logarithmic form
OF	Objective function
PDF	Probability distribution function
POT	Peak over threshold
qm	quantile-mapping
RCM	Regional Climate Model
RCP	Representative Concentration Pathways
SDS	statistical downscaling
SMILE	Single Model Initial Condition Large Ensemble
SRES	Second Report on Emission Scenarios

1 Introduction

1.1 Hydrological extreme events in Bavarian catchments

The landscapes of Bavaria are characterized by a vast river network of major rivers such as the Danube and Main river as well as their important tributaries of minor order such as the Lech, Iller, Isar, and Inn south of the Danube, and the Regnitz south of the Main river. Since not all of these streams spring within the political Bavaria but in adjacent German states or countries, the hydrological catchments span a larger area than the political Bavaria. Therefore, the total area of all catchments of rivers flowing through Bavaria is referred to as 'Hydrological Bavaria' and covers roughly 100,000km^2.

In the last three decades severe hydrological disasters frequently occurred in various scales within Bavarian catchments. Table 1 shows two types of hydrological extreme events, floods and droughts. Throughout this thesis, the term 'hydrological extremes' and related synonyms further refer to flood events only. In populated areas these events impose harm to the economy, infrastructure, food supply, biodiversity, livestock, and human life (Brunner et al., 2021a). Thus, in the immediate aftermath of these events, governmental institutions, the general public, and different science disciplines put their focus on their detailed analysis and reporting (McCollum & Beighley, 2019). On one hand, these short term analysis and reports target to answer the questions about the circumstances which caused such an event (e.g., Blöschl et al., 2013), what the economic and ecological damage was (e.g., Thieken et al., 2016), and whether these events are already influenced by climate change (attribution research, e.g., Kreienkamp et al., 2021; Szymczak et al., 2022). Furthermore, the return period of the event is usually estimated based on the existing record of observations. However, this estimate is prone to statistical and measurement uncertainties due to mostly short time series of observed values and flaws in the derivative of discharge values. Further uncertainties arise from the natural variability of peak discharges. Depending on the period selected from the time series of observations the estimate of the discharge of high return periods may vary substantially (Schulz & Bernhardt, 2016). Additionally, even if one opts to derive the estimate from the entire available time series, this estimate is only representative for this period. On the other hand, long term research seeks to clarify whether and how climate change affects the development of extreme hydrological events in terms of its frequency and intensity (e.g., van der Wiel et al., 2019; van Kempen et al., 2021).

Table 1: Severe hydrological extreme events in Bavaria of the last three decades.

Year	Extreme Type	Months	Affected Area	Source
2019	Drought	Summer	Bavaria	Bayerisches Landesamt für Umwelt (2020)
2018	Drought	Summer	Bavaria	Bayerisches Landesamt für Umwelt (2020)
2016	Flood	May/June	Bavaria	Bayerisches Landesamt für Umwelt (2017a, 2017b)
2013	Flood	June	Bavaria	Bayerisches Landesamt für Umwelt (2016)
2011	Drought	Spring & late fall	Bavaria	Kohn et al. (2014)
2007	Drought	Summer	Bavaria	Bayerisches Landesamt für Umwelt (2007b)
2005	Flood	August	Danube, Isar, Inn	Bayerisches Landesamt für Umwelt (2007a)
2003	Drought	Summer	Bavaria	Bayerisches Landesamt für Umwelt (2016)
2002	Flood	August	Danube, Iller, Lech, Isar, Inn	Gewässerkundlicher Dienst Bayern (2002)
1999	Flood	May	Danube, Iller, Lech, Isar	Bayerisches Landesamt für Wasserwirtschaft (2003)

The Intergovernmental Panel on Climate Change (IPCC) regularly publishes reports comprising the current state of scientific results regarding the effects of climate change on the economy, ecology, and biodiversity, as well as providing the physical basis for these changes on continental and regional scale. While previous reports mainly focused on changes in mean conditions the 6^{th} Assessment Report (AR6, Bednar-Friedl et al., 2022) addresses the effects of climate change on extreme events as well. The AR6 which is based on a large number of scientific publications indicates a significant increase in frequency and intensity of extreme precipitation events until the end of this century over almost the entire globe (Bednar-Friedl et al., 2022). Depending on the underlying emission scenario, projections of future climate imply temperatures to climb at higher or lower rates towards the end of the century. According to the Clausius-Clapeyron scaling, which describes the relation between temperature and moisture content of the atmosphere, the capacity of an air column available to accumulate moisture increases by close to 7% for each degree of global mean surface temperature warming (O'Gorman & Schneider, 2009; Vergara–Temprado et al., 2021). Furthermore, associated increases in extreme precipitation values scale globally close to this ratio (E. M. Fischer & Knutti, 2016; Rajczak & Schär, 2017; Vergara–Temprado et al., 2021; Westra et al., 2013; Wood et al., 2021). Therefore, as the IPCC's worst case scenario projects global average temperatures being more than 4°C higher by the end of the century compared to the period of 1850 to 1900 (IPCC, 2021), the atmosphere's potential to accumulate moisture is increased. According to Vergara–Temprado et al. (2021), in Europe the rate at which extreme precipitation events intensify in volume in the future is close to the Clausius-Clapeyron rate and can locally even exceed it (Wood & Ludwig, 2020).

The Hydrological Bavaria is situated within Central Europe. For this region projections from climate models employing a strong radiative forcing depict that extreme precipitation events are likely to occur more frequently and yield higher intensities (Martel et al., 2020; Poschlod et al., 2021). This development affects various spatio-temporal scales, from small scale short lived convective events to large scale long lasting stationary weather patterns. For the Hydrological Bavaria the Vb weather pattern is responsible for severe floods across multiple catchments and projected to form more often under ever increasing emissions in the future (Mittermeier et al., 2019; Stahl & Hofstätter, 2018).

How these changes in extreme precipitation events might translate into a hydrological response for the Hydrological Bavaria is yet uncertain as existing projects and related studies put their focus on mean conditions (QBIC[3], GLOWA-DANUBE: www.glowa-danube.de). Thus, modelling chains for the assessment of climate change impacts (CCI) on mean catchment hydrology rely on a multi-model ensemble comprised by a limited number of model simulations (members) of a combination of global and regional climate models (GCM, RCM) as well as different hydrological models working on rather coarse spatio-temporal scales. Therefore, this approach is less applicable to assess changes in hydrological extremes due to its narrow database available for the estimation of discharges of high return periods. Hence, as for the short observational records, the available data from these model experiments is prone to statistical uncertainties and natural variability (Aalbers et al., 2018).

However, a method to robustly estimate the discharge of high return periods (e.g., the 100-year flood) is required since they are used as a design criterion for flood protection or hydro-power infrastructure (Brunner et al., 2021a). Since the base for this design criterion - i.e., peak discharges - may change in the future, it is not stationary in time. Thus, robust projections of how these extreme values are likely to change have to be taken into account for the design of new or adaptation of existing structures.

To robustly investigate these projections of changes in extreme values, the database has to be enlarged. Advances in available computational resources (power and storage) allowed for the introduction of single model initial condition large ensembles (SMILEs) of GCMs and RCMs (Maher et al., 2021). Each SMILE comprises several members of a single climate model. Each member uses a different initial condition but the same radiative forcing (Maher et al., 2021). GCM SMILEs (often referred to as large ensembles, LE) have been developed and used for global climate science within the last two decades (e.g., Zelle et al., 2005). Recently, RCM SMILEs are more frequently used in regional and local climate change studies (Maher et al., 2021). Both, GCM and RCM SMILEs provide a profound database of n members times m model years to foster the analysis of natural (internal) variability and forced response (change signal), as well as extreme values (Maher et al., 2021; van der Wiel et al., 2019). The natural variability of the climate system, which originates from the complex interaction of different processes in the atmosphere, land, ocean, and cryosphere, creates an intrinsic source of uncertainty. Separating the forced response from internal variability is challenged by conventional ensembles of climate models differing in model structure and forcings (Aalbers et al., 2018). Furthermore, these conventional ensembles offer approximately the same sample size as observations (only 30 years per period) for extreme value analysis. Hence, the large database of a SMILE allows for the implicit quantification of the natural variability (Aalbers et al., 2018; Maher et al., 2019) and enlarges the sample size for the estimation of extremes (Haugen et al., 2018; Suarez-Gutierrez et al., 2018). Thereby, the difference among members only originates from internal variability due to changes in initial conditions. In contrast to uncertainties originating from the scenario and the model, the uncertainty induced by natural variability cannot be reduced (Lehner et al., 2023).

Climate change impact studies further benefit from the application of SMILEs when applied to subsequent models. Thus, by applying SMILEs to hydrological models, the resulting hydrological SMILE or hydro-SMILE may be exploited to robustly assess alterations in extreme peak discharges and respective return periods (van der Wiel et al., 2019), or to robustly assess changes in discharge regimes (Poschlod et al., 2020).

Hence, the ClimEx project (Climate Change and Extreme Events - Risks and Perspectives for Bavarian Water Management, funded by the Bavarian State Ministry for Consumer Protection and the Environment, StMUV) was found under the leadership of the Ludwig-Maximilians-Universität München to assess the dynamics of extreme hydro-meteorological events under changing climate conditions. In its first phase the project's focus was on extreme precipitation events and the response of a catchment in the form of floods. However, only major river floods were addressed in this research, excluding flash floods.

The scope of this thesis is to provide a new and robust approach to assess climate change impacts (CCI) on the dynamics of hydrological extreme events within the catchments of the entire Hydrological Bavaria. Thus, it puts the focus on certain steps of the hydrological impact modelling chain emphasizing hydrological extremes through an adapted hydrological model towards a good representation of peak flows of high return periods (publication I, chapter 2.1), an optimal method to adjust a potential bias (systematic deviation; publication II, chapter 2.2), and finally, through the application of a climatological single model initial condition large ensemble (SMILE) - the Canadian Regional Climate Model (version 5) Large Ensemble (CRCM5-LE, Leduc et al., 2019) - to drive the developed hydrological model to establish a robust estimation of the dynamics of extremes through the provision of a profound database for extreme value analysis (publication III, chapter 2.3).

1.2 Scientific foundation: hydrological impact modelling

1.2.1 The hydro-meteorological modelling chain

Modelling impacts of climate change on the hydrological system relies on a sequence of different models. The number of steps along this sequence may vary with the topic and scale of a CCI assessment study. Impact studies on climatological variables only require data from GCMs or RCMs to determine changes on global or regional scale, respectively. Subsequent studies on effects of climate change on the environment apply data from GCMs and RCMs as driver for impact models (e.g., hydrological models), also on global or regional scale, respectively.

To robustly assess the hydrological response to changes in meteorological drivers such as precipitation or temperature on a regional scale a complex hydro-meteorological modelling chain as illustrated in Figure 1 is required (Muerth et al., 2013; Teutschbein &

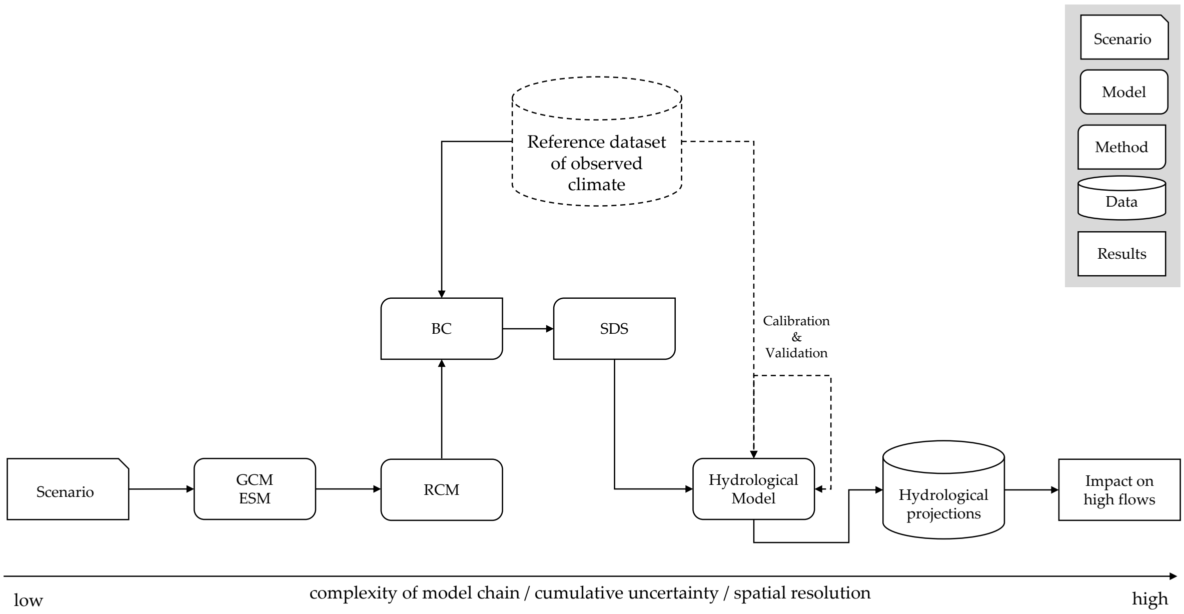


Figure 1: Hydrological climate change impact modelling chain comprised by applied change scenarios, global climate or earth system models (GCM, ESM), regional climate models (RCM), and the hydrological impact model yielding future projections on a catchment's hydrology. Data from the RCMs are processed in several steps (bias correction (BC), spatial statistical downscaling (SDS)) to meet the requirements of the hydrological model.

Seibert, 2010). It comprises model data from GCMs, the employment of a RCM to dynamically downscale the GCM data to a higher spatial resolution, correcting the RCM results to the regional climate conditions if deviations form observations are too large, further statistical spatial downscaling of the (adjusted) RCM data to match the spatial scale of the impact model using mass conserving approaches, setting up a hydrological model to best represent observations (e.g., discharge, spatial snow patterns, soil moisture), and finally, using the spatially downscaled (and adjusted) RCM data to drive the hydrological model for a reference and future period.

Along this sequence of models and processing steps the chain becomes more complex and the spatial resolution increases to allow for a more detailed analysis of processes on ever smaller scales. Since every model is prone to different sources of uncertainties (e.g. model uncertainty, statistical uncertainty, uncertainty from internal climate variability, and scenario uncertainty), each subsequent model adds to the overall uncertainty despite the increase on spatial accuracy. The contribution of each part of the modelling chain to the overall uncertainty has been subject of several studies. In general, the different GCMs, their members, and subsequent RCM realizations make up for the largest source of uncertainty in hydrological CCI studies (Gädeke et al., 2014; Muerth et al., 2012) as these components represent the (climate) model and scenario uncertainty as well as the uncertainty from the internal or natural variability of the climate system. However, their overall impact on the results of hydrological models may be less or more strong depending on which hydrological indicator is of interest. While the choice of the hydrological model may exceed the GCM-RCM combination's contribution to uncertainty for low flows, both may be equal with regards to uncertainties in high flows (Muerth et al., 2012).

Traditional hydrological CCI studies use a multi-member approach (multiple GCMs, RCMs, hydrological models) to assess the future development of a catchments hydrology in a changing climate while accounting for agreement between models. The database gained by different models and members varies due to multiple factors such as scenarios, model structure, and initial conditions. Hence, the data cannot be combined to create a single database but have to be treated independently. However, a robust assessment of potential dynamics in extreme events requires a profound database to reduce uncertainty due to statistical estimation. Thus, recent studies employ SMILEs of GCMs and RCMs for subsequent hydrological models (global to regional, lumped to semi-distributed, daily resolution, conceptual models; van der Wiel et al., 2019; van Kempen et al., 2021) to foster extreme value analysis as the differences in members only originates from internal variability (Aalbers et al., 2018; Maher et al., 2021).

This thesis focuses on the potential changes in dynamics of extreme high flows of the rivers of the Hydrological Bavaria. Therefore, a single SMILE of RCM data is used within the hydrological modelling chain to drive a distributed and process-based hydrological model for impact assessment to create a profound database of discharge values for extreme value analysis. Hence, this thesis is the first study to use this combination of a RCM-SMILE, process-based hydrological model for the catchments of the Hydrological Bavaria filling the research gap of a robust assessment of the dynamics of high flows in a changing climate for this region. Hence, this setup does not allow for the analysis of different sources of uncertainties but on uncertainties originating from internal climate variability.

More details on the individual compartments of the modelling chain are given in the following sections.

1.2.2 The climatological drivers

Data from climate models are the foundation of CC studies. Hydrological CCI studies further require observed meteorological values to train hydrological models towards an adequate representation of catchment behavior, e.g., runoff formation and discharge, and to perform a bias correction on the output of climate models.

Meteorological reference

Meteorological reference datasets either consist of gridded interpolations of in-situ measurements from meteorological stations (e.g. HYRAS, Rauthe et al., 2013), reanalysis data (e.g. ERA5, Hersbach et al., 2023) which use different data assimilation techniques to constrain a weather forecasting model for the past (hindcast), or satellite derived meteorological data (e.g., GPM, Hou et al., 2014). The quality of these datasets relies on the availability and spatial distribution of stations. Within mountainous regions such as the Alps precipitation undercatch due to wind drift is a major source of uncertainty in observations (Isotta et al., 2014; Prein & Gobiet, 2017). Hence, this error results in an underestimation of liquid or solid precipitation amounts (Poschlod et al., 2020). Thus, this error has to be taken into consideration within the steps of the modelling chain where the reference data is required (e.g. bias correction, hydrological model setup).

Some hydrological models offer integrated spatial interpolation schemes (e.g., inverse distance weighting or altitude dependent regression) to directly interpolate point observations for each time step. Thus, if these models provide the spatial interpolation results as an output, they can be used to create a climatological reference dataset.

The selection of an appropriate meteorological reference dataset is a crucial task as it affects bias-adjusted projections of climate model outputs, especially extreme precipitation values (Gampe et al., 2019). Since regional reference datasets cover large areas, their spatio-temporal resolution usually is set to a daily timestep and $1km^2$ or coarser. Thus, a new meteorological reference dataset was created within the scope of the ClimEx project to allow for model applications on a higher spatio-temporal resolution (Brunner et al., 2021b; Poschlod et al., 2020; Willkofer et al., 2020).

Climate models

Throughout different cycles of the World Climate Research Programme's (WRCP) Coupled Model Intercomparison Project (CMIP, Meehl et al., 2000), a large number of global climate models have been developed using different types of scenarios describing the future development. Subsequent modelling initiatives such as the Coordinated Regional Model Experiments initiative (CORDEX, Giorgi et al., 2009) employ these GCMs to create ensembles of RCMs through dynamical downscaling. On the global scale, GCMs and Earth System Models (ESM) provide the foundation for the analysis of climate change impacts depicted in the regular iterations of the IPCC Assessment Reports.

Along the development of GCMs and ESMs different scenario types have been introduced to assess the effects of different paths of socio-economic development and associated development in emissions on the future climate. Hence, they comprise different scenarios of greenhouse gas (GHG) emissions or their concentrations to investigate a variety of possible future developments (Collins et al., 2013) due to changes in radiative forcing. The Special Report on Emission Scenarios (SRES, Nakicenovic and Swart, 2000) formed the underlying scenarios based on socio-economic storylines for GCMs used to analyze effects of climate change within the IPCC's 4^{th} Assessment Report (AR4, Solomon et al., 2007). Since the models use diverse implementations of the carbon cycle or chemistry schemes, the GHG and aerosol concentrations could vary even when applying the same SRES scenario (Cubasch et al., 2013). Hence, a new type of scenarios was introduced for the application within models of the 5^{th} iteration of CMIP, the Representative Concentration Pathways (RCP; Collins et al., 2013; Meinshausen et al., 2011; Moss et al., 2010; Taylor et al., 2012; van Vuuren et al., 2011). Apart from not being based on socio-economic storylines directly, RCPs consider short-lived gases and land use changes (Cubasch et al., 2013). Results from these models were used to analyze CCI shown in the IPCC's 5^{th} Assessment Report (AR5 Stocker et al., 2013). In order to account for changes in society (population, education, consumption) as well as climate change mitigation and adaptation (D. Chen et al., 2021) a new set of scenarios - so called Shared Socio-economic Pathways - were developed and form the foundation for models in CMPI6 and for CC analysis presented in the IPCC's AR6 (IPCC, 2022; Meinshausen et al., 2020).

The spatial resolution of GCMs usually is too coarse (between 100 and 300km) to adequately represent regional heterogeneity, especially in mountainous regions (Di Virgilio et al., 2022; Maraun, 2016; Teutschbein & Seibert, 2012; Themeßl et al., 2011). Hence, RCMs are employed to dynamically downscale GCM outputs over a certain region providing a finer resolution (Gampe et al., 2019; Giorgi & Gutowski, 2015). Instead of using statistical methods to increase the spatial resolution, RCMs employ GCM outputs at its boundaries (Salathé et al., 2010), referred to as one-way nesting approach (Giorgi & Gutowski, 2015). It allows the RCM to generate its own fine-scale climate based on physical processes while being consistent with the patterns from its driving GCM (Giorgi, 2019).

The realizations (members) of the different RCMs applied for this dissertation originate from the 3^{rd} and 5^{th} CMIP iteration since data of the latest scenarios were not available by the time the model simulations for this thesis were produced. Hence, the results are based on SRES (A1B) and RCP (RCP8.5) scenarios. While the A1B scenario, representing a middle-of-the-road scenario, was used to determine effects of RCM correction on the climate change signal (CCS) of hydrological indicators, the more pessimistic scenario of the RCP8.5 was selected to account for distinct signals towards more severe changes in hydro-meteorological extremes.

Within the modelling framework of the ClimEx project a SMILE of regional climate simulations was created by dynamically downscaling 50 members of the Canadian Earth System Model version 2 (CanESM2) (Arora et al., 2011; Fyfe et al., 2017; Kirchmeier-Young et al., 2017) using the Canadian Regional Climate Model version 5 (CRCM5) further referred to as CRCM5 Large Ensemble (CRCM5-LE) (Leduc et al., 2019). All members are based on the RCP8.5 emission scenario and the differences among the individual members originate from differences in the initial conditions in the driving GCM, and thus represent differences originating from internal climate variability.

1.2.3 Bias correction and spatial downscaling

Along the hydro-meteorological modelling chain the output from RCMs is used to drive the hydrological model in order to assess the impact of climate change on a catchment's hydrology. However, the RCMs may exhibit more or less strong systematic deviations in seasonality and/or magnitude from observed climate for a given reference period (Rajczak et al., 2016; Smiatek et al., 2016). These deviations are referred to as bias (Ehret et al.,

2012) and are usually neglected in studies solely focusing on climate change (Themeßl et al., 2011). However, climate change impact studies typically serve to derive adaptation strategies towards climate change. Using biased data for subsequent applications (e.g., hydrological models) may result in biased responses of impact models as well. Thus, a bias correction is considered necessary for climate change impact studies (Gampe et al., 2019), despite a diversity of associated shortcomings such as the disruption of physical consistence between meteorological parameters or changes in climate change signals (Enayati et al., 2021; Maraun, 2016). Bias correction methods build on a statistical relationship between climate model output and observations for a given reference period, which are used to establish transfer functions applied to the climate model simulations to adjust the bias (Rajczak et al., 2016). Among the most common methods for BC are local lntensity scaling (Schmidli et al., 2006), linear scaling (Lenderink et al., 2007), power transformation (Leander & Buishand, 2007; Leander et al., 2008), and various forms of distribution mapping (quantile-mapping, distribution mapping (Teutschbein & Seibert, 2012); daily translation (Mpelasoka & Chiew, 2009)) for precipitation and variance scaling (J. Chen et al., 2011b, 2011a) for temperature. The quantile-mapping approach can be applied for other variables as well (Gampe et al., 2019). Hence, it is often preferred over other methods. A crucial aspect of all BC methods is the assumption of a stationary bias (Ehret et al., 2012; Huang et al., 2014; Stagl & Hattermann, 2015). Hence, the systematic deviation of the reference period also applies to the future periods. However, this assumption is still debated and might not be justified (Huang et al., 2014).

An important prerequisite for the adjustment of precipitation is the elimination of the fraction at the lowest end of the distribution from the RCM data (Gampe et al., 2019). Since the driving GCM can produce precipitation of low intensities (drizzle) too frequently, the resulting precipitation output from an RCM may inherit this drizzle (Dai, 2006; Ehret et al., 2012; Piani et al., 2010). This drizzle causes the climate model to produce more wet days compared to the observations (Suman et al., 2022). Hence, this low intensity precipitation is usually removed using a threshold derived from a distribution analysis (typically around 1mm/day) in order to match the number of wet days within the observations (Kjellström et al., 2010).

BC as a post-processing step is crucial for CCI studies and derived adaptation strategies to enhance trust in the results from the impact model. However, if those CCI studies solely focus on change signals, raw RCM data may be applied as well to reduce uncertainty and prevent deviations in change signals (Muerth et al., 2013).

The question about which BC approach yields the best adjustment of hydrological simulations using corrected RCM data to simulations using observations, as well as the effect of BC on the climate change signal of hydrological indicators are part of the scope of this thesis.

The resolution of RCMs (usually 12km) is still to coarse for its direct application in high resolution distributed hydrological models (typically 1km or higher) (Cloke et al., 2013). Hence, the RCM resolution may yield biases in hydrological simulations due to the reduced topographic detail affecting the meteorological fields (Kleinn, 2005). Thus, the RCM's resolution has to be further increased to avoid this bias and to match the resolution of the hydrological model. A simple spatial disaggregation using classical interpolation approaches (e.g., inverse distance weighting) however, does not account for the spatial variability of small precipitation events (Gagnon et al., 2012). Therefore, several statistical approaches have been developed to further downscale outputs from RCMs which differ in complexity due to employing either complex statistical disaggregation (e.g., Gagnon et al., 2012), regression (e.g., nonlinear Artificial Neural Networks, Sharifi et al., 2019), or rainfall patterns (e.g., TopoSCALE, Fiddes and Gruber, 2014; SCALMET, Marke, 2008; REGNIE, Rauthe et al., 2013) (Kay et al., 2023). The more simple approaches using rainfall patterns further account for conservation of precipitation amounts, thus, leaving extremes and dry periods unchanged (Marke, 2008; Rauthe et al., 2013). Apart from precipitation, these methods allow for spatial disaggregation of other meteorological parameters such as temperature as well or may be adapted towards it in order to provide a consistent dataset. Furthermore, these approaches allow for a more sophisticated interpolation of point observations as well (e.g., Rauthe et al., 2013) thus, avoiding simple interpolation approaches provided by the hydrological model. In this thesis adaptations of the REGNIE approach were employed for the spatial interpolation of observed values and the spatial disaggregation of the meteorological outputs of the CRCM5-LE required by the hydrological model, which in both cases were precipitation, air temperature, relative air humidity, incoming shortwave radiation, and wind speed.

1.2.4 Hydrological modelling

Along the hydro-meteorological modelling chain the hydrological model represents the last component. There is a diversity of models available covering a variety of different approaches and scales. While in some cases these models have been developed for a specific purpose, they all reproduce the hydrological cycle of a catchment, though the degree of details may vary.

Hence, there are many ways to categorize hydrological model types. The following description of model types follows the classification after Wheater et al. (1995). Although great progress in model development due to advances in computational performance has been made over the last decades, the classification scheme still applies.

Model types

The complexity of a hydrological model is characterized by its spatial discretization, process description, and the degree of determinism (Devia et al., 2015; Pechlivanidis et al., 2011).

In terms of spatial discretization a hydrological model may be considered lumped, semi-, or fully-distributed. Lumped models treat entire catchments as an enclosed entity disregarding any spatial details (Beven, 2001; Y. Chen, 2017). Semi-distributed models allow for a more complex spatial description of a catchment and are most commonly represented by areas of similar characteristics forming a single unit, the so called hydrological response units (HRU) (Paul et al., 2019). An even higher degree of spatial precision is achieved using fully-distributed models as they discretize catchments using regularly distributed units (e.g., grid cells) and solve equations representing various processes for these spatial units (Pechlivanidis et al., 2011). Thus, distributed models may capture the variability of processes, catchment characteristics, and model inputs (Pechlivanidis et al., 2011). The spatial accuracy of distributed models is however determined by the spatial resolution of the available input data as these may be averaged over several grid cells due to model scale and resolution (Beven, 2001; Pechlivanidis et al., 2011).

There are many processes involved in the formation of catchment runoff such as canopy interception, evapotranspiration, infiltration, soil-water or ground water fluxes. However, not all of these processes can be properly resolved at all scales and resolutions.

Furthermore, there may be catchment inherent processes a which are not yet fully understood. Hence, different hydrological models offer variations of methods to describe a process in more or less details or even focus on a single method.

Thus, a further distinction of hydrological models is done according to their degree of process description. On the one side, empirical (or data-driven, black-box) models employ statistical relations between observed input (precipitation) and output (discharge) (Wheater et al., 1995). Representatives of this model type are for example the antecedent precipitation index, regression models, or fuzzy logic models (Xu et al., 2017). More enhanced modelling approaches are based on artificial intelligence such as Artifical Neural Networks (Xu et al., 2017). Therefore, these models do not allow for a detailed analysis on catchment processes (Liu et al., 2017). On the other side, conceptual (or gray-box) models rely on descriptions of catchment processes prior to a model run based on relations between hydrological variables derived empirically or from observations (Pechlivanidis et al., 2011; Xu et al., 2017). Physical processes such as evapotranspiration, infiltration, soil moisture storage, and runoff generation and routing are considered in a simplified manner and thus, require a certain understanding of the catchments underlying physical and hydrological condition (Liu et al., 2017). Finally, physically- or process-based models employ non-linear partial differential equations partly derived from laboratory or hillslope scale field experiments (e.g., Penman-Monteith, Richards, Darcy) or their explicit forms to describe catchment processes such as evapotranspiration, infiltration, incerception, snowmelt, and soil-water fluxes within the unsaturated and saturated zone (Y. Chen, 2017). These complex approaches require a large database since the algorithms demand a vast number of parameters. Furthermore, to fully exploit the enhanced process description, these models usually operate on fully-distributed descretization (Y. Chen, 2017). Thus, physically-based models are also more demanding in terms of the required computational performance than conceptual or empirical models.

Despite the effort for setting up a physically-based hydrological model, their capability to more accurately represent hydrological processes than other model types can justify their application (Devia et al., 2015; Kunstmann et al., 2006). Especially, if a deeper understanding of underlying processes of discharge generation are of interest.

Hence, the selection of a model type should match the scope of the research, the spatial scale, and the quality of the available observations and input data as even the most sophisticated approach is prone to the four major sources of uncertainty: natural, data, model parameter, and model structure uncertainty (Pechlivanidis et al., 2011).

A third classification of models considers their stochasticity. Hence, hydrological models are either classified as deterministic or stochastic. A deterministic model yields the same result for a single set of input parameters (Pechlivanidis et al., 2011). In order to account for process uncertainty stochastic models employ random variables, thus yielding a different model output for each iteration while using a single set of parameters (Pechlivanidis et al., 2011).

For this thesis the deterministic, fully-distributed, and process-based model WaSiM (Schulla, 2021) was chosen as it allows for an analysis of catchment processes involved in the genesis of extreme high flows, e.g. as a response to precipitation extremes. Hence, further statements of this thesis will focus on this model type. In a climate change impact context the physically-based models are considered advantageous since changes in individual processes can be modeled. Furthermore, the finer spatial discretization allows for a higher spatial resolution to better resolve small scale processes in the topographically complex region of the Hydrological Bavaria and a better representation of intense locally defined precipitation events.

Model parametrization

In theory, physically-based models require no calibration of model parameters as they are derived through a detailed catchment analysis (Y. Chen, 2017). However, most models described as physically-based often incorporate some form of conceptual approaches (Pechlivanidis et al., 2011). Hence, these models employ two different types of parameters, physical and process (empirical) parameters (Pechlivanidis et al., 2011). Process parameters are considered free parameters as they often cannot be measured and thus, must be calibrated as they cannot be measured directly (Y. Chen, 2017; Pechlivanidis et al., 2011). Although physical parameters can be obtained through measurements, they usually must be adjusted through calibration as well due to a mismatch in scaling between the (point) measurements and the model (grid) scale (Madsen, 2003; Pokhrel et al., 2012).

Model calibration is one part of a split-sampling test (Klemeš, 1986) to determine a model's performance of simulating the hydrology of a catchment (Pechlivanidis et al., 2011). The calibration is then followed by a model validation. The validation serves to evaluate the model performance for the parameter set derived during the calibration to verify if the parameters are transferable and the model is robust to changes in inputs (Arsenault et al., 2018; Shen et al., 2022). There are three major approaches for this split-sampling test: the two-period approach (two periods of close to equal length), the mixed-bag approach (years for calibration and validation are sampled randomly from the available data), and the odd-even approach (calibration on odd years, validation on even years) (Arsenault et al., 2018). The two-period method is most commonly used according to literature (Arsenault et al., 2018). For this approach, the calibration period should cover at least 5 to 10 years in order to capture most of the flow variance (including high and low flows) (Anctil et al., 2004; Brath et al., 2004; Merz et al., 2011).

However, Arsenault et al. (2018) and Shen et al. (2022) found that classical split sample approach should be omitted and suggest to use as many data of a time series as possible, if not the entire time series for calibration as this approach yields the most robust results, thus, skipping the validation entirely. While this approach could be feasible for conceptual models or small-scale physically-based models, its expected benefit for the purpose of the study for large-scale physically-based models should be measured against the requirements of vast computational resources and time. Hence, as the Hydrological Bavaria is considered a large scale catchment and a fully-distributed physically-based model is used for simulations, this thesis employs the classical two-period split-sample approach to reduce computational demands and time needed for calibration.

In order to measure the performance of a parameter set, different goodness-of-fit criteria are available to compare model output with observed data. These objective functions (OF) usually focus on differences of observed and simulated discharge. Depending on how they handle these differences the various OFs are more or less suitable for certain purposes as they are sensitive to different types of errors or systematic deviations (Jackson et al., 2019; Krause et al., 2005). On the one hand, the widely employed Nash-Sutcliff Efficiency (NSE; Nash and Sutcliffe, 1970) is sensitive to high flows as it uses squared differences between measured and simulated discharge (Gädeke et al., 2014; Jackson et al., 2019; Krause et al., 2005). On the other hand, its logarithmic form (logNSE) is sensitive to low flows despite belonging to the same metric family since the values of observations and predictions are used in logarithmic form (Gädeke et al., 2014; Krause et al., 2005). Recently, the Kling and Gupta Efficiency (KGE; Gupta et al., 2009) is getting more attention in hydrological modelling since it incorporates different components (bias ratio of the mean, the variance, and dynamics or correlation). These components may be weighted to focus on different aspects of the flow regime (e.g., high flows) (Jackson et al., 2019; Mizukami et al., 2019). In general, it is recommended to use more than one OF to measure the performance of a model as a single criterion imposes a bias on a certain feature of the hydrograph (Jackson et al., 2019; Pechlivanidis et al., 2011). Furthermore, apart from discharge, other outputs of distributed hydrological models can be compared with spatial or point observations (e.g., soil-moisture, snow depth, groundwater depth) to improve the model performance (Tong et al., 2021). However, there are limits to this comparison due to a difference in spatio-temporal scales. Spatial validation data derived from remote sensing data may lack a reasonable spatial resolution (Wagner et al., 2007), and point observations from different gauges are used for comparison with the value of a considerably larger model grid cell (Beven, 2019).

Since physically-based models include many parameters which can't be derived from observations, and a calibration on all these parameters is not feasible, a sensitivity study is often employed to reduce their number to the most sensitive (Pechlivanidis et al., 2011). The sensitivity study may be omitted by employing expert knowledge on these parameters provided by literature or model developers.

However, even with the reduction of model parameters there are major issues with the parameterization, e.g., parameter uncertainty (due to limited data availability), over-parameterization (exploiting too many free and sensitive parameters), and equifinality (equally good, non-unique parametersets) (Bárdossy, 2007; Beven, 2001; Pechlivanidis et al., 2011). The individual parameterization of multiple catchments despite their close spatial proximity within the same region emphasizes these issues. Hence, a global calibration approach was introduced aiming towards the identification of a single parameter set which is valid for all considered gauges of the domain of interest (Gaborit et al., 2015; Ricard et al., 2013). It trades an allegedly better local performance for regionally consistent parameters (Ricard et al., 2013), thus, helping to diminish over-parameterization and equifinality by reducing well performing non-unique paremtersets.

There are two approaches for model calibration, manual and automatized. While manual calibration involves changing parameters and evaluating model results for each iteration by the user, automated calibration uses optimization algorithms to optimize (minimize or maximize) a pre-defined OF by iteratively changing the model parameters until a termination criterion (e.g., difference in OF lower than a threshold) is fulfilled (Pechlivanidis et al., 2011). Approaches for an automated calibration such as the Shuffled Complex Evolution (Duan et al., 1993) or Dynamically Dimensioned Search (Tolson & Shoemaker, 2007) differ in complexity and computational demands. Since even less complex approaches can require a substantial amount of time for processing, especially when used to calibrate large-scale physically-based models on high spatio-temporal resolution, they are often paired with a manual calibration to reduce the number of parameters they use and limit the number of iterations. Furthermore, they are not able to fully replace human judgment yet (Pechlivanidis et al., 2011).

In this thesis the objective function focuses on discharge. During the parameterization other variables have been regarded for their applicability (e.g., groundwater measurements, soil moisture at various depths, snow depth and snow water equivalent). However, most of this data was neglected for a multi-criteria approach as either the model structure was not able to reproduce the observations (e.g., groundwater depths below 20m), the spatial distribution and availability of observations was scarce (e.g., soil moisture, snow depth and water equivalent), or the available data was too fragmented (e.g., snow water equivalent).

1.2.5 Hydrological Extremes: definition and modelling

Hydrological extreme events occur on different spatial and temporal scales. Flash flood events are usually limited to a single small catchment and caused by spatially stationary intense convective rainfall over a short period of time (Blöschl et al., 2015; Garambois et al., 2013). River flood events can affect multiple catchments over several days as a response to perseverative steady rain which lasts several days yielding high volumes of precipitation falling on partly or fully saturated soils, sometimes in association with snow melt (Berghuijs et al., 2019).

By definition, such extreme hydrological events are rare which is also expressed by their respective return period being the expected time interval between events of the same magnitude (Salas et al., 2018; Slater et al., 2021). Thus, existing time series of observations exhibit only a few of these events, especially if these time series only cover a few years or decades. In Bavaria, most of the available stations provide data for up to 50 years. Only a few stations offer longer time series of up to 120 years (GRDC, 2021). Hence, there are large uncertainties when it comes to ascribing a certain return period to a recent event as the available data do not allow for a robust estimation of extreme return periods of 100-years and above (Schulz & Bernhardt, 2016). In general, if the estimation of the return period is based on annual maximum discharges, the available discharge time series should at least span a third of the years of the return interval to be estimated in order to reduce uncertainties (Maniak, 2010). Furthermore, discharge values for high flows of a given return period are usually not stationary in time as natural and anthropogenic changes affecting the hydrological cycle also affect extreme hydrometeorological events (Salas et al., 2018). Thus, the location of the period along the time series as well as its length chosen to estimate these values has a significant impact on it (Schulz & Bernhardt, 2016). However, the discharge value of a 100-year flood is often referred to as a threshold for a design criterion for flood protection and hydro-power facilities (Brunner et al., 2021a; Slater et al., 2021). This stationary threshold is based on estimates using available observed data. Therefore, as the threshold varies with time, it might not be suitable as design criterion for these structures in the future since the structures are adapted to the current representation of extremes.

A flood frequency analysis (FFA) is typically used to determine the relation between the magnitude of a flood and its frequency of occurrence (Gharib et al., 2017; van Kempen et al., 2021). Estimates of high return periods which exceed the length of recorded events require an extrapolation through the application of a extreme value distribution (EVD) function (or probability distribution function, PDF) fitted to the distribution of peak flow samples (Curceac et al., 2020; E. M. Fischer & Knutti, 2016). There are two commonly used approaches for sampling time series of high flow events: the block-maxima method (e.g., annual maximum, AM) and the peak over threshold (POT) method (Zhao et al., 2019). PDFs based on samples derived using the AM approach (e.g., Weibull, Pearson III, log Pearson III, GEV, Gumbel) are still widely used and recommended by various countries (Lawrence, 2020; Meresa et al., 2022; Pan et al., 2022; Salas et al., 2018). However, their estimates for higher return periods can be rather uncertain due to short sample sizes and inevitable extrapolation (S. Fischer & Schumann, 2022; Lawrence, 2020).

Furthermore, the distribution of AM values may include events which would not be considered as an actual flood, while multiple peaks within a year that are higher than those of other years are omitted (Zhao et al., 2019). Samples of peak flows obtained by the POT approach containing independent and identically distributed values follow the Generalized Pareto Distribution (GPD) (Solari & Losada, 2012). This EVD is most commonly used for these partial time series of peak discharges which exceed a threshold despite other existing approaches (e.g., exponential distribution being the reduced form of the GPD) (Gharib et al., 2017; Pan et al., 2022). However, the determination of a threshold is crucial to offer a large enough database that follows the GPD (Curceac et al., 2020). Despite the increase in robustness for high return periods when employing the POT approach and the GPD due to the larger sample size, short time series of observations still don't provide an adequate number of events for stable estimates of higher return periods (E. M. Fischer & Knutti, 2016).

PDFs are fitted to the distribution of the samples by adjusting their parameters (e.g., shape, skewness, location, scale), where different PDFs require more or less parameters (Cassalho et al., 2018). To determine optimal parameters to fit the EVD to the distribution of the samples a diversity of parameter estimation algorithms can be applied. These are, among others, the method of moments, the maximum likelihood estimator, the probability of weighted moments or L-moments (Curceac et al., 2020; Salas et al., 2018; Zhao et al., 2019). From these approaches the L-moments algorithm is commonly used as it is less sensitive to outliers and thus, often yields a more robust fit of the EVD for both, AM and POT approaches (Durocher et al., 2019; S. Fischer & Schumann, 2022).

Recent research seeks to avoid the application of statistical estimates through the creation of large time series of discharge values using (usually conceptual) hydrological models in combination with climatological SMILEs (van der Wiel et al., 2019; van Kempen et al., 2021). The hydrological model translates the vast meteorological information of a SMILE (usually a GCM) into a hydrological response resulting in a profound hydrological database of hundreds of model years. The discharge time series of these hydrological models can then be used to empirically derive events of high return periods (van Kempen et al., 2021). Thus, hydrological modelling serves as a tool to enhance the database and foster extreme value analysis. Furthermore, this approach allows for a thorough and robust analysis on the dynamics of intensity and frequency of extreme flows in a changing climate. However, applying a single SMILE only allows for the analysis of natural variability but not for an analysis of variability according to different models or scenarios (Schulz & Bernhardt, 2016; Sperna Weiland et al., 2022).

A similar approach is to employ enlarged time series from statistical weather generators based on extrapolations of current (historical) conditions and future periods under climate change conditions to drive a hydrological model also resulting in long time series suited for FFA (Hattermann et al., 2018). However, this approach does not account for a physical representation of meteorological variables and uses independent historical and future periods which does not allow for a transient analysis of changes in hydrological processes and discharge.

Therefore, as mentioned in chapter 1.2.1, for this thesis a new approach was established to foster FFA of extreme peak flows using the output of 50 members of a RCM, the CRCM5-LE, to drive a fully-distributed, process-based hydrological model for the Hydrological Bavaria. This approach provides transient simulations for the period between 1961 and 2099 for climate change analysis and a total of 1,500 model years per 30-year evaluation period (30 years times 50 members). Thus, it allows for a thorough exploitation of the data for FFA based on empirical probabilities. It is the first time a RCM SMILE was used in combination with a distributed hydrological model.

Since this model type is demanding in terms of computational performance and time and both of these resources are usually limited, a compromise between spatio-temporal resolution and expected loss or gain in information from simulations of peak discharges should be considered. While a higher spatial resolution allows for a more accurate representation of small scale convective precipitation events, the expected computation time increases close to a squared ratio. However, a higher temporal resolution only results in a linear increase in computation time, but allows for a better approximation of actual observed peak flows as exemplary shown in Figure 2a for a flood event within the Isar catchment at the gauge Plattling (8,613km^2), situated within the Hydrological Bavaria.

Hence, for this thesis, an a priori study on an optimal spatio-temporal resolution was conducted using the same distributed hydrological model as for all other analyses to find a compromise between computational demand and accuracy in peak discharge simulation (see Fig. 2b). The experiment employed model setups in different spatio-temporal resolutions (100m to 1000m; 1h to 24h) for the Isar catchment up to the gauge Plattling.

The model parameters for the setup with the highest spatio-temporal resolution (100m, 1h) were considered the best parameters and not changed for the remaining setups to assess the effects of changes in spatio-temporal resolution. The empirical cumulative distribution function (ECDF) of the modeled discharges of the experiment revealed that spatial resolution has a minor effect on the model results (2b, differently dashed lines). However, changes in temporal resolution (Fig. 2b, differently colored lines) lead to dif-

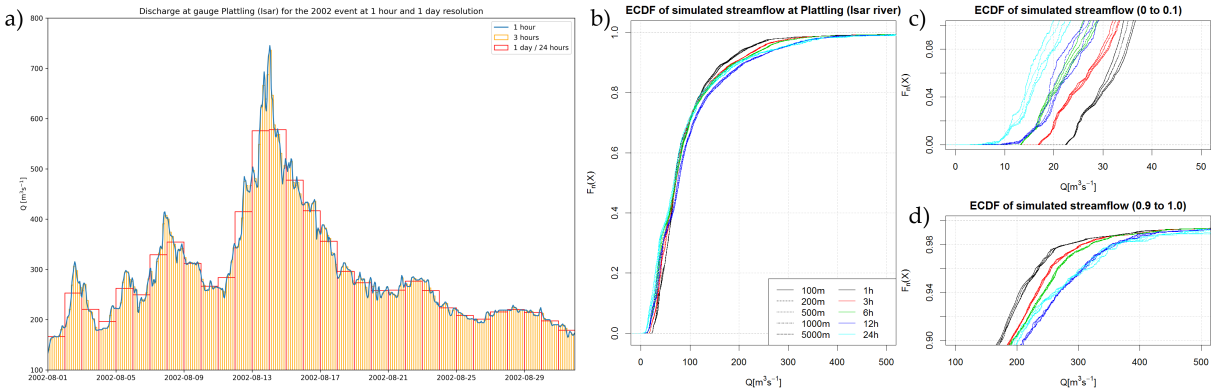


Figure 2: Effects of spatio-temporal resolution on the representation of peak discharges. a) illustrates the difference in peak discharges between hourly observations and 3-hourly and daily aggregation. b) illustrates the effects of changes in spatio-temporal resolution on discharge simulations of hydrological models for the entire ECDF, c) shows the lower part of the ECDF (i.e., low flows), d) shows the upper part of the ECDF (i.e., high flows).

ferences in the lowermost (Fig. 2c) and uppermost (Fig. 2d) percentiles of the distribution. Peak discharges (highest percentiles) are continuously underestimated with decreasing temporal resolution. For the low flows the coarser temporal resolutions result in lower discharge values than compared to the 1h resolution. Since the focus of this thesis is the assessment of CCI on the dynamics of peak discharges of larger catchments rather than small scale events within the Hydrological Bavaria, a high spatio-temporal resolution of 500x500m^2 and 3 hours was used for the setup of the distributed hydrological model.

1.2.6 The role of high performance computing

The creation of a 50 member RCM, its adjustment to observed conditions, the further rescaling to the hydrological model resolution, and its application for the hydrological modelling itself requires extensive computational resources in terms of performance and storage. In order to complete these tasks in a timely manner, currently available personal computers lack the required performance nor provide robust and sufficient storage. Hence, the ClimEx project partnered with the Leibniz Supercomputing Centre (LRZ) of the Bavarian Academny of Science to cope with the intense demand in computational resources. Their high performance computing (HPC) and scientific storage solutions offer vast resources for Bavarian universities. To create the CRCM5-LE a total of 1,000,000 core hours (CH) of processing time (CPU time) and 400 Terabytes (TB) of storage was granted by the Gauss Centre for Supercomputing and spent on the HPC systems of the LRZ. Additionally, 100,000 CH and 72 TB of storage were used to create the hydrological large ensemble through the application of the CRCM5-LE as climatological driver for the hydrological model.

The LRZ provides support for the adaptation and improvement of the source code of models to be applicable for their HPC systems. However, since RCMs and hydrological models heavily depend on reading input and writing output (I/O) the scalability - capability of a model to split a task over as many cores as possible without losing performance - of these model types is limited which is referred to as bottleneck. Hence, to enhance the usage of cores, a so called job scheduler was provided by the LRZ to run as many jobs (unit which describes a separate action, e.g., running a model) in a single submission (referred to as embarrassingly parallel execution). Furthermore, the LRZ offers meta data mapping for the distribution of selected variables through the download portal GLOBUS as well as permanent storage of data on tape for irregular access.

1.3 Scientific scope of the thesis

The scope of this dissertation is to develop a robust approach to determine how hydrological extreme events in the Hydrological Bavaria might respond to projected changes of future climate conditions, especially of extreme precipitation.

In order to provide more robust estimates of this development, novel approaches have been introduced to the common CCI modelling chain. These approaches focus on the application of the CRCM5-LE as a driver for a hydrological model to exploit the simulated runoff projections for the robust estimation of peak flows of extreme return periods in the Hydrological Bavaria. Thus, the research questions address the hydrological model setup to create the base for the analysis, an optimal method to adjust the climate model output to the observations, and the benefit of a hydrological large ensemble for the analysis of extreme values and their dynamics under a strong climate change scenario (RCP8.5).

To assess the CCI on extreme peak flows, at first, a hydrological model for the Hydrological Bavaria has been developed in high spatio-temporal resolution to avoid related shortcomings for a robust estimation, e.g., underestimation of discharge peaks. A single set of regionalized parameters was employed over a heterogeneous region comprising different landscapes and flow regimes. Thus, the research questions regarding capabilities of the model to represent peak flows are as follows:

Q1: *Can a holistic model setup, parameterized towards high flow representation, provide sufficient performance across heterogeneous catchments?*
Q2: *Does this holistic model setup qualify for the simulation of flood events with higher return periods?*

Hydrological CCI studies rely on the meteorological output from climate models (GCM or RCM). While the variability from observations for a certain reference period is inevitable at the respective time step, the long term regime of a variable of the model should not exhibit a large systematical deviation from the observed regime. Thus, the model output has to be adjusted to eliminate this bias. Many different bias correction approaches are available which usually focus on precipitation or air temperature. However, an adjustment alters the output from climate models and should be avoided if the bias in magnitude and seasonal course is negligible. If the bias is non-negligible and has to be corrected for CCI studies, the altered values may affect CCS of the climate model and consequently of the hydrological model. Thus, the research questions regarding the suitability of different bias correction methods are as follows:

Q3: *Which method for bias correction is recommended in terms of best representation of observed discharge regimes and hydrological indicators?*
Q4: *How does bias correction affect the climate change signal of hydrological indicators?*

The overarching aim of this thesis addresses the question of how climate change affects the dynamics of extreme peak flows in terms of intensity and frequency. The choice of climatological drivers for the hydrological model either facilitates or antagonizes a development towards these extreme events. Hence, the CRCM5-LE is forced by the RCP8.5 emission scenario which represents a path of severe changes towards the end of the 21^{st} century. Furthermore, the CRCM5-LE with its 50 members represents a SMILE. Its subsequent application for hydrological modelling results in a hydrological SMILE (hydro-SMILE) of 50 equally probable representations of the hydrology of the Hydrological Bavaria. Thus, exploiting this profound database of 1,500 models years for a 30-year period offers an extraordinary opportunity for EVA. Therefore, the research questions regarding the capabilities of hydro-SMILEs and their robust analysis of flood characteristics are as follows:

Q5: *Can SMILEs of RCMs contribute to facilitate robust estimates of peak discharges of high return periods?*
Q6: *Will the frequency and intensity of peak discharges of high return periods decrease or increase due to climate change?*

2 Scientific Publications

This dissertation builds on three scientific publications published in or submitted to peer reviewed journals relevant to the field of research. These publications are presented in the following sections. Paper I and II have been published already, while paper III has been submitted. The publications presented in this thesis are introduced by a brief and concise description comprised by its reference, a short transition in the form of a short summary indicating its placement within the thesis and among the other papers, its publication status, its publisher and associated impact factor according to the Journal Citation Reports (JCR) by Clarivate Analytics as stated by the respective journal. Furthermore, the contribution of each author to the respective publication is stated. All papers address specific steps of the hydrological impact modelling chain as shown in Figure 3 and the relevant research questions associated to the respective step.

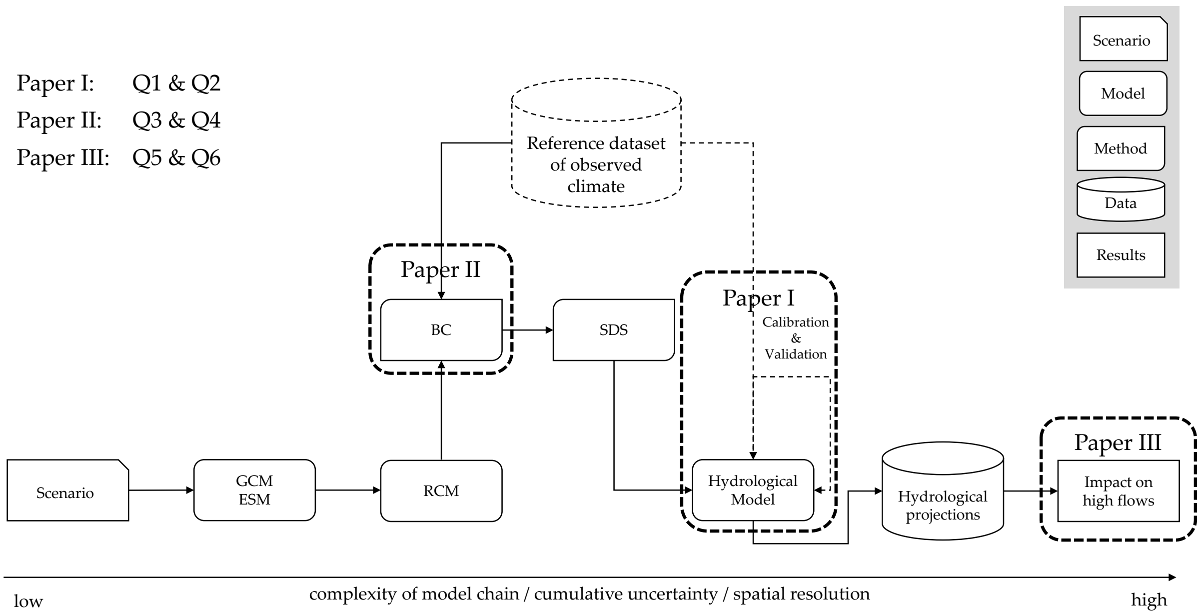


Figure 3: Placement of the scientific contributions of this thesis within the hydrological impact modelling chain.

The publications are presented in a non-chronological order to be consistent with the stages of the hydrological impact modelling chain. Paper I deals with the development of a hydrological model setup, paper II seeks to determine an optimal method for bias correction for hydrological modelling, and paper III addresses the potential dynamics of extreme high flow events due to climate change through hydrological modelling. Papers II and III employ data from different ensembles of RCMs. In paper II a multi-model ensemble of three RCMs (two of them also multi-member models) forced by the SRES A1B emission scenarios were applied. In paper III a single model initial condition large ensemble (SMILE) of 50 members of the CRCM5 forced by the RCP8.5 emission scenario was used as the climatic input for the hydrological model described in paper I which resulted in the hydrological large ensemble (hydro-LE or hydro-SMILE). However, the different emission scenarios in paper II and III do not affect the conclusions drawn from paper II as these focus only on the effect of bias correction on the CCS and not absolute magnitudes of the projections.

2.1 Paper I: A Holistic Modelling Approach for the Estimation of Return Levels of Peak Flows in Bavaria

Reference: Willkofer, F., Wood, R. R., von Trentini, F., Weismüller, J., Poschlod, B., & Ludwig, R. (2020). A Holistic Modelling Approach for the Estimation of Return Levels of Peak Flows in Bavaria. Water, 12(9), 2349. https://doi.org/10.3390/w12092349

Transition to paper I: This publication provides a novel approach to hydrological modelling of severe floods for 98 catchments of the Hydrological Bavaria. The approach comprises a single setup for a hydrological model in high spatio-temporal resolution for a more detailed representation of processes contributing to extreme discharges. A single set of parameters was determined among all the different catchments using a mixed approach of manual and automated calibration techniques. Commonly applied objective measures to determine the model's performance were selected according to their affinity towards high flows as well as overall discharge representation. The paper further introduces a Level of Trust (LOT) for the capability of the model to represent observed high flows of up to a 20-year event. Extreme values were estimated using the Generalized Pareto Distribution (GPD) on samples of flood events derived using the peak over threshold (POT) approach. The model introduced in this paper forms the basis for the analysis on the dynamics of extreme peak discharges presented in paper III.

Author's contribution: FW designed the concept for this publication, performed the formal analysis and investigation, and wrote the original draft. FW further was responsible for the visualization of the presented results. FW and BP selected the methods applied in this study. FW developed the hydrological model while JW was responsible for the adaptation of the model code for the high performance computing systems. FW, RRW, and FT performed the model validation and were responsible for data curation. RL provided the resources, supervised the research, and was responsible for project administration and funding acquisition. RRW, FT, BP, JW, and RL reviewed and edited the manuscript.

Status:	published
Journal:	Water (MDPI)
Impact factor:	3.4

water

MDPI

Article

A Holistic Modelling Approach for the Estimation of Return Levels of Peak Flows in Bavaria

Florian Willkofer [1,*], Raul R. Wood [1], Fabian von Trentini [1], Jens Weismüller [2,3], Benjamin Poschlod [1] and Ralf Ludwig [1]

1 Department of Geography, Ludwig-Maximilians-Universität München, 80333 Munich, Germany; raul.wood@lmu.de (R.R.W.); fabian.trentini@lmu.de (F.v.T.); benjamin.poschlod@lmu.de (B.P.); r.ludwig@lmu.de (R.L.)
2 Leibniz Supercomputing Centre, Garching, 85748 Munich, Germany; jweismueller@bfs.de
3 Federal Office for Radiation Protection, 85764 Oberschleißheim, Germany
* Correspondence: florian.willkofer@lmu.de

Received: 20 July 2020; Accepted: 18 August 2020; Published: 21 August 2020

check for updates

Abstract: This study introduces a holistic approach for the hydrological modelling of peak flows for the major Bavarian river basins, referred to as Hydrological Bavaria. This approach, intended to develop a robust modelling framework to support water resources management under climate change conditions, comprises a regionalized parameterization of the water balance simulation model (WaSiM) for 98 catchments in high temporal (3 h) and spatial (500 m) resolution using spatially coherent information and an automatized calibration (dynamically dimensioned search–simulated annealing, DDS-SA) for storage components. The performance of the model was examined using common metrics (Nash & Sutcliffe Efficiency (NSE), Kling-Gupta Efficiency (KGE)). The simulations provided the means for the calculation of a level of trust (LOT) by comparing observed and simulated high flows with a five, ten, and 20-year return period. These estimates were derived by the Generalized Pareto Distribution (GPD) applying the peak over threshold (POT) sampling method. Results show that the model overall performs well with regard to the selected objective measures, but also exhibits regional disparities mainly due to the availability of meteorological inputs or water management data. For most catchments, the LOT shows moderate to high confidence in the estimation of return periods with the hydrological model. Therefore, we consider the holistic modelling approach applicable for climate change impact studies concerned with dynamic alterations in peak flows.

Keywords: hydrological modelling; holistic parameterization; return levels; peak flows; Bavaria; dynamically dimensioned search; GPD

1. Introduction

Within the last three decades hydrological extreme events (i.e., major floods and droughts) frequently occurred in Bavaria (floods: 1999, 2002, 2005, 2013, 2016; droughts: 2003, 2007, 2011, 2015, 2018). Such events impose severe risk and damage to infrastructure, economies, and civil security and–according to the Intergovernmental Panel on Climate Change (IPCC)–are likely to amplify in the future due to changes in extreme precipitation [1,2]. Providing a hydrological modelling scheme that will remain applicable under changing (climatic) driving conditions requires a high-resolution, process-based, and spatially explicit and thus often computationally demanding solution tool. The ever-increasing computational power (especially the application of high-performance computing (HPC)) allows for new approaches to be developed and applied to investigate current and future extreme events. Most hydrological models-usually driven by daily data on the mesoscale (i.e., 10^2 to 10^4 km^2 [3])–are supposed to sufficiently simulate mean hydrological conditions and long-term water balances using different approaches. However, observed hydrological extreme events show, that over complex terrain processes on shorter timescales are important and that peak flows are underestimated on daily timescales.

In order to investigate the possible changes in extreme hydrological events for Bavarian catchments the ClimEx project (Climate Change and Extreme Events–Risks and Perspectives for Water Resources Management in Bavaria and Québec) was founded to investigate and assess changes in hydro-meteorological extreme events induced by a changing climate. In this project a complex model chain was introduced comprising the creation of a large ensemble of a single initial condition climate model [4] employing the IPCC emission scenario (Representative Concentration Pathways RCP8.5 [5]). The project concludes in an application of these data for a hydrological model to investigate the impacts of climate change on the hydrology of Bavaria.

In recent years, several other studies focused on the analysis of the impacts of climate or global change on the water resources of the Danube [6–8]. Therefore, these studies employed different climate models to drive different hydrological models (e.g., PROMET [9], WaSiM, former WaSiM-ETH [10]) to assess climate induced changes in the hydrology of the upper Danube basin by measures of mean yearly and monthly flows. Apart from the model PROMET, the hydrological models applied in these studies were locally calibrated for each sub-catchment and were set up on rather coarse spatio-temporal resolutions. In Bavaria changes in frequency and intensity of hydrological extreme events have so far not been extensively investigated. In a recent study by Hattermann et al. [11] the semi-distributed hydrological model SWIM (Soil and Water Integrated Model) [12] was employed to investigate possible future changes of flood events (frequency and intensity) in the entire Danube basin divided into 50 sub-catchments.

Hydrological simulations for individual catchments within a larger river basin are prone to equifinality of the parameters gained by local calibration since these parameters are non-unique and may vary from catchment to catchment [13–15]. However, Samaniego et al. [16] and Kouchi et al. [17] summarize that parameters for process based hydrological models operating on the mesoscale have to be calibrated as they cannot be measured. To overcome the issue of equifinality (i.e., multiple sets of parameters provide similar model performance [15]) and over-parameterization (i.e., fitting parameters to countervail flaws in the model structure and observations [18,19]) there are several approaches proposed by various authors. One of these approaches is the application of a macroscale model that employs a vast amount of gauges for a local calibration [18,20]. Another approach proposed by Gaborit et al. [18] is the use of a physically based hydrological model that does not require calibration over a large area (e.g., PROMET). However, this approach requires a profound database and is demanding in terms of labor and time [18,20]. Finally, the regionalization of model parameters among various but similar catchment is widely adopted for hydrological modelling on the mesoscale [14,16,18,19,21], also referred to as global parameterization [14,18]. This approach assumes that similar catchments exhibit a similar runoff behavior, with similarity being defined by spatial proximity or catchment characteristics (i.e., land use, topography, soil types) [15,18,19].

In this study, the hydrological model WaSiM is set up for the major Bavarian river basins, i.e., upper Danube (comprising Danube, its tributaries and the Inn), Main and small parts of the Elbe (hereafter referred to as Hydrological Bavaria) covering basins from different German states and adjacent countries (Austria and Switzerland). Following the work of Ricard et al. [14] and Gaborit et al. [18], this approach comprises the identification of a set of model parameters which are valid for similar catchment characteristics of the Hydrological Bavaria. Since the regionalization is conducted over an entire region rather than pre-selected individual catchments, we further refer to this approach as holistic. Hence, we opt for global model settings (i.e., one model approach representing a specific process of the hydrological cycle) if the model structure supports it. Therefore, the hydrological model WaSiM is set up using a database comprised of a single data set for each spatially distributed input (soil, land use, hydrogeology, and meteorology) to foster a holistic parameterization approach during the calibration.

To our knowledge, this is the first time a single hydrological model is set up for the 98 selected catchments of the entire Hydrological Bavaria in high spatial (500 m) and temporal (3 h) resolution. Hence, compared to preceding studies [6,8,9], the focus on the upper Danube basin is expanded with the Main basin. With regard to Hattermann et al. [11], the study area is narrowed but displayed in more detail in terms of examined gauges and sub-catchments. The global parameterization approach presented by Ricard et al. [14] and Gaborit et al. [18] is extended to interconnected rather than independent catchments an incorporated into a single model setup.

With regard to the aforementioned projected changes in hydrological extremes, the intention of the presented analysis in this paper is to provide a modelling framework to support the estimation of robust return levels of peak flows to foster climate change impact studies concerning high flows. Therefore, this paper seeks to provide insight about the capability of the holistic modelling approach (i.e., one set of parameters applied to a variety of catchments) to sufficiently represent the observed discharge, with a focus on peak flows and their return levels, at the 98 presented gauges distributed across the Hydrological Bavaria.

2. Study Area, Data, and Methods

2.1. Study Area

The study area illustrated in Figure 1 encompasses the catchments of three major river systems and its tributaries flowing inward and outward of Bavaria. Namely the upper Danube upstream of Achleiten, the river Main upstream of Frankfurt Osthafen, and the Inn river to its confluence with the Danube at the city of Passau. Furthermore, it comprises three minor tributaries to the Elbe catchment. The Hydrological Bavaria covers parts of multiple German states (Bavaria, Baden-Württemberg, Hessen, and Thuringia) along the upper Danube, Main and Elbe tributaries as well as parts of Austria and Switzerland on the Inn. For this study, the Hydrological Bavaria was further divided into 98 sub-catchments due to their importance for flood protection and representation of the respective catchment characteristics.

With an overall catchment size of approximately 103,200 km^2 the Hydrological Bavaria includes different complex landscapes: the Alps in the South, the Alpine foreland north of the Alps bounded by the course of the Danube, the southern German escarpment and the eastern mountain ranges to the East. The elevation ranges from 90 m.a.s.l. at the gauge Frankfurt Osthafen in the North to 4049 m.a.s.l. at Piz Bernina (Switzerland) in the Southwest. Hence, the Hydrological Bavaria is characterized by a complex geography featuring a variety of runoff regimes [22] as well as different climatological conditions.

The rivers of the Hydrological Bavaria are in parts severely influenced by structural retention facilities (i.e., reservoirs for flood protection or hydro power generation) or transfer systems for drinking water supply or raising low flow levels in the Regnitz and Main river (e.g., Main-Donau-Channel). Only a few rivers remain in their natural condition. Besides these artificial retention structures, also natural lakes influence the discharge of rivers. These natural and artificial structures must be considered for the hydrological modelling to avoid large deviations between simulated and observed runoff.

The mean climate of the Hydrological Bavaria is characterized by a general increase of mean annual precipitation (Figure 2a) sums from north (500–1100 mm, the southern German escarpment) to south (1500–2500 mm, the northern Alps). However, within the Austrian Alps the Inn catchment represents a more arid region with its dry valley encompassing precipitation sums between 600 and 1400 mm, whereas outside the valley exceeding 1400 mm. The catchment of the upper Salzach exhibits higher yearly mean precipitation sums between 800 and 2100 mm. The annual mean temperatures (Figure 2b) also depict a North-South gradient with higher values in the Main catchment (10 °C) and lower values in the Alps (5 °C). While the lowest values are observed on alpine summits (−8 °C), the highest values occur along the Main valley as well as near the Main outlet (10.5 °C) Annual air temperatures for catchments of the upper Danube north of the Alps vary between 6 °C in the vicinity of the Alps and 9 °C for the remaining regions.

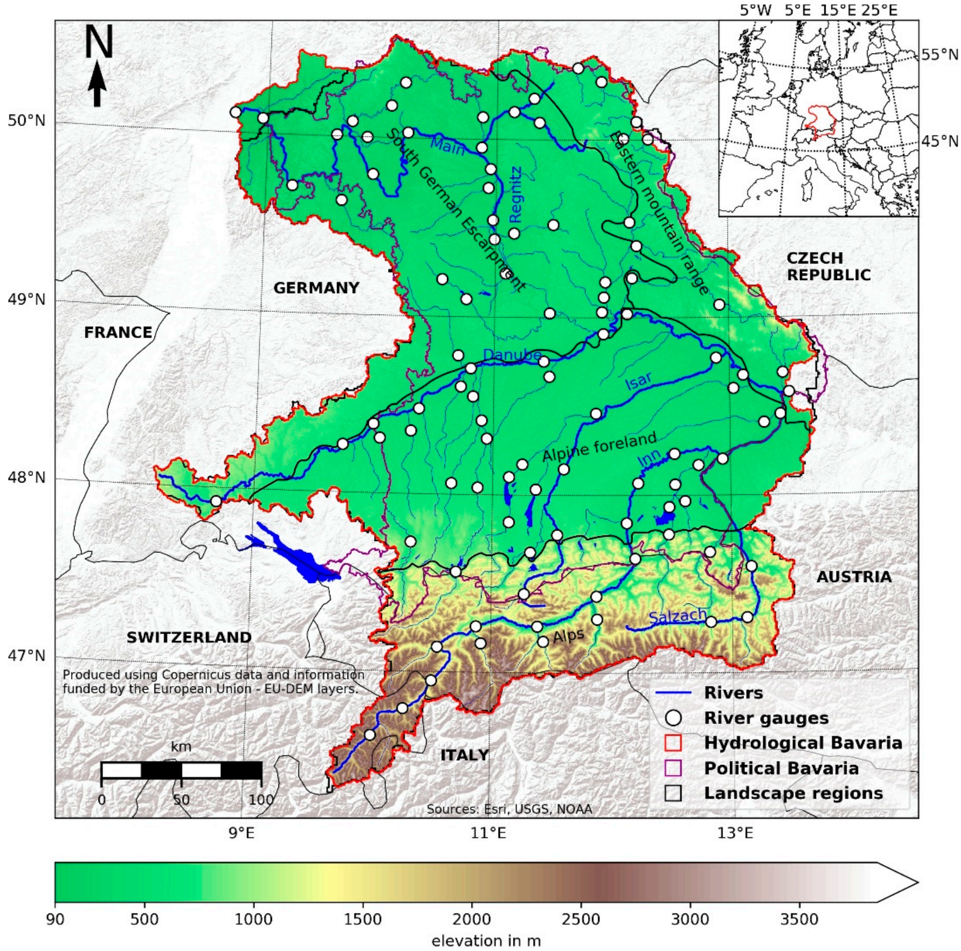
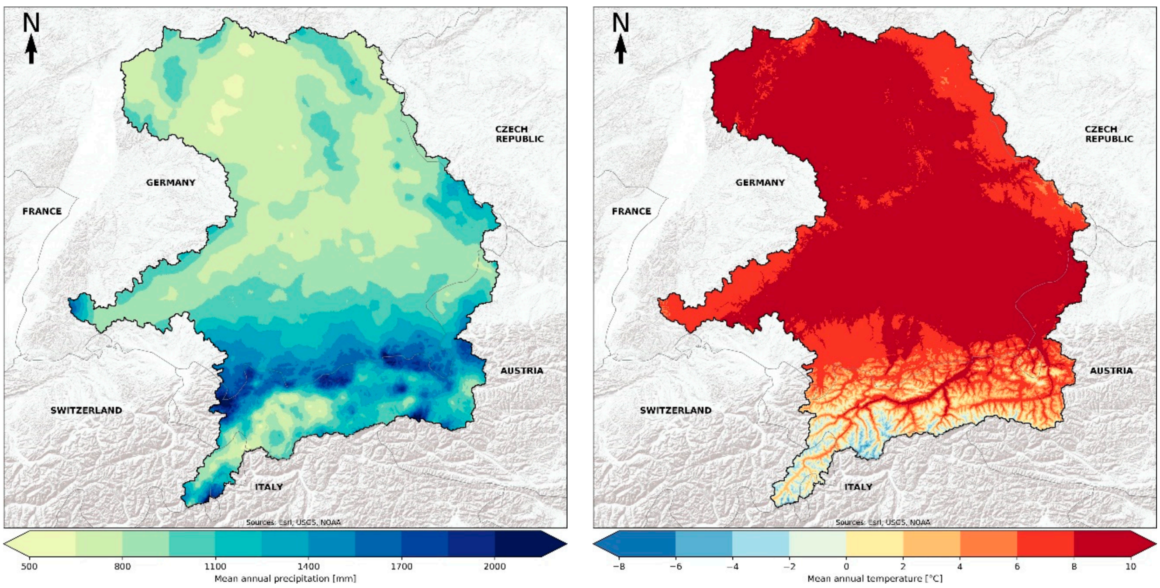


Figure 1. Map of the Hydrological Bavaria (red line) comprising different landscape regions (black lines) and the political Bavaria (purple line). The map further indicates the 98 gauges used in this study (white dots) at their respective rivers (blue lines).



(**a**) (**b**)

Figure 2. Mean annual precipitation [mm] (**a**) and air temperature [°C] (**b**) of the Hydrological Bavaria for the reference period 1981–2010 derived from the meteorological reference dataset (see Section 2.2.2).

2.2. Data

2.2.1. Spatial Model Inputs and Discharge Data

The applied hydrological model requires a plethora of data to perform the simulation in its most physically based process representation. Table 1 gives an overview of the data used for this study either as model input (e.g., spatially distributed inputs) or for validation purposes (e.g., discharge data). The digital elevation model forms the basis for the hydrological simulations since it is used to derive important model inputs (e.g., slope, exposition, catchment delineation). For this study the Digital Elevation Model over Europe (EU-DEM) [23] was used. Information on land use was obtained from the Corine Land Cover 2006 (CLC) [24] dataset. Similar land use types were merged to a single category to avoid redundancy and overfitting of land use parameters. Soil data from the European Soil Database (ESDB v2.0) [25,26] was used to provide a common basis for the political Bavaria and adjacent states and countries. This dataset fits into the study's framework of a holistic parameterization approach as the provided parameters were derived by homogenizing particle-size and hydraulic properties of soils across European datasets yielding the Hydraulic Properties of European Soils (HYPRES) database [27]. HYPRES provides Mualem–van Genuchten parameters for eleven soil textural classes (5 for topsoil, 5 for subsoil and one organic) required for the hydrological modelling. These parameters were used to describe the hydraulic properties for each individual soil type and their two different horizons. Apart from data for land use and soil texture, information regarding groundwater conductivity was required to model groundwater fluxes. While detailed data was available for the German regions of the Hydrological Bavaria from the Hydrogeologische Übersichtskarte 200 (HÜK200) [28], for regions outside of Germany the information on groundwater conductivity was retrieved from the coarser resolution International Hydrogeological Map of Europe (IHME1500 v1.1) [29]. This coarser resolution IMHE1500 was further refined based on the slope, with steeper slopes receiving lower values of conductivity values and flat areas higher values. The rationale behind this approach is that shallow areas within the Alps (mainly valleys) tend to accumulate gravel and other coarse material with higher hydraulic conductivity than steep areas with raw soils or bare rocks. Overall, the resulting spatial pattern of hydraulic conductivity classes was comparable with that of the IHME1500 dataset. All spatial data were scaled to a spatial resolution of 500 m × 500 m.

Table 1. Input and validation data for the hydrological model.

Type	Name	Resolution	Source
Land use	Corine Land Cover 2006 v17 (CLC)	100 m × 100 m	[24]
Soil	European Soil Database v2.0 (ESDB)	1:1,000,000	[25,26]
Digital elevation model (DEM)	Digital Elevation Model over Europe (EU-DEM)	1′ (≅25 m)	[23]
Hydrogeology	Hydrogeologische Übersichtskarte 200 (HÜK200) v2.5/International Hydrogeological Map of Europe 1:1,500,000 (IHME1500 v1.1)	1:200,000/1:1,500,000	HÜK200 © BGR & SGD 2011, [28]/IHME1500 v1.1 © BGR, Hannover, 2014, [29]
Meteorological data	Sub Daily Climate Reference (SDCLIREF)	500 m × 500 m	
Discharge	Gauging stations		Bavarian Environment Agency (LfU)

For this study discharge data of 98 gauges for the period of 1 January 1980 to 31 December 2010 was provided by the Bavarian Environment Agency (Bayerisches Landesamt für Umwelt-LfU) in a temporal resolution of 1 h. This data was aggregated to a 3-hourly timestep.

2.2.2. Meteorological Data

In the framework of this study a new spatially distributed dataset of meteorological observations was created. Therefore, available meteorological data comprising point measurements of precipitation (P), air temperature (T), relative humidity (RH), incoming shortwave radiation (R), and near surface wind speed (WS) at various temporal scales (daily, hourly, sub-hourly) were homogenized (temporal scale) and spatially interpolated. Sub-daily records of precipitation data began in 1989. However, the spatial coverage and the number of available stations was insufficient for a robust spatial interpolation. Hence, daily observations were temporally disaggregated to time series of 3-hourly resolution using the method of fragments (MOF) approach [30] to extend the otherwise scarce data base. The disaggregation is based on resampling sub-daily fragments for a daily observation by choosing from a set of ratios (fragments) of sub-daily to daily precipitation. Compared to other stochastic methods for disaggregation of precipitation (e.g., Bartlet–Lewis or cascade-based methods) MOF does not require parameterization and is supposed to outperform more complex methods [31]. For this study the regionalized MOF after Westra et al. [32] was applied as it provides a larger sampling size by including records of nearby stations and also acknowledges the characteristics of preceding and successive rainfall of a prolonged event. The applied MOF framework is also described in Poschlod et al. [33]. The approach was further adjusted regarding the temporal course to be applicable to other meteorological data. The received time series of P, T, RH, R and WS were spatially interpolated according to the REGNIE (REgionaler NIEderschlag; eng.: regionalized precipitation) method [34] on a spatial resolution of 500 m. Originally developed as an interpolation scheme for precipitation it combines inverse distance weighting and a multiple linear regression which considers orographical conditions. Adjustments regarding the statistics of the remaining meteorological variables were made to provide a consistent approach. This new high-resolution set of meteorological data is hereafter referred to as Sub-Daily CLImate REFerence (SDCLIREF).

2.3. Methods

2.3.1. The Holistic Modelling Approach

The goal of the holistic modelling approach is to find a single set of parameters for the hydrological model, which is valid across the entire domain of the Hydrological Bavaria. This approach is similar to a global parameterization which might be performed using independent catchments as well as model setups [14,18]. However, it is considered as holistic as it derives model parameters using a single model setup for several interconnected (and thus interdependent) sub-catchments. Hence, this approach encompasses two major premises: the application of homogeneous spatially distributed input data as well as a set of homogeneous model parameters for spatially distributed components of the hydrological model. The holistic approach should make the results from individual catchments more comparable across the domain by reducing the risk of over-parameterization often seen in single modeling setups. Especially in hydrological setups focused on the impacts from external forcing (e.g., climate change) a holistic approach is preferable. Furthermore, by avoiding over-parameterization the holistic approach may contribute to a convergence of the different philosophies of climate and hydrological modelling.

To assure the homogeneity of spatially distributed input data, we opted for data sets covering the entire Hydrological Bavaria. In limited cases where no high detailed data was available for the entirety of the domain, we have merged datasets and applied an additional set of refinement steps in post-processing (e.g., for the hydrogeology inputs).

The parameters for spatially distributed model components (i.e., parameters for land use, snow, and glaciers) were manually calibrated over the entire Hydrological Bavaria. Other parameters directly influencing the shape of discharge compartments (recession parameters for inter-, and direct flow) which are dependent on the unique features of each catchment, were automatically calibrated for each gauge separately. Hence, the final parameter set of the model can be divided in a global and local component.

2.3.2. The Hydrological Model and Parameterization Approach

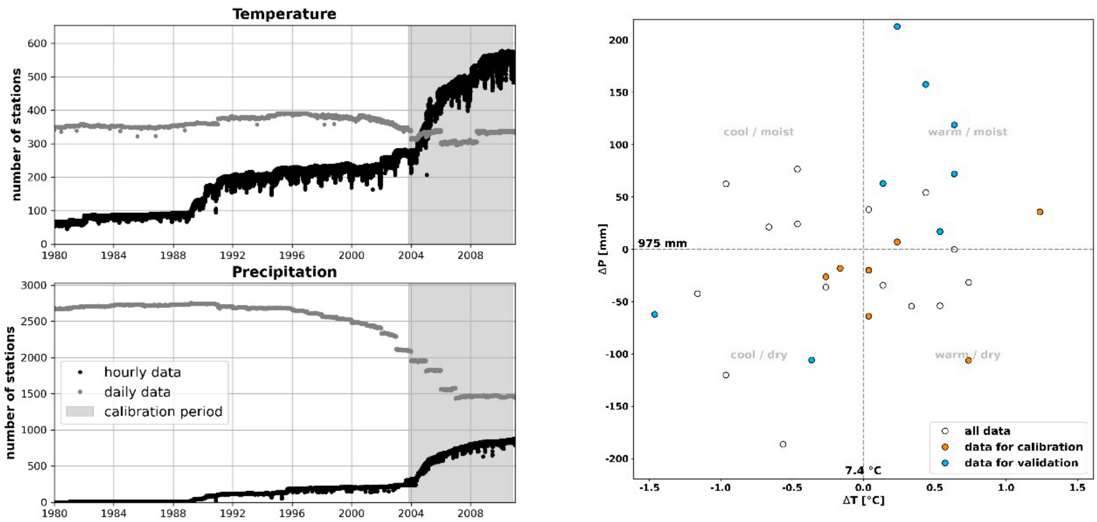
In this study the water balance simulation model WaSiM [10] was applied to perform the hydrological simulations. This model is characterized as being physically based, fully distributed, and deterministic, operating on constant time steps ranging from minutes to days. WaSiM is parallelized for distributed memory architectures using the message passing interface (MPI) and shows a good scalability on modern high-performance clusters. This enables the model to simulate physical processes with a high degree of accuracy using complex methods for the representation of hydrological processes (see Table 2). It also enables high spatial and temporal resolutions for this study. For this study, the hydrological model runs have been performed on the high-performance hardware on the Leibniz Supercomputing Centre (LRZ).

Table 2. Applied processes and their associated approaches within the WaSiM setup.

WaSiM Module	Approach	Source (If Available)
Evapotranspiration	Penman-Monteith	[35]
Snowmelt	Enhanced energy balance with snow redistribution by wind and gravitation	[36]
Soil water movement	Richards equation	[37]
Groundwater movement	Darcy equation	
Soil parameterization	Van Genuchten parameter	[27]

For the parameterization of the model a split sample approach was used for calibration and validation. Therefore, two periods comprising the hydrological years (1 November to 31 October) from 2004 to 2010 and 1995 to 2002 were selected from the reference period from 1981 to 2010 for calibration and validation, respectively. The calibration period was set at the end of the reference period, since here a larger availability of observed sub-daily meteorological records can be found (Figure 3a). This should minimize errors in the simulation due to possible changes in the temporal distribution imposed by the temporal disaggregation of observations. The calibration period (Figure 3b, orange dots) covers rather average or below average years in terms of annual precipitation and air temperature with regard to the entire reference period (1981–2010). Only two years exhibit moister conditions than the mean of the reference period. Contrary, the validation period (Figure 3b, blue dots) covers mainly above average years (six years), with three of the six years showing a mean deviation greater than 100 mm. Hence, the results of the hydrological simulations for the validation period as well as for the entire reference period should indicate whether the parameters received by the calibration are suitable for the simulation of above average moist conditions. Whether the parameters are time invariant regarding future periods cannot be decided based on the presented data. However, it should be noted that the choice of reference period might affect climate change signals [38].

The calibration was performed in parts by manual alternation of parameters for spatially distributed processes (i.e., parameters regarding evapotranspiration, infiltration rate, groundwater fluxes, snowmelt, and glacier dynamics), while four free parameters directly affecting the shape of the discharge (i.e., drainage density, recession factors for direct and interflow, snow fraction factor) were automatically estimated for each of the 98 individual catchments separately. Here, the adjusted dynamically dimensioned search (DDS) algorithm after Tolson and Shoemaker [39] was used for the automatic parameter estimation. This approach can be described as a simple stochastic and adaptive algorithm and is supposed to provide robust parameter sets if the number of iterations or evaluations is small as it adapts its search strategy according to the predefined number of iterations [40–42]. Preliminary tests have shown that an evaluation budget of 200 is sufficient for finding appropriate combinations of the four free parameters. Since the applied model was complex and computationally demanding, this resulted in run times of approximately 48 h per run. The adjusted algorithm applies a simulated annealing (SA) approach to diminish the limits of variation for the parameters to be calibrated with increasing iterations. Since the algorithm starts its search globally and becomes more and more local with increasing iterations [41], the SA further enhances the transition from a global to a local search. This approach is further referred to as DDS-SA.



(**a**) (**b**)

Figure 3. Selection of a calibration period based on meteorological data. (**a**) Number of stations available for spatial interpolation and temporal disaggregation for air temperature and precipitation. Black dots represent stations with hourly records, gray dots represent daily measurements. (**b**) Spatially averaged annual precipitation and temperature anomalies w.r.t 1981–2010 for the Hydrological Bavaria. Orange dots represent years selected for calibration.

In this study, several different objective measures (OM) were applied to evaluate the capability of the model to reproduce observed discharge values at the 98 gauges. The OM include the Nash and Sutcliffe Efficiency and its logarithmic form (NSE; logNSE, [43]), the Kling Gupta Efficiency (KGE, [44]), and the ratio of the root mean square error to standard deviation of measured data (RSR, [45]). The respective equations are presented in the Supplementary Materials. The choice of these efficiency measures is based on their respective focus on the evaluation of observed and simulated discharge (NSE: high flows; KGE: shape and bias; logNSE: low flows; RSR: volume error) [45,46]. Since the automatic calibration requires a single value to be optimized (here minimized), the efficiency criteria were combined to an overall objective measure ($OM_{overall}$) using different weights according to their contribution to calibrate the model towards high flows (see Equation (1)). A perfect fit would receive a value of zero.

$$OM_{overall} = 0.5 \times (1 - NSE) + 0.25 \times (1 - KGE) + 0.15 \times (1 - logNSE) + 0.1 \times RSR \quad (1)$$

In order to avoid downstream propagation of simulation errors, observed discharges were used as external input for individual downstream catchments during the automatized calibration.

2.3.3. Evaluation of Simulated Return Levels

In the field of flood frequency analysis (FFA) there are two different approaches to estimate return levels: the annual maximum (AM) method relying solely on the yearly peak flow, and the peak over threshold method (POT, also known as partial duration series approach) that considers all discharge values exceeding a predefined threshold [47–50]. While the AM method is still used as a standard in many countries, its application is often criticized due to a loss of information by only employing the annual maximum peak flows of a complete time series [47,50]. Hence, low peak flow values, which are not considered an actual flood, are used to fit an extreme value distribution function (EVDF).

Since historical records of stream flow are often incomplete or short, the AM method does not provide a sufficient sample size. The POT approach offers more flexibility in the representation of a flood by including all values above a chosen threshold into consideration yielding a larger sample size for an EVDF fit [47–50]. Therefore, in this study we opt for the POT method. According to [51,52] a series of samples of independent and identically distributed (iid) data which exceed a high enough threshold follow a Generalized Pareto Distribution (GPD) [48,50,53,54]. This theorem emphasizes two major requirements for the POT approach: independence of events (i.e., independent clusters of excess values) and the careful selection of a threshold. Although a plethora of methods for the estimation of a suitable threshold (e.g., graphical methods, analytical methods, mixture models) [48,50,55,56] exists to date, the selection remains subjective to an extent where several values may lead to similar results [49]. The independence criteria is described by a minimum time span between events and determined by the nature of the underlying physical process [49,50,54].

In this study we use a common procedure for all observed and simulated data as it fits the scope of a holistic modelling and evaluation approach. Nevertheless, we are aware, that this might lead to deviations from an optimal fit of the GPD at some gauges as the choice of threshold affects the bias and/or variance of the fit [48,53], and that the inherent relation between discharge and return levels for a certain gauging period may be disturbed [47]. At first, a series of independent events is created by de-clustering the observations using a physically based threshold u_p [54] which is defined by the average number of events per year λ_T [49] (see Equation (2)). Here, we opt for a value $\lambda_T = 8$.

$$u_p = 1 - \left(\frac{\lambda_T}{365.25}\right) \tag{2}$$

The events are further separated by an intermittence time of 5 days following [48,49,57,58]. The resulting series of separated events (clusters) form the basis for the GPD fit. A statistical threshold u_s [54] equal to the 90th highest value of the clustered data is applied for the model fit to ensure that only values of an actual flood event are included for the estimation. Thus, the requirement of the GPD for a sufficiently high threshold is met. The GPD parameters (shape and scale) were estimated using the method of L-moments (LM) [59] as it shows good performance for small and moderate sample sizes [60,61]. Figure 4 illustrates the different steps of our workflow to derive return levels from the complete time series of discharge.

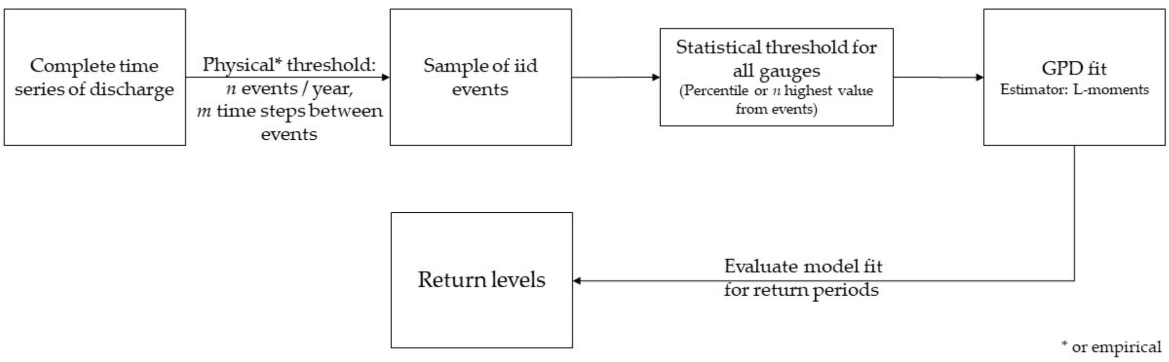


Figure 4. Scheme of the applied approach to estimate return levels for the 98 gauges.

The applied thresholds u_p (λ_T, respectively) and u_s are seemingly large compared to other studies [47,49,54,57,62,63]. However, a sensitivity analysis using several thresholds for the evaluation of the GPD model fit showed the best results for this combination across all gauges (see Supplementary Materials).

Further, we use the GPD model fit to estimate the return levels of five, ten, and 20-year return periods for both observations and simulated data, respectively. The 20-year return period is deemed as an appropriate level for comparison, since for 90% of the gauges at least once the associated 20-year return level occurred within the observational record, considering a 10% error for observed maximum flows. Choosing a return period beyond 20-years means relying on an extrapolation for most gauges (e.g., more than 50% percent of gauges at a 50-year return period). A level of trust (LOT) is then calculated by a simple comparison after Equation (3) using the estimates of the return periods, where HF_x represents the high flow values of return interval x, for the simulated sim data and observations *obs*.

$$\mathrm{LOT} = \mathrm{abs}\left(\frac{\mathrm{HF}_{x_{\mathrm{sim}}} - \mathrm{HF}_{x_{\mathrm{obs}}}}{\mathrm{HF}_{x_{\mathrm{obs}}}}\right) \times 100 \quad (3)$$

The LOT is subdivided into the qualitative categories very high (LOT ≤ 10%), high (10% < LOT ≤ 20%), moderate (20% < LOT ≤ 30%), low (30% < LOT < 50%), and very low (LOT ≥ 50%). These categories are considered to be feasible for the comparison of high flows and return levels due to the uncertainties immanent to high flow observations which are imposed by the calculation of discharge using the water level to discharge relationships established for the individual gauges. These relationships might change over time (e.g., with changes in the riverbed at the gauge) and are further limited by a maximum water level which might be exceeded by severe floods (overflowing).

3. Results

3.1. Results from the Holistic Modelling

Although multiple performance criteria were used during the calibration, the focus in this section is on the NSE and KGE. These two metrics provide a thorough overview of the quality of the model. In addition, the NSE provides a first hint on the model's capability to reproduce observed high flows, since it is sensitive to deviations in high values of a time series. The KGE is less sensitive to these high value deviations, and hence often indicates a better performance than the NSE.

The results for the periods of model calibration, validation, and the reference period for all gauges of the Hydrological Bavaria are shown in Figure 5 (top). Overall, the model shows sufficient values in both criteria.

Regarding the NSE, the model calibration yields sufficient representations of observed discharge for more than 75% of the gauges (values above 0.5). Furthermore, the holistic model setup results in a very good fit for almost half of the gauges of the Hydrological Bavaria. However, for some gauges, the proposed approach leads to an insufficient model performance with values below zero. The KGE indicates a better overall model fit than the NSE. 75% of the gauges exhibit a very good performance according to the KGE. Only two gauges show values below 0.5. Furthermore, a smaller variance in performance for the KGE than for the NSE indicates that the model is more robust regarding the overall regime than discharge extremes.

The validation of the model also illustrates a sufficient model performance with respect to the NSE, as more than 75% of the gauges of the Hydrological Bavaria exhibit values greater than 0.5. Compared to the calibration period, the model overall represents the observations better with more than 50% of gauges showing values above 0.7. However, the simulations show insufficient NSE values (NSE < 0) for three gauges. The KGE shows an overall better fit of the model to the observations as well. A significantly worse performance is also recognizable for two gauges for the KGE and follows the development of NSE for the validation period.

The model performance of the reference period is particularly interesting for the subsequent estimation of return levels based on the simulations and the comparison with return levels derived from observations. The simulation results for this period indicate an overall slight decrease in model performance for the NSE and KGE, respectively. However, more than 75% of all gauges exhibit sufficient values for both objective measures. Almost 50% of the simulated discharges exhibit a very good model performance in terms of NSE and almost 75% in terms of KGE. As for the validation period, the simulation results for three gauges yields insufficient model fit regarding the NSE.

Figure 4 further illustrates the model performance for all gauges separated by their location within the greater landscape.

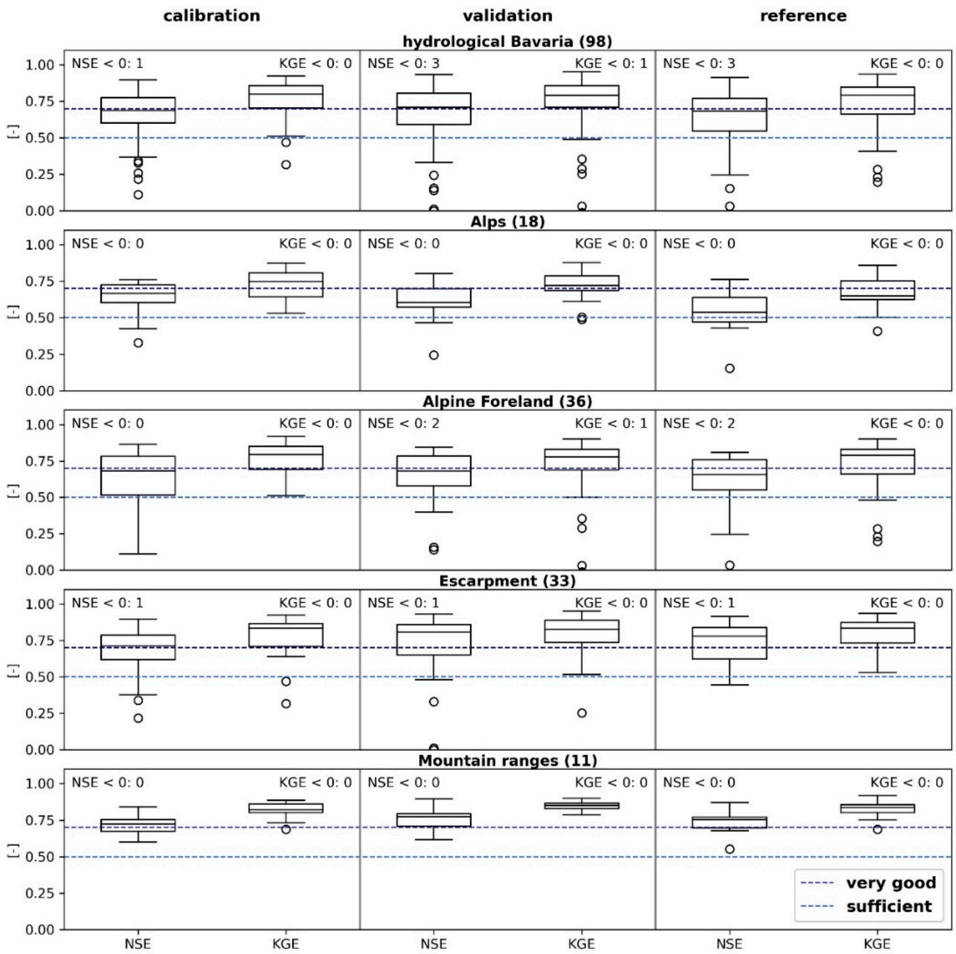


Figure 5. WaSiM model performance (NSE and KGE) for the three different periods (calibration, validation, reference period) for gauges of the Hydrological Bavaria and further subdivided by the four different landscapes (Alps, Alpine foreland, escarpment, mountain ranges). The number of gauges in each subregion is marked in brackets. The dashed lines indicate margins of a sufficient (light blue) and very good (dark blue) model fit. The values at the top of each frame indicate the number of gauges where the model performance of NSE or KGE is below zero.

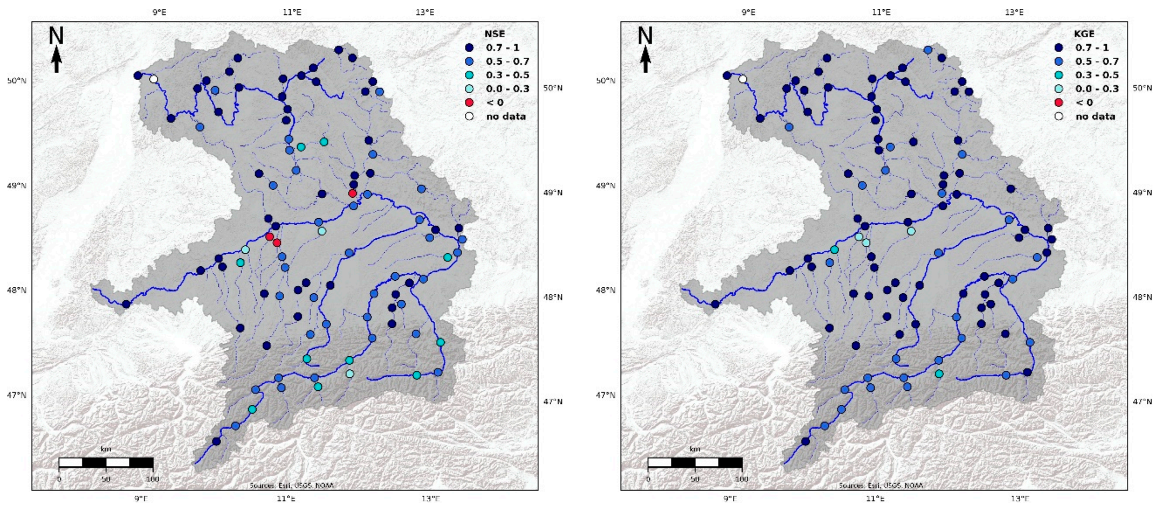
The gauges within the Alps show a sufficient performance at most gauges for the calibration and validation period. Here, the model performance of the validation period is worse than for the calibration. The reference period exhibits the worst values, particularly for the NSE where the model yields sufficient results for slightly more than 50% of the gauges. Again, the variance in the criteria increases from calibration to the reference simulation. This in parts unsatisfactory model performance for the Alps may be ascribed to three major factors. First, meteorological observation is the Alps are scarce and most gauges are located in the valleys rather than in high altitudes with steep slopes or even glaciers. Hence, there is an undercatch in the precipitation measurements [22,34,64,65], which influences runoff generation. Furthermore, since the model also employs a dynamic glacier routine to simulate glacier runoff, the globally defined parameters may not be suitable for every glacier within the study area. At last, the discharge within the Alps is severely influenced by water management infrastructure (e.g., hydro power generation, detention basins) and operation data of these facilities is usually not available. In this study, only the most dominant structures were implemented.

The model overall performs better for gauges within the Alpine foreland. However, two gauges show an insufficient model fit for the validation and reference period regarding the NSE. These gauges exhibit a poorly distinctive annual runoff regime compared to the remaining gauges of the Hydrological Bavaria. Hence, some unique processes leading to these regimes might not be reproduced by the holistic modelling approach. Apart from those two gauges, more than 75% of the gauges show at least a sufficient agreement between simulations and observations. Furthermore, the rivers of the Alpine foreland are affected by water management structures as well. In many cases, the modes of operation are estimated and stationary with time as explicit operation data are again not available.

For gauges situated in the escarpment, the model performance shows better results compared to the Alps and the Alpine foreland throughout all simulation periods. The NSE and KGE are sufficient at most gauges with only a few outliers. These underperforming simulation results are either induced by a lack of information about the fuzzy control mechanisms of the Main-Danube transfer system or an insufficient parameterization of karstic soils of the Franconian Jura. Along the Main there are several hydro power plants as well. However, since these structures are run-of-river power plants, they have less impact on the natural runoff regime than the structures in the Alps and Alpine foreland. Hence, the model yields better representations of observed runoff for the Main.

Gauges situated in the eastern mountain ranges exhibit the best model performance among all gauges. Across all periods, the NSE and KGE never fall below 0.5. Furthermore, the variance of these values is rather small compared to those of the gauges within other landscapes. An explanation for the better performance is, that almost all gauges are unaffected by water management and thus considered as natural.

Figure 6 shows the spatial distribution of the NSE and KGE (adapted from Poschlod et al. [22]) for the reference period. Despite the efforts taken to reduce forward propagation of model errors from upstream to downstream gauges (i.e., use of observed discharge as input for the automatic calibration of a downstream sub-catchment), the gauges along the Danube show a decrease in performance from its source to the outlet. The model performance in most cases is better for headwater catchments. However, the figure illustrates an overall sufficient performance for most of the gauges.



(**a**) (**b**)

Figure 6. Maps of the WaSiM model performance for the 98 gauges ((**a**) NSE, (**b**) KGE) for the reference period (1981–2010). Blueish colors indicate values above zero with better performance indicated by darker colors. Red dots represent gauges with values below zero and white dots gauges with no observations available for the reference period.

The two illustrated model criterions exhibit–with the exceptions mentioned–only little variation for the respective modelling period. Hence, the model parameters are to some extent robust towards changes in land use, water management and other non-stationary conditions affecting the runoff formation. However, the presented reference period is rather short and thus, these changes might be moderate.

An indication for the applicability of the holistic model setup to estimate return levels is provided by the relative deviations between the observed and simulated mean high flows (MHF, i.e., average of annual maximum peaks of discharge over a given period, here 30 years of the reference period) illustrated in Figure 7. The deviations of MHF between simulations and observations are diverse within the entire Hydrological Bavaria. However, there are regions that exhibit considerable larger deviations than others. Simulated MHF for gauges situated in the Alps tend to underestimate the MHF derived from observations with values below −20% and up to −10% difference. This underestimation in MHF is ascribed to the underrepresentation of precipitation in the meteorological data that results in a smaller storage of snow and ice during the winter and thus a reduced discharge from snow and ice melt together with precipitation during spring and summer. For gauges in the vicinity of the Alps and south of the Danube, there are only few gauges that depict a larger deviation in MHF. For the majority of these gauges, the deviation is between −10% and +10%. However, along the Danube and the Isar, the differences in MHF become larger with distance from the source and with the increasing number of tributaries contributing to the Danube's discharge. These deviations originate to some extent from the automatized calibration approach using observed discharge of tributaries as input for downstream catchments. Since the discharge of a downstream catchment is to a certain extent governed by the discharge of its tributaries (stronger dependency with larger upstream catchments), this approach leads to very good model fits during the calibration regardless of the automatized algorithm efforts to find an optimal parameter set. On the other hand, some tributaries to the Danube and Inn exhibit a stronger underestimation of simulated MHF (<−20%). In these cases, the model in general performs rather poorly leading to these deviations. For the gauges of the escarpment we generally find an underestimation along the upper Main, an overestimation along the lower Main, and small deviations for the gauges of the Regnitz river system. For the mountain ranges the model performs sufficiently well for most gauges.

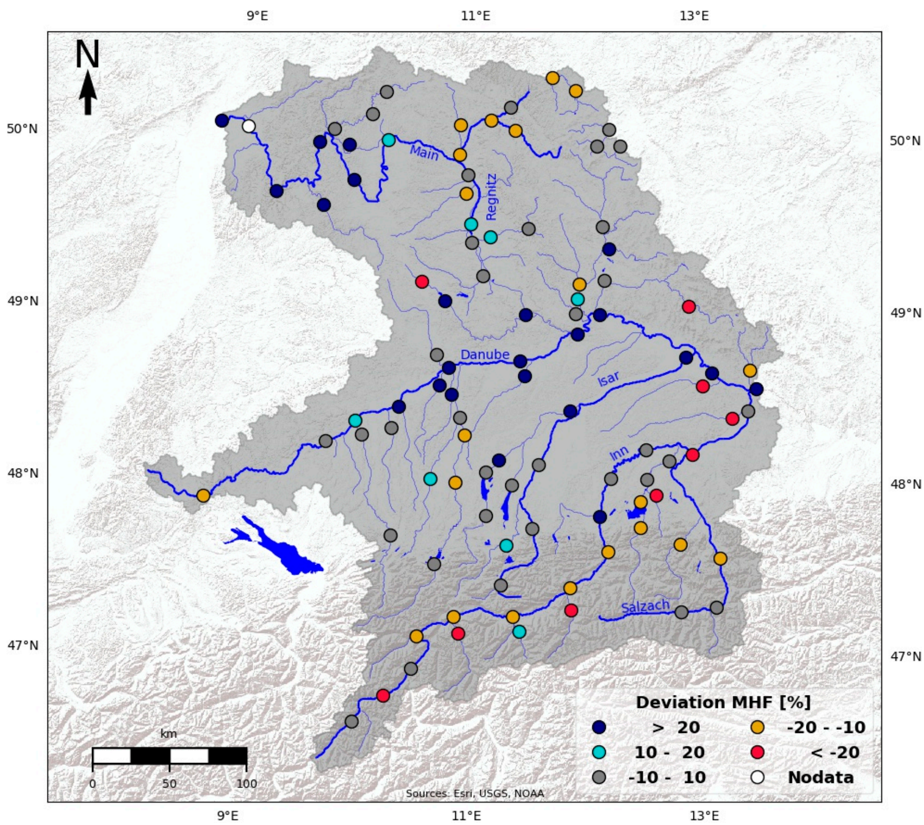


Figure 7. Relative deviations between observed and simulated mean high flows (MHF) for the 98 gauges of the Hydrological Bavaria.

3.2. Qualitiy Assessment of the Representation of Return Levels

The quality of the estimates of return levels, based on the model results and observations in the reference period is illustrated by LOT. The LOT is analyzed for the return periods of five, ten, and 20 years. Higher return periods are not considered here, as their estimate would be a result of an extrapolation of the GPD fit. The high flow values of the prior mentioned return periods are further referred to as HF5, HF10, and HF20, respectively.

The LOT of the HF5 for all gauges are shown in Figure 8. Most gauges of the Danube basin exhibit a moderate to very high LOT. Exceptions with little to no trust in absolute values of the simulations are found for eight gauges, most of them situated at the Danube or its tributaries. Gauges within the Alps in most cases show a moderate to very high LOT as well, whereas only two values indicate a low confidence. The Main and Elbe basins depict a moderate to very high confidence in the representation of the HF5. Only four gauges downstream the Main show a low to very low LOT as well as one tributary to the Regnitz.

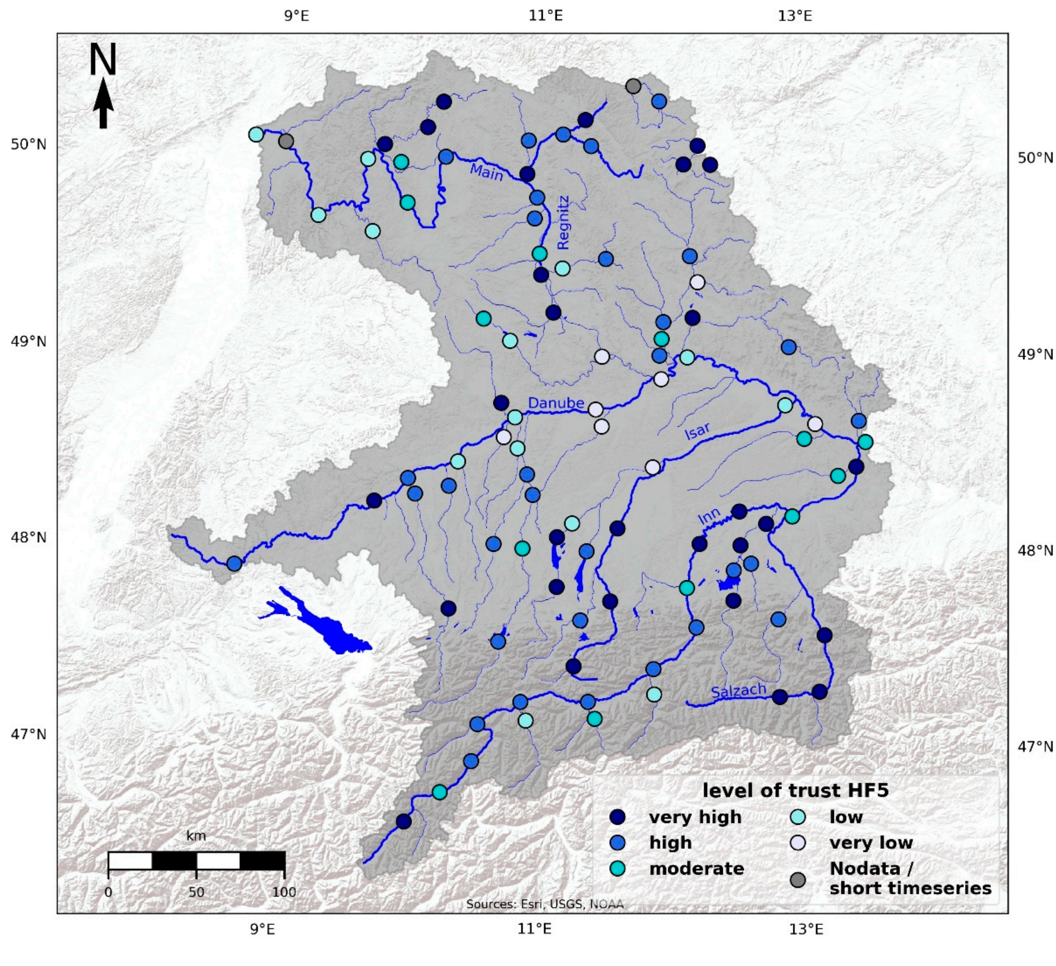


Figure 8. The level of trust (LOT) for high flows with a 5-year return period (HF5). Gray dots indicate gauges with no data or short timeseries. Qualitative categories: very high (LOT ≤ 10%), high (10% < LOT ≤ 20%), moderate (20% < LOT ≤ 30%), low (30% < LOT < 50%), and very low (LOT ≥ 50%).

The LOT for high flow events with a return period of 10 years (HF10) is given in Figure 9. Here, a moderate to high LOT may be declared for gauges within the Danube basin. As it was the case for the HF5, the model does not allow for statements about actual values of the HF10 for some gauges of the Danube and its southern confluxes since the LOT is low or worse. The gauges of the basins of the Main and Elbe depict a decent confidence in HF10 values derived from simulations (moderate to very high LOT). Just a few gauges show a low confidence level. Compared to the HF5 the LOT diminishes for some gauges within the Alps, but still shows moderate to very high confidence for most gauges except for two smaller Austrian tributaries to the Inn.

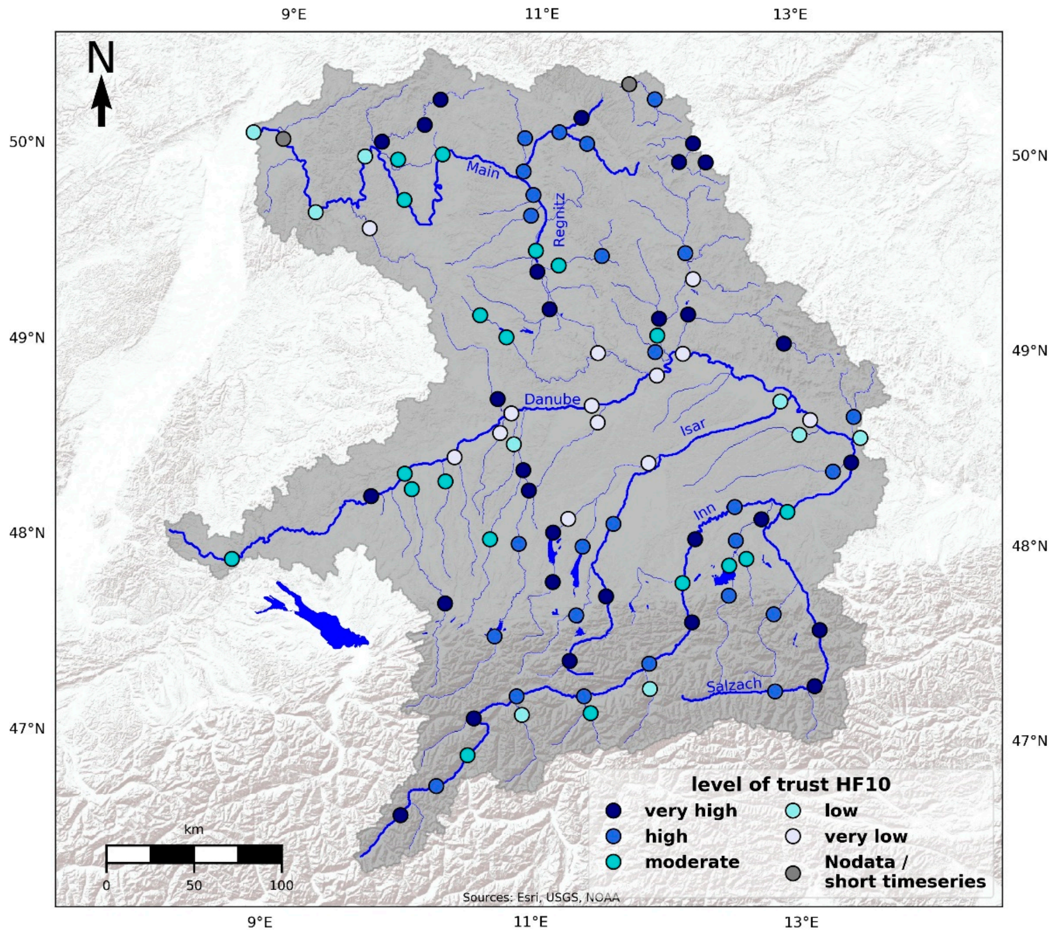


Figure 9. The level of trust (LOT) for high flows with a 10-year return period (HF10). Gray dots indicate gauges with no data or short timeseries. Qualitative categories: very high (LOT ≤ 10%), high (10% < LOT ≤ 20%), moderate (20% < LOT ≤ 30%), low (30% < LOT < 50%), and very low (LOT ≥ 50%).

Figure 10 shows the LOT for high flows with a 20-year return period (HF20). Compared to the HF5 and HF10, the overall confidence in simulated HF20 values decreases slightly. However, the confidence in the simulations is still moderate to very high for most gauges of the Danube with exceptions already noted for the HF5 and HF10. Hence, the simulations for these gauges are not suitable for the estimation of actual values of recurrence periods of any annuality. The simulations for the Main and Elbe basins still yield moderate to very high confidence in the representation of the HF20 as well. There are only minor shifts in confidence for the individual gauges and those showing a low or very low LOT for HF5 and HF10 remain within this category for the HF20, too. The model results for alpine gauges show a sustained moderate to very high confidence in the estimates of HF20. However, there is a shift from high to low confidence for one gauge of the Salzach catchment, indicating a divergence of the GPD between observations and model data for higher annualities.

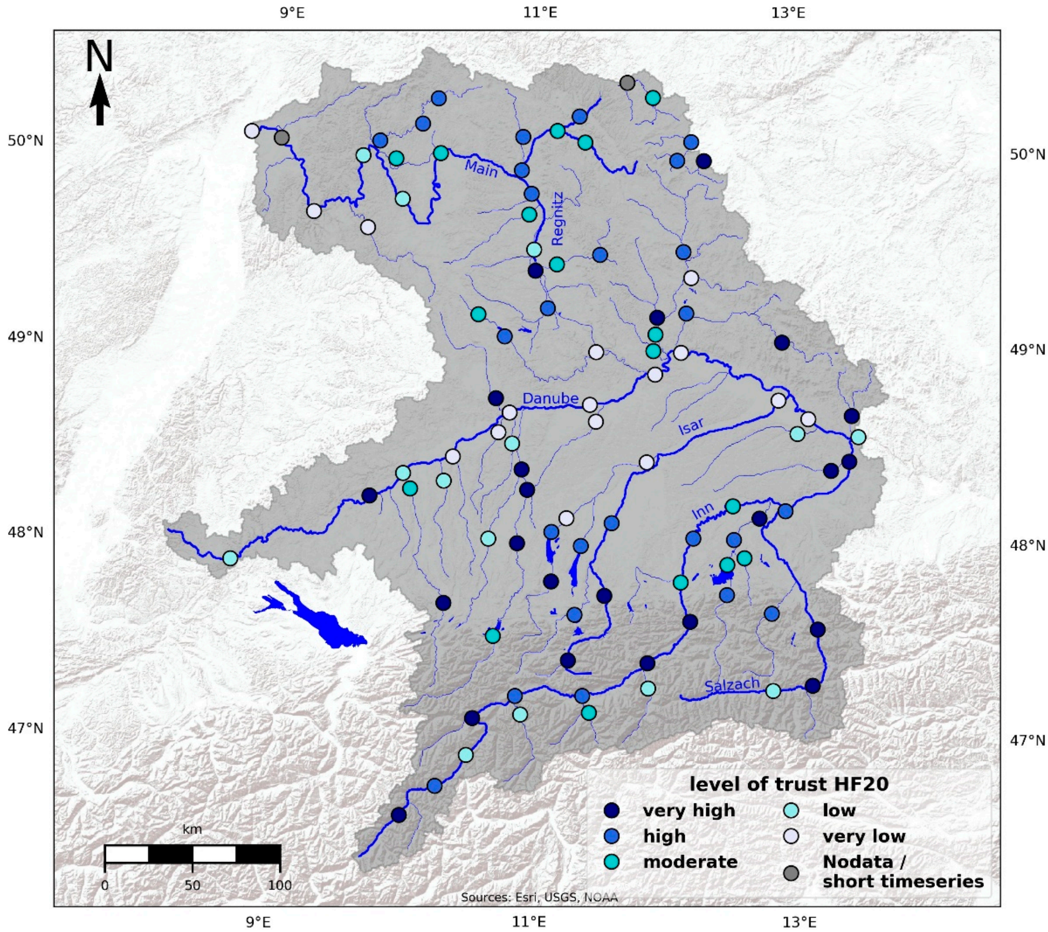


Figure 10. The level of trust (LOT) for high flows with a 20-year return period (HF20). Gray dots indicate gauges with no data or short timeseries. Qualitative categories: very high (LOT ≤ 10%), high (10% < LOT ≤ 20%), moderate (20% < LOT ≤ 30%), low (30% < LOT < 50%), and very low (LOT ≥ 50%).

In some cases, the LOT does not show a distinct pattern of decrease or increase for individual gauges with higher annualities, but instead gains or loses confidence independently from the annuality. Hence, the GPD features a distinct difference in shape between observations and modelled discharge for these gauges.

4. Discussion

4.1. On the Holistic Model Approach

The holistic modelling approach applied in this study to simulate the discharges for 98 catchments allows for a sufficient representation of the observations with regard to the presented objective measures of the NSE and KGE. Furthermore, the model provides sufficient results over different runoff regimes, including glacio-nival to nival regimes in the Alps, pluvio-nival in the Pre-Alps, nivo-pluvial in the Alpine foreland, and pluvial regimes of the Main and its tributaries [22]. Since the results are similar throughout the simulated periods for most of the individual gauges, the model is considered robust (i.e., unlikely over-parameterization [19]) towards changes in land cover and use, water management, and other processes affecting runoff formation. However, there are catchments with poor performance in every simulation period. Here, the model may not be able to represent the observations using parameters that are considered valid for various regions of the Hydrological Bavaria. This loss of performance when applying regionalized parameters is reported in several studies (e.g., [14,18,19,44]).

These studies seek to find optimal regionalized or globally applicable parameters by analyzing catchments which are either not affected by water management or other kinds of anthropogenic alteration, or the implemented structures only have minor effects on runoff formation. Hence, the parameters reflect the physical processes of these catchments. However, the catchments of the Hydrological Bavaria are severely influenced by different water management structures (e.g., detention basins, hydropower plants) or governed by the discharge of natural lakes. For the political Bavaria, the Bavarian Environment Agency (LfU) lists more than 4200 hydropower plants, 25 reservoirs or detention basins governed by the Bavarian state, and 150 larger natural lakes [66]. Since not every structure or natural lake affects the discharge of the catchments presented in this study, only the most important were included in the modeling approach. These structures might have a certain impact on the global definition of parameters. This is especially the case for the Main-Danube transfer system which is a complex structure comprising the transfer channel as well as the transfer of water between three artificial lakes which are additionally used for low flow elevation.

The holistic modelling approach might further benefit from an improved data basis for soils and hydrogeology as this information is crucial for the hydrological simulation with WaSiM. Since the data provided by the ESDB comprises only eleven texture classes (five for top- and sub-soils each and one organic) [27] for European soils, the parameterization of the individual soil types is limited to these classes. To avoid larger uncertainties due to over-parameterization, these parameters were not adjusted for most catchments of the Hydrological Bavaria. However, for three catchments the calibration indicated that the parameter values for a dominating soil type were responsible for an insufficient performance. By comparing the grain size distribution of the USDA metric [67] with the common soil texture metric for Germany [68] these EDSB parameters have been replaced.

Furthermore, the hydrological model itself could be improved to further enhance the overall performance of the holistic modelling approach. WaSiM employs a single selected routine for the simulation of snowfall and -melt, as well as glacier development. For this study we opted for a method after Warscher et al. [36] which includes a high degree of physical processes (i.e., radiation-based melt, distribution of snow by gravity and wind). This approach was developed for high Alpine terrain and thus might not yield any benefit in regions outside the Alps. Hence, the hydrological model could be adapted to employ several different snow melt approaches suitable for different landscapes.

Usually, automatized calibration approaches like SCE-UA or DDS are used to facilitate the parameterization of a model [14,18,19,39,69–71]. However, in most cases these algorithms are applied to single headwater catchments. The automatic calibration approach for downstream catchments applied in this study using observed discharge as inflows seems to affect the performance of gauges with further distance from its headwater. The intension of this practice was to avoid the propagation of modelling errors. Since the discharge at gauges defining larger catchments is mostly governed by its tributaries, the objective metrics during the calibration using the DDS-SA approach were close to optimal by definition. Hence, the parameters gained by the automatized calibration for downstream catchments do not reflect the physical processes of that sub-catchment but rather an attempt to adjust the cumulative discharge of its tributaries.

Nonetheless, a poor performance at individual gauges is overall outweighed by a gain in spatial consistency, reduction of equifinality by avoiding over-parameterization, and the provisioning of a holistic simulation approach [14].

4.2. On the LOT

The four performance criteria used in this study were selected due to their focus on different regions of the statistical distribution of discharge values according to [44–46]. Furthermore, for the automatic calibration the NSE and KGE were emphasized by assigning higher weights. Since the model was targeted at representing the high flows, we have assigned the highest weights to the NSE. The NSE is known to be more sensitive for deviations in high flows [46] due to its sensitivity to deviations between higher discharge values, received the highest weight. However, Mizukami et al. [72] state that a parameterization based on the KGE leads to a better simulation of high flow values as it considers the mean and variance. The performance of a model with respect to high flows might further increase when weights are applied to the compartments of the KGE (correlation coefficient, bias of mean and variance). Hence, the selected weights for the automatic calibration might be misleading for the parameterization of the model towards a sufficient representation of high flows. On the other hand, Knoben et al. [73] point out that the NSE and the KGE should not be interpreted the same way as the "KGE does not have an inherent benchmark against which flows are compared". For the NSE this benchmark is represented by the mean of the observations, whereas for the KGE the benchmark must be specified by the user. Therefore, the weights of the overall calibration metric can be considered reasonable.

The LOT of return periods strongly depends on the model's capability to sufficiently simulate high flows. Furthermore, it also relies on the approach selected for the estimation of EVD. In this study we opt for the GPD in conjunction with the POT to estimate for the reference period from 1981 to 2010. The POT method is supposed to yield a better EVD than when applying AM, especially if the time series is short (i.e., less than 14 years) [62]. However, an estimation of higher return periods requires a larger sample size that cannot be provided by AM in most cases. Since the reference period comprises 30 years, the sample of 30 annual peaks is still not sufficient for the estimation of longer return periods. Therefore, the POT method was chosen to form a sufficient sample size for the EVD. The GPD is sensitive to the threshold defining the sample size as well as to the statistical threshold defining the value when an event is considered extreme [47–49,54]. Thus, an individual threshold for each gauge of the Hydrological Bavaria might result in an increase in confidence in the presented estimated return periods. Nevertheless, these thresholds derived by either graphical or analytical methods are still prone to the subjective choice by the user as there is no optimal value [48,49,53]. Furthermore, an individual threshold might still not increase the LOT for the worst performing gauges.

A low LOT implies that the model is not applicable to estimate actual values of return levels. Here, the holistic modelling approach does not provide parameters that result in a sufficient representation of high flows. This behavior is also reported by Ricard et al. [14] when applying regionalized parameters to model various catchments.

Further improvements for the presented hydrological model comprise the adjustment of the undercatch of precipitation in the Alps and the adjustment of operating rules for the implemented reservoirs and lakes. These features affect the runoff formation and thus the simulation of high flows.

5. Conclusions

A first holistic modelling approach intended to develop a robust toolset to support water resources management under climate change conditions with robust estimates of return levels was introduced. The approach comprises a semi-global and semi-automatized calibration for the 98 catchments of the Hydrological Bavaria. The model applies a regionalized set of parameters that exhibit satisfactory results with regard to the presented evaluation metrics and simulation periods, with the values of the NSE and KGE exceeding 0.6 for the majority of the 98 gauges. For many gauges, the holistic modelling approach lead to satisfactory results for the MHF where relative deviations between observed and simulated values were kept at around ±10%. However, there are gauges where deviations are larger in MHF and for various return levels (>20%), thus creating lower confidence. Deviations of such magnitude reduce the trustworthiness of model performance; in consequence, a low level of confidence should be markedly illustrated when estimating changes in absolute values of peak flow return levels. Most of these poor performances can be attributed to deficiencies in the provided input data (e.g., lack of operational rulesets for water management structures, soil data, meteorological data) rather than the holistic modelling approach.

Even though individual adjustments could be made to improve the performance of single catchments, the overall benefits gained from a holistic modelling approach outweighs the shortcomings of a few gauges. By using a holistic modelling approach, it is possible to model larger catchments without the need to sacrifice model complexity.

For many applications, CCI studies are interested in the course of future trends compared to a certain reference period rather than in absolute values [18]. Changes in streamflow are a response to changes in climate conditions (or changes in land use or water management). Hence, we consider the model to perform sufficiently well to derive relative changes in the frequency and intensity from simulations.

For future applications of the model, there is potential to further improve its performance. One would be to increase the volume of precipitation over the Alps in the meteorological reference SDCLIREF to reduce the underestimation of runoff in this region. The model could also benefit from an increase of the level of detail of homogenized soil information, and a better representation of humanmade hydraulic/hydrologic structures.

Supplementary Materials: The following refer to [43–45,49] and are available online at http://www.mdpi.com/2073-4441/12/9/2349/s1, Equations (S1a) and (S1b): The Nash & Sutcliffe efficiency (a) and its logarithmic form (b), Equations (S2a), (S2b), and (S2c): The Kling-Gupta Efficiency (a) and its compartments (b) and (c). Equation (S3): the ratio of root mean square error to standard deviation of observations, Equation (S4): Normalization of return levels of discharge, Figure S1: Mean and standard deviation for standardized return levels of the HF5 over all 98 gauges, Figure S2: Mean and standard deviation for standardized return levels of the HF10 over all 98 gauges, Figure S3: Mean and standard deviation for standardized return levels of the HF20 over all 98 gauges.

Author Contributions: Conceptualization, F.W.; methodology, F.W., B.P.; software, F.W., R.R.W. and J.W.; validation, F.W. and F.v.T.; formal analysis, F.W.; investigation, F.W.; resources, R.L.; data curation, F.W., R.R.W. and F.v.T.; writing—original draft preparation, F.W.; writing—review and editing, R.R.W., F.v.T., B.P., J.W. and R.L.; visualization, F.W.; supervision, R.L.; project administration, R.L.; funding acquisition, R.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Bavarian State Ministry for the Environment and Consumer Protection (StMUV). Computations with WaSiM-ETH for the ClimEx project were conducted on the SuperMUC supercomputer of the Leibniz Supercomputing Centre (LRZ) of the Bavarian Academy of Sciences and Humanities. The operation of this supercomputer was granted and funded via the Gauss Centre for Supercomputing (GCS) by the German Federal Ministry of Education and Research and the Bavarian State Ministry of Education, Science, and the Arts.

Acknowledgments: The authors thank Gilbert Brietzke and André Kurzmann from the LRZ Garching for their technical support and advice using the LRZ's supercomputing facilities. Furthermore, the authors would like to acknowledge the LRZ for the provisioning of extensive computation resources to perform the hydrological modelling. We also thank Jörg Schulla for his efforts in improving the hydrological model WaSiM and his advice of its application. Also, we thank Holger Komischke (LfU, StMUV) for his support regarding data acquisition.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Collins, M.; Knutti, R.; Arblaster, J.; Dufresne, J.-L.; Fichefet, T.; Friedlingstein, P.; Gao, X.; Gutowski, W.J.; Johns, T.; Krinner, G.; et al. Long-term climate change: Projections, commitments and irreversibility. In *Climate Change 2013-The Physical Science Basis: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., Eds.; Cambridge University Press: Cambridge, UK, 2013; pp. 1029–1136, ISBN 9781107057991.
2. IPCC. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Pachauri, R., Meyer, L., Eds.; IPCC: Geneva, Switzerland, 2014; p. 151.
3. Dooge, J.C.I. Looking for hydrologic laws. *Water Resour. Res.* **1986**, *22*, 46–58. [CrossRef]
4. Leduc, M.; Mailhot, A.; Frigon, A.; Martel, J.-L.; Ludwig, R.; Brietzke, G.B.; Giguère, M.; Brissette, F.; Turcotte, R.; Braun, M. The ClimEx Project: A 50-member ensemble of climate change projections at 12-km resolution over Europe and Northeastern North America with the Canadian regional climate model (CRCM5). *J. Appl. Meteorol. Climatol.* **2019**, *58*, 663–693. [CrossRef]
5. Van Vuuren, D.P.; Edmonds, J.; Kainuma, M.; Riahi, K.; Thomson, A.; Hibbard, K.; Hurtt, G.C.; Kram, T.; Krey, V.; Lamarque, J.-F.; et al. The representative concentration pathways: An overview. *Clim. Chang.* **2011**, *109*, 5. [CrossRef]

6. Mauser, W.; Prasch, M. (Eds.) *Regional Assessment of Global Change Impacts: The Project GLOWA-Danube*; Springer: Berlin/Heidelberg, Germany, 2015. [CrossRef]
7. Muerth, M.; Gauvin St-Denis, B.; Ricard, S.; Velázquez, J.A.; Schmid, J.; Minville, M.; Caya, D.; Chaumont, D.; Ludwig, R.; Turcotte, R. On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 1189–1204. [CrossRef]
8. Velázquez, J.A.; Schmid, J.; Ricard, S.; Muerth, M.J.; St-Denis, B.G.; Minville, M.; Chaumont, D.; Caya, D.; Ludwig, R.; Turcotte, R. An ensemble approach to assess hydrological models' contribution to uncertainties in the analysis of climate change impact on water resources. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 565–578. [CrossRef]
9. Mauser, W.; Bach, H. PROMET–Large scale distributed hydrological modelling to study the impact of climate change on the water flows of mountain watersheds. *J. Hydrol.* **2009**, *376*, 362–377. [CrossRef]
10. Schulla, J. Model Description WaSiM: (Water Balance Simulation Model). Available online: www.wasim.ch (accessed on 9 June 2020).
11. Hattermann, F.F.; Wortmann, M.; Liersch, S.; Toumi, R.; Sparks, N.; Genillard, C.; Schröter, K.; Steinhausen, M.; Gyalai-Korpos, M.; Máté, K.; et al. Simulation of flood hazard and risk in the Danube basin with the Future Danube Model. *Clim. Serv.* **2018**, *12*, 14–26. [CrossRef]
12. Krysanova, V.; Hattermann, F.; Wechsung, F. Development of the ecohydrological model SWIM for regional impact studies and vulnerability assessment. *Hydrol. Process.* **2005**, *19*, 763–783. [CrossRef]
13. Beven, K.; Freer, J. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J. Hydrol.* **2001**, *249*, 11–29. [CrossRef]
14. Ricard, S.; Bourdillon, R.; Rousse, L.D.; Turcotte, R. Global Calibration of Distributed Hydrological Models for Large-Scale Applications. *J. Hydrol. Eng.* **2013**, *18*, 719–721. [CrossRef]
15. Bárdossy, A. Calibration of hydrological model parameters for ungauged catchments. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 703–710. [CrossRef]
16. Samaniego, L.; Kumar, R.; Attinger, S. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resour. Res.* **2010**, *46*. [CrossRef]
17. Kouchi, D.H.; Esmaili, K.; Faridhosseini, A.; Sanaeinejad, S.H.; Khalili, D.; Abbaspour, K.C. Sensitivity of Calibrated Parameters and Water Resource Estimates on Different Objective Functions and Optimization Algorithms. *Water* **2017**, *9*, 384. [CrossRef]
18. Gaborit, É.; Ricard, S.; Lachance-Cloutier, S.; Anctil, F.; Turcotte, R. Comparing global and local calibration schemes from a differential split-sample test perspective. *Can. J. Earth Sci.* **2015**, *52*, 990–999. [CrossRef]
19. Merz, R.; Blöschl, G. Regionalisation of catchment model parameters. *J. Hydrol.* **2004**, *287*, 95–123. [CrossRef]
20. Xu, C.-Y.; Widén, E.; Halldin, S. Modelling hydrological consequences of climate change—Progress and challenges. *Adv. Atmos. Sci.* **2005**, *22*, 789–797. [CrossRef]
21. Blöschl, G.; Sivapalan, M. Scale issues in hydrological modelling: A review. *Hydrol. Process.* **1995**, *9*, 251–290. [CrossRef]
22. Poschlod, B.; Willkofer, F.; Ludwig, R. Impact of Climate Change on the Hydrological Regimes in Bavaria. *Water* **2020**, *12*, 1599. [CrossRef]
23. European Environment Agency. Digital Elevation Model over Europe (EU-DEM). 2013. Available online: https://www.eea.europa.eu/data-and-maps/data/eu-dem (accessed on 9 June 2020).
24. European Environment Agency. Corine Land Cover 2006 v17. 2013. Available online: https://www.eea.europa.eu/data-and-maps/data/clc-2006-raster-3 (accessed on 9 June 2020).
25. European Commission and the European Soil Bureau Network. The European Soil Database Distribution Version 2.0 EUR 19945 EN. 2004. Available online: https://esdac.jrc.ec.europa.eu/content/european-soil-database-v20-vector-and-attribute-data (accessed on 9 June 2020).
26. Panagos, P. The European soil database. *GEO Connex.* **2006**, *5*, 32–33.
27. Wösten, J.H.M.; Lilly, A.; Nemes, A.; Le Bas, C. Development and use of a database of hydraulic properties of European soils. *Geoderma* **1999**, *90*, 169–185. [CrossRef]
28. Dörhöfer, G.; Hannappel, S.; Reutter, E.; Voigt, H.-J. Die Hydrogeologische Übersichtskarte von Deutschland HÜK200. *Z. Für Angew. Geol.* **2001**, *47*, 153–159.
29. BGR; UNESCO (Eds.) *International Hydrogeological Map of Europe 1:1,500,000 (IHME1500 v1.1)*; Bundesanstalt für Geowissenschaften und Rohstoffe (BGR): Hannover, Germany; Paris, France, 2014.

30. Shamra, A.; Srikanthan, S. Continuous rainfall simulation: A nonparametric alternative. In Proceedings of the 30th Hydrology & Water Resources Symposium: Past, Present & Future, Launceston, Australia, 4–7 December 2006.
31. Pui, A.; Sharma, A.; Mehrotra, R.; Sivakumar, B.; Jeremiah, E. A comparison of alternatives for daily to sub-daily rainfall disaggregation. *J. Hydrol.* **2012**, *470*, 138–157. [CrossRef]
32. Westra, S.; Mehrotra, R.; Sharma, A.; Srikanthan, R. Continuous rainfall simulation: 1. A regionalized subdaily disaggregation approach. *Water Resour. Res.* **2012**, *48*. [CrossRef]
33. Poschlod, B.; Hodnebrog, Ø.; Wood, R.R.; Alterskjær, K.; Ludwig, R.; Myhre, G.; Sillmann, J. Comparison and Evaluation of Statistical Rainfall Disaggregation and High-Resolution Dynamical Downscaling over Complex Terrain. *J. Hydrometeorol.* **2018**, *19*, 1973–1982. [CrossRef]
34. Rauthe, M.; Steiner, H.; Riediger, U.; Mazurkiewicz, A.; Gratzki, A. A Central European precipitation climatology—Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS). *Meteorol. Z.* **2013**, *22*, 235–256. [CrossRef]
35. Monteith, J.L. Evaporation and environment. *Symp. Soc. Exp. Biol.* **1965**, *19*, 205–234.
36. Warscher, M.; Strasser, U.; Kraller, G.; Marke, T.; Franz, H.; Kunstmann, H. Performance of complex snow cover descriptions in a distributed hydrological model system: A case study for the high Alpine terrain of the Berchtesgaden Alps. *Water Resour. Res.* **2013**, *49*, 2619–2637. [CrossRef]
37. Richards, L.A. Capillary conduction of liquids through porous mediums. *Physics* **1931**, *1*, 318–333. [CrossRef]
38. Liersch, S.; Drews, M.; Pilz, T.; Salack, S.; Sietz, D.; Aich, V.; Larsen, M.A.D.; Gädeke, A.; Halsnæs, K.; Thiery, W. One simulation, different conclusions—The baseline period makes the difference! *Environ. Res. Lett.* **2020**. [CrossRef]
39. Tolson, B.A.; Shoemaker, C.A. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resour. Res.* **2007**, *43*. [CrossRef]
40. Behrangi, A.; Khakbaz, B.; Vrugt, J.A.; Duan, Q.; Sorooshian, S. Comment on "Dynamically dimensioned search algorithm for computationally efficient watershed model calibration" by Bryan A. Tolson and Christine A. Shoemaker. *Water Resour. Res.* **2008**, *44*. [CrossRef]
41. Huang, X.; Liao, W.; Lei, X.; Jia, Y.; Wang, Y.; Wang, X.; Jiang, Y.; Wang, H. Parameter optimization of distributed hydrological model with a modified dynamically dimensioned search algorithm. *Environ. Model. Softw.* **2014**, *52*, 98–110. [CrossRef]
42. Arsenault, R.; Poulin, A.; Côté, P.; Brissette, F. Comparison of stochastic optimization algorithms in hydrological model calibration. *J. Hydrol. Eng.* **2014**, *19*, 1374–1384. [CrossRef]
43. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [CrossRef]
44. Gupta, H.V.; Kling, H.; Yilmaz, K.K.; Martinez, G.F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* **2009**, *377*, 80–91. [CrossRef]
45. Moriasi, D.N.; Arnold, J.G.; van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* **2007**, *50*, 885–900. [CrossRef]
46. Krause, P.; Boyle, D.P.; Bäse, F. Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* **2005**, *5*, 89–97. [CrossRef]
47. Bezak, N.; Brilly, M.; Šraj, M. Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. *Hydrol. Sci. J.* **2014**, *59*, 959–977. [CrossRef]
48. Curceac, S.; Atkinson, P.M.; Milne, A.; Wu, L.; Harris, P. An evaluation of automated GPD threshold selection methods for hydrological extremes across different scales. *J. Hydrol.* **2020**, *585*, 124845. [CrossRef]
49. Lang, M.; Ouarda, T.B.M.J.; Bobée, B. Towards operational guidelines for over-threshold modeling. *J. Hydrol.* **1999**, *225*, 103–117. [CrossRef]
50. Solari, S.; Losada, M.A. A unified statistical model for hydrological variables including the selection of threshold for the peak over threshold method. *Water Resour. Res.* **2012**, *48*. [CrossRef]
51. Pickands, J. Statistical Inference Using Extreme Order Statistics. *Ann. Stat.* **1975**, *3*, 119–131. [CrossRef]
52. Balkema, A.A.; de Haan, L. Residual Life Time at Great Age. *Ann. Probab.* **1974**, *2*, 792–804. [CrossRef]
53. Scarrott, C.; MacDonald, A. A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT Stat. J.* **2012**, *10*, 33–60.

54. Bernardara, P.; Mazas, F.; Kergadallan, X.; Hamm, L. A two-step framework for over-threshold modelling of environmental extremes. *Nat. Hazards Earth Syst. Sci.* **2014**, *14*, 635. [CrossRef]
55. Coles, S.; Bawa, J.; Trenner, L.; Dorazio, P. *An Introduction to Statistical Modeling of Extreme Values*; Springer: Berlin/Heidelberg, Germany, 2001. [CrossRef]
56. Gharib, A.; Davies, E.; Goss, G.; Faramarzi, M. Assessment of the Combined Effects of Threshold Selection and Parameter Estimation of Generalized Pareto Distribution with Applications to Flood Frequency Analysis. *Water* **2017**, *9*, 692. [CrossRef]
57. Cunnane, C. A note on the Poisson assumption in partial duration series models. *Water Resour. Res.* **1979**, *15*, 489–494. [CrossRef]
58. Fischer, S.; Schumann, A. Robust flood statistics: Comparison of peak over threshold approaches based on monthly maxima and TL-moments. *Hydrol. Sci. J.* **2016**, *61*, 457–470. [CrossRef]
59. Hosking, J.R.M. L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *J. R. Stat. Soc. Ser. B* **1990**, *52*, 105–124. [CrossRef]
60. Hosking, J.R.M.; Wallis, J.R. *Regional Frequency Analysis. An Approach Based on L-Moments*; Cambridge University Press: Cambridge, UK, 1997. [CrossRef]
61. Reiss, R.-D.; Thomas, M. *Statistical Analysis of Extreme Values. with Applications to Insurance, Finance, Hydrology and Other Fields*, 3rd ed.; Birkhäuser Verlag AG: Basel, Switzerland, 2007. [CrossRef]
62. Robson, A.; Reed, D. *Flood Estimation Handbook Volume 3: Statistical Procedures for Flood Frequency Estimation*; Centre for Ecology and Hydrology: Wallingford, UK, 1999; ISBN 0948540915.
63. Bačová-Mitková, V.; Onderka, M. Analysis of extreme hydrological Events on the Danube using the Peak Over Threshold method. *J. Hydrol. Hydromech.* **2010**, *58*, 88–101. [CrossRef]
64. Prein, A.F.; Gobiet, A. Impacts of uncertainties in European gridded precipitation observations on regional climate analysis. *Int. J. Climatol.* **2017**, *37*, 305–327. [CrossRef] [PubMed]
65. Meyer, J.; Kohn, I.; Stahl, K.; Hakala, K.; Seibert, J.; Cannon, A.J. Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments. *Hydrol. Earth Syst. Sci.* **2019**, 1339–1354. [CrossRef]
66. Bayerisches Landesamt für Umwelt. Allgemeine Daten zur Wasserwirtschaft. Available online: https://www.lfu.bayern.de/wasser/allgemeine_daten_wasserwirtschaft/index.htm (accessed on 2 July 2020).
67. United States Department of Agricultire (USDA). *Soil Survey Manual*; Government Printing Office: Washington, DC, USA, 1951.
68. Deutscher Verband für Wasserwirtschaft und Kulturbau Fachausschuss Verdunstung. *Ermittlung der Verdunstung von Land- und Wasserflächen*; Wirtschafts- und Verl.-Ges. Gas und Wasser: Bonn, Germany, 1996; Volume 238, pp. 1–135, ISBN 389554034X.
69. Tolson, B.A.; Shoemaker, C.A. Efficient prediction uncertainty approximation in the calibration of environmental simulation models. *Water Resour. Res.* **2008**, *44*. [CrossRef]
70. Gupta, H.V.; Sorooshian, S.; Yapo, P.O. Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration. *J. Hydrol. Eng.* **1999**, *4*, 135–143. [CrossRef]
71. Seong, C.; Her, Y.; Benham, B. Automatic Calibration Tool for Hydrologic Simulation Program-FORTRAN Using a Shuffled Complex Evolution Algorithm. *Water* **2015**, *7*, 503–527. [CrossRef]
72. Mizukami, N.; Rakovec, O.; Newman, A.J.; Clark, M.P.; Wood, A.W.; Gupta, H.V.; Kumar, R. On the choice of calibration metrics for "high-flow" estimation using hydrologic models. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 2601–2614. [CrossRef]
73. Knoben, W.J.M.; Freer, J.E.; Woods, R.A. Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 4323–4331. [CrossRef]

2.2 Paper II: The impact of bias correcting regional climate model results on hydrological indicators for Bavarian catchments

Reference: Willkofer, F., Schmid, F.-J., Komischke, H., Korck, J., Braun, M., & Ludwig, R. (2018). The impact of bias correcting regional climate model results on hydrological indicators for Bavarian catchments. Journal of Hydrology: Regional Studies, 19, 25–41. https://doi.org/10.1016/j.ejrh.2018.06.010

Transition to paper II: In paper II the impact of employing bias-adjusted outputs of RCMs as a driver for a hydrological model on climate change signals (CCS) of hydrological indicators was investigated. Therefore, a validated hydrological model was driven by the raw output of three RCMs (RACOM2, CCLM, REMO-UBA) for a reference period to illustrate deviations from observations and thus, the necessity for bias correction (BC). The RCMs were bias-adjusted using three methods (linear scaling, local intensity scaling, quantile-mapping; yearly and monthly correction factors). The results of the hydrological model driven by observed meteorology and adjusted RCM outputs were compared by means of best adjustment of the RCM to observations and its impact on the CCS using a set of hydrological indicators. For this thesis the impact of BC on the CCS of extreme indicators is of particular interest especially for events which are larger than those analyzed in paper I. Thus, the findings of this paper influenced the decision to opt for a quantile-mapping approach adapted to sub-daily time steps to adjust the output of the CRCM5-LE used to produce the data presented in paper III.

Author's contribution: FW developed the concept for this study, performed the formal analysis and investigation, visualized the data and wrote the original manuscript. FW and FJS were responsible for data curation and validation of the analysis and developed the required software tools. FW, FJS, and MB developed the employed methodology for the analysis. RL acquired the funding for this study, was responsible for the project administration and resources, and supervised the research. FJS, MB, HK, JK, and RL reviewed and edited the manuscript.

Status: published
Journal: Journal of Hydrology: Regional Studies
Impact factor: 4.7

Journal of Hydrology: Regional Studies 19 (2018) 25–41

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

The impact of bias correcting regional climate model results on hydrological indicators for Bavarian catchments

Florian Willkofer[a,*], Franz-Josef Schmid[a], Holger Komischke[b], Jane Korck[b], Marco Braun[c], Ralf Ludwig[a]

[a] Department of Geography, Ludwig-Maximilians-Universtiät München, Luisenstraße 37, 80333, Munich, Germany
[b] Bavarian Environment Agency (LfU), Hans-Högn-Str. 12. 95030, Hof, Germany
[c] Ouranos, 550 Rue Sherbrooke West, West Tower, 19th Floor, Montreal QC H3A 1B9, Canada

ARTICLE INFO

Keywords:
Bias correction
Regional climate model
Climate change signal
Hydrological modeling
Runoff indicators
Bavaria

ABSTRACT

Study region: The Mindel river catchment, gauge Offingen, Bavaria, Germany.
Study focus: The study investigates the potential interference of climate change signals (CCS) in hydrological indicators due to the application of bias correction (BC) of regional climate models (RCM). A validated setup of the hydrological model WaSiM was used for runoff modeling. The CCS, gained by the application of three RCMs (CCLM, REMO-UBA, RACMO2) for a reference period (1971–2000) and a scenario period (2021–2050), are evaluated according to eight hydrological indicators derived from modeled runoff. Three different BC techniques (linear scaling, quantile mapping, local intensity scaling) are applied.

New hydrological insights for the region: Runoff indicators are calculated for the investigated catchment using bias corrected RCM data. The quantile mapping approach proves superior to linear scaling and local intensity scaling and is recommended as the bias correction method of choice when assessing climate change impacts on catchment hydrology. Extreme flow indicators (high flows), however, are poorly represented by any bias corrected model results, as current approaches fail to properly capture extreme value statistics. The CCS of mean hydrological indicator values (e.g. mean flow) is well preserved by almost every BC technique. For extreme indicator values (e.g. high flows), the CCS shows distinct differences between the original RCM and BC data.

1. Introduction

In recent years, large efforts have been made in climate research to improve process understanding and advance computation power to allow for higher resolution dynamical regional climate models (RCM) (Kotlarski et al., 2014). Meanwhile, a large number of RCM results have been made available to a growing user community, showing a broad range of variability and bias (Christensen et al., 2008; Giorgi et al., 2009; Kotlarski et al., 2014; van der Linden and Mitchell, 2009). Reasons for deviations from observations are manifold and encounter various sources of uncertainty, such as errors in reference data sets (Ehret et al., 2012), the spatio-temporal scale gap between RCMs and observations, differences in model parameterizations (e.g. for convection) (Maraun et al.,

* Corresponding author.
E-mail addresses: florian.willkofer@lmu.de (F. Willkofer), f.schmid@iggf.geo.uni-muenchen.de (F.-J. Schmid), holger.komischke@lfu.bayern.de (H. Komischke), jane.korck@lfu.bayern.de (J. Korck), braun.marco@ouranos.ca (M. Braun), r.ludwig@lmu.de (R. Ludwig).

https://doi.org/10.1016/j.ejrh.2018.06.010
Received 28 September 2017; Received in revised form 20 June 2018; Accepted 23 June 2018
Available online 31 July 2018

2010). The selection of SRES emission scenarios (SRES, Nakicenovic (2000)) or recently developed representative concentration pathways (RCP, van Vuuren et al. (2011)), however, affects the climate change signal for the future period. RCM data is made freely available through various data bases (ENSEMBLES (SRES) (van der Linden and Mitchell, 2009), CORDEX (RCP) (Giorgi et al., 2009)) and evermore climate change impact studies apply these data to assess the effects of potential alterations in climate on various physical, ecological and/or socio-economic aspects (e.g. runoff regimes, extreme discharge, biodiversity, water management) (Hattermann et al., 2014; Lenderink et al., 2007; Majone et al., 2012; Stagl and Hattermann, 2015). However, the increasing resolution of RCMs is mostly still too coarse for smaller scale investigations in hydrology, so additional downscaling techniques must be applied (Cloke et al., 2013). Besides this scale issue, RCMs often exhibit pronounced systematic deviations from any given reference period which are considered as bias (Ehret et al., 2012; Kotlarski et al., 2014; Maraun, 2016). If large enough, these biases can result in significantly and often non-linearly different outputs from subsequent models (e.g. for hydrological models) (Chen et al., 2011) which are usually calibrated against observations. Thus, the bias between the observations and the models has to be removed before the data is applicable for impact models. Several methods have been developed for this purpose and are often critically discussed (Ehret et al., 2012; Maraun et al., 2010). Recent studies indicate, that bias correction (BC) methods can have different effects on the distribution of any given parameter (e.g. precipitation), and can thus particularly impact its extreme values (Hagemann et al., 2011; Mudelsee et al., 2010). The underlying principle and thus the most crucial assumption is that the bias correction factors retrieved by any such methods must necessarily be considered valid for the future, assuming a temporal stationarity and thus introducing another, yet often neglected source of uncertainty (Teutschbein and Seibert, 2012). Hence, it must be argued that BC methods might falsify the original climate change signal (CCS) of RCMs with extreme values being stronger affected than means (Themeßl et al., 2012). Regarding the influence of the use of bias corrected data on hydrological modeling, Muerth et al. (2012) point out that individual simulations with a strong inherent bias visibly affect the CCS of hydrological indicators. The overall mean CCS of large RCM ensembles (i.e. multiple member of a RCM driven by the same GCM with changing initial conditions) however seem to be less sensitive to BC.

Many studies investigated the removal of bias in RCMs, resulting in a myriad of methods and various performances for specific purposes (e.g. Maraun et al. (2010); Themeßl et al. (2012)). The study by Muerth et al. (2012) investigated the influence of BC on the representation of observed runoff, the impact of CC on the runoff regime and the effect of BC on the future change in hydrological indicators over a single catchment in Bavaria. Hagemann et al. (2011) state that the hydrological CCS at certain locations and for specific seasons might be affected by the BC of raw GCM data. This impact of BC on the CCS of hydrological indicators is also significant if outputs from corrected RCMs are applied as a meteorological driver of hydrological models (Muerth et al., 2012). Cloke et al. (2013) investigated the impact of BC on the CCS of extreme discharges for the Upper Severn catchment, England, and found that it is even stronger than for mean flows.

To further investigate this specific topic in the course of its routine operations in water resources management (e.g. design of flood detention basins based on a threshold for extreme high flows), the Bavarian Environment Agency (Bayerisches Landesamt für Umwelt (LfU)) requested to analyze the performance of three bias correction methods (local intensity scaling, quantile mapping, linear scaling) for multiple Bavarian catchments in the framework of the BI-KLIM[1] project. These specific BC approaches are chosen for being considered state-of-the-art methods to adjust the systematic differences between RCM data and observations (Ehret et al., 2012). Hence, the purpose of this study was:

a) to determine the most sufficiently performing bias correction method as a standard approach for the Bavarian domain (see Fig. 1, upper left) and
b) to quantify and evaluate the effects of bias correction on the CCS of specific hydrological indicators for river catchments located in Bavaria, Germany.

This paper focuses on the effects of bias correction on the CCS of hydrological indicators. The climate simulations ensemble for this study includes three different RCMs: the COSMO-CLM (CCLM 4.8, Berg et al. (2013); Wagner et al. (2013)) of the Karlsruhe Institute of Technology (KIT)[2], the REMO-UBA[3], and RACMO (v2.1) of the KNMI (van Meijgaard et al., 2008), all driven by the same global circulation model (GCM, ECHAM5, Roeckner et al. (2003)) (further referred to as: CCLM, REMO, RACMO). The hydrological model WaSiM (Schulla, 2012) was applied to determine the impacts of BC to the CCS in the hydrology of several selected Bavarian catchments.

The performance of BC methods is evaluated by comparing long term flow regimes as well as specific flow indicators resulting from the hydrological modeling. A reference data set of observed data was set up at the beginning of the project. This dataset is further used as the observational reference for hydrological modeling and bias correction. The effects of BC on the CCS of the catchment's hydrology are investigated using the same hydrological indicators.

[1] Einfluss der Biaskorrektur dynamischer regionaler Klimamodelldaten auf die Wasserhaushaltsmodellierung und Klimafolgenabschätzung in Bayerischen Flussgebieten (BI-KLIM) (Impact of bias correction of dynamic regional climate model data on water balance modeling and assessment of climate impacts for Bavarian catchments).

[2] Institute of Meteorology and Climate Research, Department Troposphere Research (IMK-TRO) of the KIT, 2011. Provision of CCLM forcing data, version 4.8, calculated by the KIT for runoff models for KLIWA. Unpublished report on behalf of the Bavarian Environment Agency (LfU), Measurements and Environmental Protection Baden-Württemberg, and Water Management and Factory Department Rheinland-Pfalz.

[3] Max-Planck-Institute (MPI) under contract to the German Federal Environment Agency, 2006.

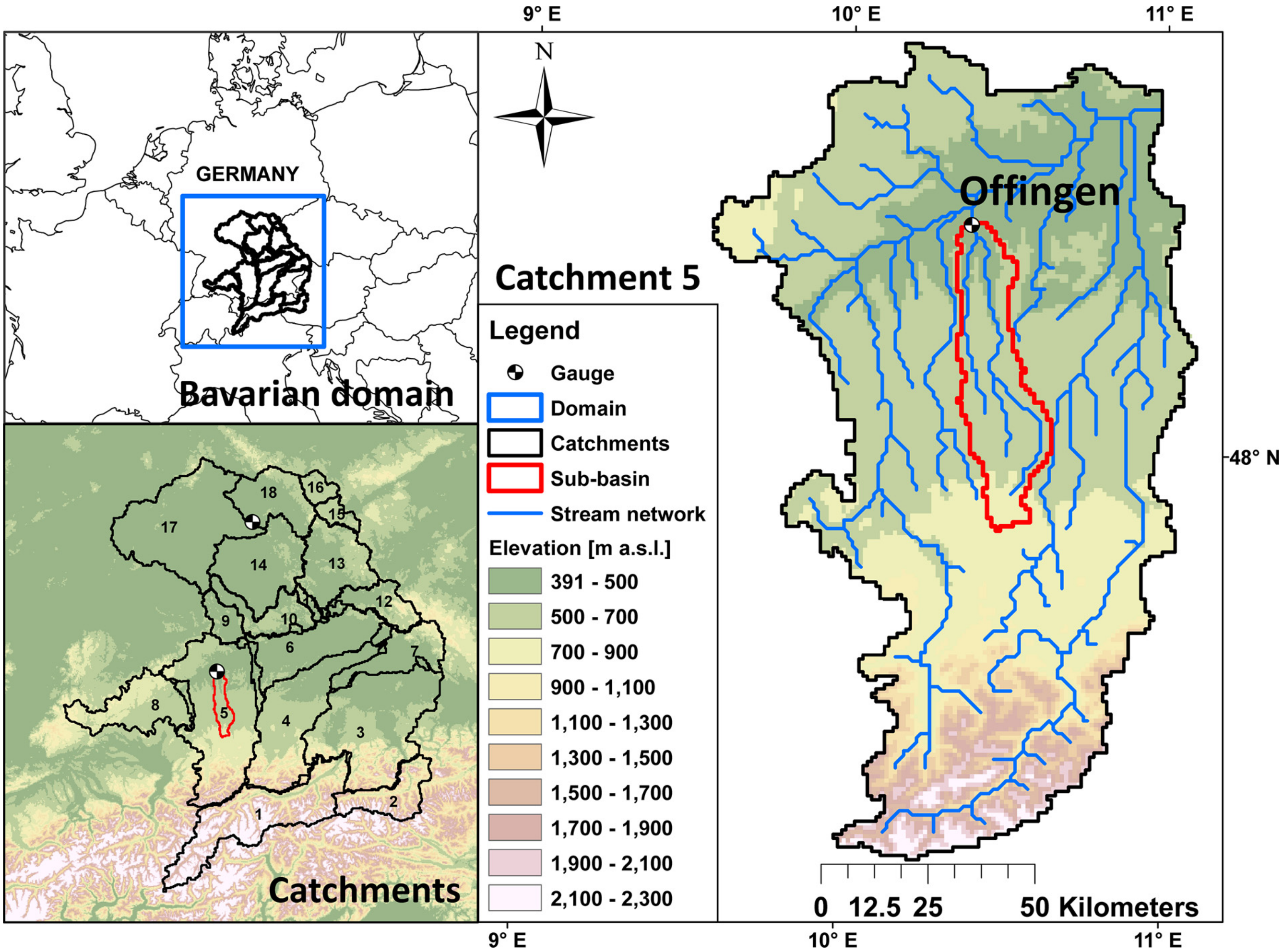


Fig. 1. Catchments of the hydrological Bavaria and the surrounding domain needed for scaling purposes. The Mindel sub-basin to the gauge Offingen (right, red boundary) is situated within the Iller-Lech catchment (5, lower left). The blue box in the upper left depicts the domain used for bias correction and spatial downscaling of RCM data.

2. Study area, data and methods

2.1. Study area

The study area covers the major Bavarian river basins including their headwaters in southern Germany and partly adjoining states (Austria to the south, Czech Republic to the east), furtherly referred to as "hydrological Bavaria". It comprises 18 hydrological catchments modeled separately with the Water balance Simulation Model (WaSiM, Schulla (2012)) at the LfU as illustrated in Fig. 1 (left). Furthermore, this figure shows the surrounding domain (upper left, blue box) used for the bias correction of the RCM data for the Bavarian catchments (lower left). The following sections will focus on catchment 5 representing the Iller-Lech river system and parts of the Danube. In particular, results are shown for the Mindel river sub-basin up to the gauge Offingen (Fig. 1, right red outline) covering an area of about 929 km^2, since it represents a relatively pristine basin with only limited effects from water management infrastructure. Other catchments (Lech river to the East or Iller river to the West) are heavily impacted by artificial reservoirs and dams, which imposes additional challenges on the hydrological modeling outside the scope of this study.

This sub-basin is characterized by pre-alpine topography, showing a S-N gradient from the gauge in the north at 440 m a.s.l. to the highest peak in the south at 860 m a.s.l..The long term precipitation sums follow this gradient, ranging from 1100 mm in the southern part to 750 mm in the north. Mean temperatures range from -1 °C (January) to about 18 °C (August) and the mean annual evapotranspiration is around 570 mm. With a mean flow of 12.2 m^3/s, ranging from 11.5 m^3/s in the summer to 12.9 m^3/s in winter, the overall annual runoff variation in this pluvio-nival flow regime remains quite small.

2.2. Data

The data for this study is provided by the LfU covering the 18 catchments and including measured values from stream gauges and meteorological stations as well as grid based meteorological data.

Performing bias correction requires a meteorological reference to compute the change factors based on a distribution function or

Table 1
Meteorological dataapplied for the creation of a reference data set. Data is provided by the LfU. All data is based on interpolated measurements using different interpolation methods. The HYRAS data set was developed by the German Weather Service (DWD) and provided only for the German parts of the hydrological Bavaria.

Source	Data type	Parameter	Interpolation	References
HYRAS data set (DWD)	Raster, 1 x 1 km^2	Precipitation [mm] (1) Air temperature [°C] (2) Rel. air humidity [1/1] (3)	REGNIE (1) Optimal Interpolation (2) & (3)	Rauthe et al. (2013), Frick et al. (2014) / Gandin (1965)
Interpolation from hydrological model	Raster, 1 x 1 km^2	Precipitation [mm] Air temperature [°C] Rel. air humidity [1/1] Global radiation [Wh/ m^2] Wind speed [m/s]	Regionally different weighted combination of Inverse Distance Weighting (IDW) and altitude dependent regression	Pöhler et al. (2010)

simple deltas for the modification of the RCM values. Here, a reference data set based on a regular grid was created by combining meteorological data from different sources (Table 1) for different regions. Fig. 2 shows the spatial distribution of the different data sources and types for the climatological variables with station values covering the Danube tributaries to the south and HYRAS raster covering the northern tributaries as well as the Main catchment (except for air humidity (H) in two northern catchments). Wind speed (W) and radiation (R) are interpolated from station measurements for the entire hydrological Bavaria. All meteorological data are available on a daily basis.

The adjacent grids of each region of are spatially merged for each time step. In combination, both data sets provide a regular grid at the resolution of the hydrological model of 1 x 1 km^2 covering the entire Bavarian domain. The different data sets (HYRAS and interpolated station data) exhibit patterns due to the different development schemes. Hence, the Danube catchments show a more pronounced topographical pattern, while the Main catchment shows a more diffuse picture. The sharp transition between the regions might influence results of affected catchments. However, the different schemes have no impact on the findings shown in this study. The different sources of meteorological data are applied to the respective hydrological model for the catchments of Fig. 1.

The reference period for this study covers the period from 1971 to 2000. This period was chosen since meteorological data was available in sufficient spatial and temporal coverage. The presented investigation on the impacts of bias correction on CCS is based on data from three different RCMs, all using ECHAM5 as the driving GCM. Two were provided by the LfU: the CCLM and REMO. These RCMs are frequently used in other climate change related projects funded by the LfU (e.g. KLIWA, AdaptAlp, ClimChAlp)since their high spatial resolution is considered to be advantageous for applications in high relief terrain as Bavaria (especially over the Alps). In order to show the performance of BC on a coarse resolution model the RACMO RCM with a spatial resolution of 50 km was applied in this study. The respective characteristics of each RCM are given in Table 2.

The driving GCM and its members (i.e. GCM runs with slightly altered initial conditions) are the same for all three RCMs. Hence, the differences in the results using the different RCM ensembles (CCLM, RACMO, REMO) originate from the differences in the RCM configurations (e.g. resolution, domain size) However, as with the GCM members, variations between RCM members originate from their respective initial conditions. In contrast to RACMO and CCLM, with three members each, there is only one member available for the REMO RCM. Furthermore, the REMO precipitation shows a shift in precipitation fields in mountainous areas due to luv and lee effects. A minor precipitation event from clouds at 3000 m altitude might be shifted by up to 15 km if affected by wind speeds up to

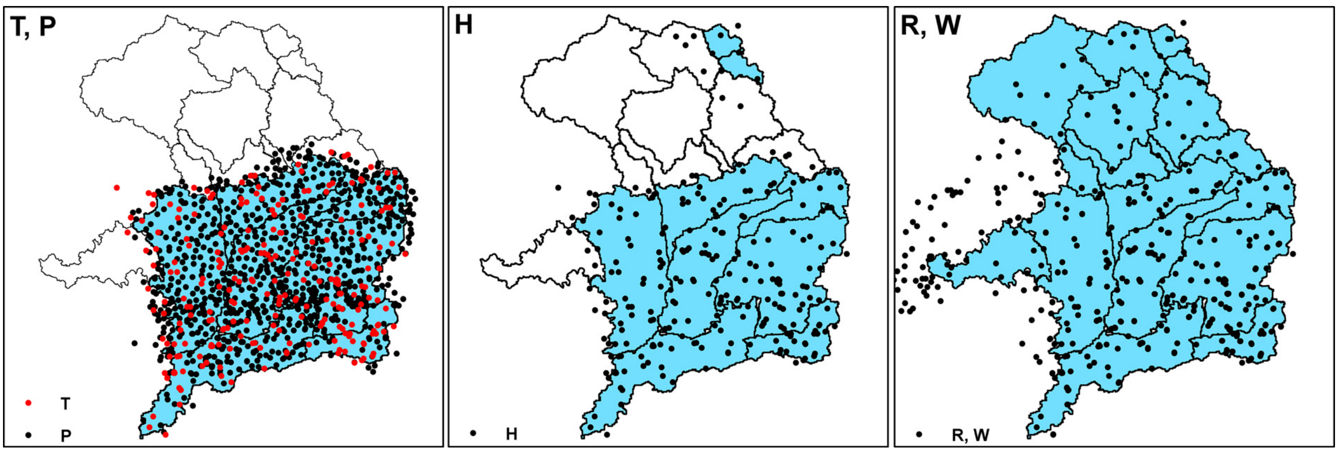


Fig. 2. Spatial distribution of the different data types available for the hydrological Bavaria. White indicates raster based interpolated measurements (HYRAS data, for air temperature (T), precipitation (P) and air humidity (H) only) whereas blue indicates interpolated point measurements from meteorological gauges. The dots in the frame show the distribution of the stations of the respective variable.

Table 2
Regional climate models applied in this study. The table shows the parameters, their spatial and temporal resolution, the period and driving GCM members available for this study.

Model	Parameter	Resolution	Period	Member (Scenario)
CCLM4.8	air temperature, precipitation, global radiation, wind speed, air humidity	7 x 7 km^2 daily	1971-2000 2011-2050	3x ECHAM5 Member 1 to 3 (A1B)
RACMO (v2.1)		50 x 50 km^2 daily	1950-2100	3x ECHAM5 Member 1 to 3 (A1B)
REMO		10 x 10 km^2 daily	1951-2100	1x ECHAM5 Member 1 (A1B)

10 m/s (Göttel, 2009). This spatial offset has to be considered in all further analysis. Table 3 illustrates the long term yearly mean values of the different meteorological variables of the reference data set and the raw RCMs as well as their respective absolute and relative biases.

The biases are considerable, especially for temperature and precipitation of all CCLM members and for temperature of the REMO RCM. Also precipitation biases for the RACMO RCM are significant (> 10%). Wind speed, global radiation and relative air humidity also exhibit strong relative biases. However, their absolute deviations are rather small. A proper correction of the bias of precipitation and air temperature is most important to allow hydrological models to produce reasonable outputs. However, since the hydrological model applied in this study requires all the above mentioned variables, they are also corrected for a better representation of the observed values.

2.3. Methods

For the purpose of analyzing the influence of the bias correction on the climate change signal a model chain was introduced (Fig. 3) with the BI-KLIM data base as central component. This data base includes all the pre- and post-processed RCM data (raw, scaled, and bias corrected). The bias correction is conducted at RCM resolution; thus, a spatial aggregation of the reference data set to the RCM scale was performed. After bias correction, the RCM data was further downscaled to the hydrological model grid, applying the scaling tool SCALMET (Marke, 2008). The influence of the bias correction on the climate change signal of the hydrological regimes was analyzed by applying all available raw and preprocessed data to the hydrological model WaSiM for the Mindel sub-basin within the Iller-Lech catchment.

2.3.1. Bias correction methods

RCM data usually display a statistical mismatch to recorded meteorological variables, a bias. In order to make the data better applicable and acceptable for users, various methods have been developed to correct such biases via transformation algorithms to statistically match the observations. A good overview of the various available approaches for bias correction is given by Teutschbein and Seibert (2012). The usual methods share the assumption that the retrieved correction factors and addendums are considered stationary in space and time. Thus, they are taken to be valid for the reference and the scenario period as well. This assumption is not challenged here, as the paper is focused on assessing the impacts of this common practice.

Table 3
Comparison (values, absolute and relative difference) of long-term yearly mean values between the reference data set and the raw RCM data.

Variable	Reference	RACMO			CCLM			REMO
		M1	M2	M3	M1	M2	M3	
Temperature [°C]	6.93	7.13	6.9	7.22	6.19	6.02	6.41	7.9
Precipitation [mm/a]	1048	1167	1175	1153	1668	1693	1674	1094
Relative air humidity [%]	75	81	81	81	86	86	85	74
Global radiation [Wh/m^2]	126	112	111	112	102	104	103	126
Wind speed [m/s]	2.05	2.92	2.9	2.91	3.56	3.55	3.54	3.41
		Absolute bias						
Temperature [°C]		0.2	−0.03	0.29	−0.74	−0.91	−0.52	0.97
Precipitation [mm/a]		119	127	105	620	645	626	46
Relative air humidity [%]		6	6	6	11	11	10	−1
Global radiation [Wh/m^2]		−14	−15	−14	−24	−22	−23	0
Wind speed [m/s]		0.87	0.85	0.86	1.51	1.5	1.49	1.36
		Relative bias						
Temperature		2.9	−0.4	4.2	−10.7	−13.1	−7.5	14.0
Precipitation		11.4	12.1	10.0	59.2	61.5	59.7	4.4
Relative air humidity		8.0	8.0	8.0	14.7	14.7	13.3	−1.3
Global radiation		−11.1	−11.9	−11.1	−19.0	−17.5	−18.3	0.0
Wind speed		42.4	41.5	42.0	73.7	73.2	72.7	66.3

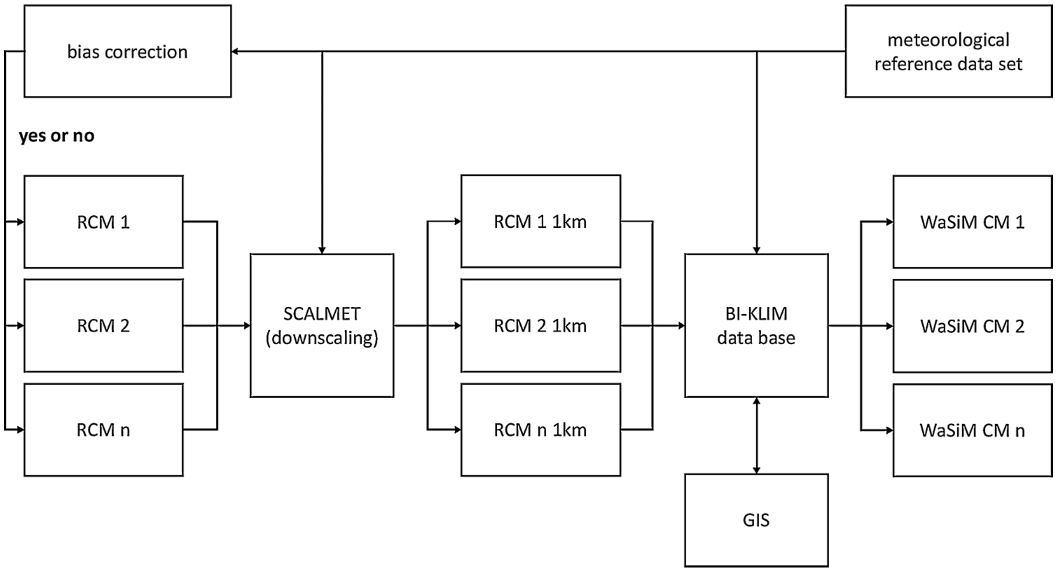


Fig. 3. Process chain with the BI-KLIM data base as most essential component including all of the processed data in the desired resolution for hydrological modeling applications. The GIS interface provides the opportunity to extract parameter values of each climate variable in table format for a single catchment. The bias correction is carried out on RCM resolution. (RCM: regional climate model; CM: climate model).

A common shortcoming of dynamic RCM data is the overestimation of the number of days with very little precipitation (Teutschbein and Seibert, 2012). This problem refers to the size of the raster cells of a RCM in combination with the convection of moist air. As the moist air reaches full saturation at a certain height with decreasing temperature, it will induce rainfall for a large area within a RCM. This process is further referred to as the 'drizzle-effect' (Dai, 2006). Dai (2006) also points out that this area-wide drizzle would not occur under natural conditions due to atmospheric instabilities and refers this effect to the model scale. Consequently, this particular portion of the RCM precipitation has to be removed in advance of the bias correction to avoid its influence on the modification factors. Kjellström et al. (2010) tested several thresholds for a minimum precipitation amount for handling the drizzle effect and found 1 mm/day to be a good value to remove excess drizzle precipitation from model data. Values up to this threshold do not significantly contribute to overall precipitation sums (Dai, 2001). Thus, this approach was applied for the elimination of the drizzle for all available RCM data in this study.

In contrast to the variability between the different RCMs, the changes in initial conditions of the driving GCM for the three members of the CCLM and RACMO induce an internal variability between these members of the particular RCM, which can be considered as natural variability (Elía and Côté, 2010; Muerth et al., 2012). To maintain this variability between the members of the CCLM and RACMO RCM, a multi member bias correction was performed. Here, a single set of correction factors is derived using the statistics of all the three respective members of the RCM (Muerth et al., 2012), instead of one set for each of the members. This allows for ascribing the differences in the annual course to the respective member of these small RCM ensembles. Furthermore, this ensures that a measure for the natural climate variability is maintained.

As mentioned above, for this study we used three different methods for bias correction which are briefly described here. The correction factors are calculated on a monthly (1 factor per month) as well as on a yearly (one factor per year) basis. Additionally, the multi member approach is applied to either of the sets of correction factors. Furthermore, values of relative air humidity are corrected in terms of dew point temperature, applying the Magnus formula for conversion. Since air temperature is required for the transformation, a good match between those two variables is maintained.

2.3.1.1. Linear scaling (ls). Linear Scaling is applied according to Lenderink et al. (2007), with slight changes regarding the long-term averages. For this approach we used the additive (air temperature [°C] (1)) or multiplicative (precipitation (2)) differences between the monthly (yearly) averages of the reference and the RCM data for the reference period similar to Teutschbein and Seibert (2012). The resulting correction factors are then applied to each daily (t) value of the entire time series of the RCM by addition or multiplication depending on the climate parameter to be corrected.

$$T_{RCM,cor}(t) = T_{RCM}(t) + (\overline{T_{obs}} - \overline{T_{RCM}}) \tag{1}$$

$$P_{RCM,cor}(t) = P_{RCM}(t) \bullet (\overline{P_{obs}} / \overline{P_{RCM}}) \tag{2}$$

The multiplicative approach also applies for the parameters wind speed and global radiation since these parameters have an absolute zero value like precipitation, whereas for air humidity the additive approach is used.

2.3.1.2. Quantile mapping (qm). The quantile mapping approach attempts to adjust the distribution function of values of the RCM to match the distribution function of observed values for the reference period (Sennikovs and Bethers, 2009; Teutschbein and Seibert, 2012). Thus, the correction factors depend directly on the values of both time series.

This study uses a modified empirical quantile mapping approach based on an daily translation after Mpelasoka and Chiew (2009). Apart from the usual multiplicative correction factors, the adapted approach of this study also provides additive factors to adjust temperature and relative air humidity (via dew point temperature). The distribution function for RCM and observed values is created by a division of the values using percentiles. In a first step, the values of each percentile i (with $i = 2k+1$, $k = [0,49]$) for both time series (observations and raw model data) are defined. If the percentile is in between two values of a time series a weighted mean will be calculated. The second step performs a cubic interpolation of the predefined percentile values to n percentiles. In order to prevent sharp edges between percentiles, the cubic interpolation to represent the fitting of the 50 percentile points is preferred over a linear interpolation in this study. The number n of percentiles can be altered and typically ranges between 0 and 100. For this investigation a value of 50 was chosen as this number was considered to sufficiently represent the distribution. Since every value of the time series is affected by the correction, also extreme values will be adjusted. Those new extreme values are achieved by an extrapolation of the percentile values > 99% and < 1% for the corrected model time series. This allows for the calculation of correction factors for the lowermost and uppermost percentile. In the last step the n percentiles are derived from the time series to be corrected. This also applies for the raw model time series for the future period. Hence, there are n values of the time series to be corrected and n correction factors for the respective percentiles. Afterwards, the correction factors closest to the respective percentile are assigned to the values of the original RCM time series. All these steps also apply for time series of single months which leads to $12n$ correction values.

2.3.1.3. Local intensity scaling (LOCI). The local intensity scaling method (Schmidli et al., 2006) only applies for precipitation values. This approach is based on a scaling factor depending on wet-day intensities (3) and a wet-day threshold (WDT) derived from the wet-day frequency of daily observed (P_{OBS}) and model data (P_{RCM}).

$$s = \frac{\overline{P_{OBS}\colon P_{OBS} \geq P_{OBS}^{WDT}} - P_{OBS}^{WDT}}{\overline{P_{RCM}\colon P_{RCM} \geq P_{RCM}^{WDT}} - P_{RCM}^{WDT}} \tag{3}$$

The corrected time series is then calculated as follows:

$$P_{RCM,cor} = \max(P_{OBS}^{WDT} + s(P_{RCM}(t) - P_{RCM}^{WDT}), 0) \tag{4}$$

After the bias correction the new model data by definition have the same wet-day frequency and intensity as the observed time series (Schmidli et al., 2006). However, the overall precipitation sums may differ as for this method only the targeted statistics will match the statistics derived from the observation values (Muerth et al., 2012). In order to draw conclusions about the effects of this approach, the remaining parameters are corrected with the qm method.

2.3.2. The hydrological model WaSiM

The Water balance Simulation Model (WaSiM) was employed to perform the hydrological modeling. WaSiM is characterized as a distributed (grid based (regular / irregular)), mainly physically based, deterministic type of model using constant time steps with internally flexible sub time steps (Schulla, 2012). It is frequently applied for various climate change impact studies (Foltyn et al., 2017; Kleinn et al., 2005; Rößler and Löffler, 2010) or for the analysis on the need for bias correction (e.g. Muerth et al. (2012)). In this study, we applied existing calibrated and validated configurations of WaSiM for the catchments of the hydrological Bavaria and the following results focus on the results for the Mindel sub-basin (gauge Offingen) calculated using the model setup for the Iller-Lech river system (catchment 5, Fig. 1). The model was set up in 1 km spatial resolution and a daily time step and the parameters were derived for the calibration period from 1994 to 1998 and validated for the consecutive period between 1998 and 2003. These time slices were chosen since the number of available meteorological input data was larger. The authors of the model (UDATA, Pöhler et al. (2009)) evaluated the modeled discharge fit by the Nash & Sutcliffe Efficiency (Nash and Sutcliffe, 1970) for raw (NSE) and logarithmic (logNSE) model outputs. Furthermore, a long-term simulation run from 1971 to 2003 was evaluated to test the overall model performance including years with less available input data. The results for the different modeling periods (Table 4) for the Mindel catchment show a fairly good representation of the observed runoff by the model (NSE > 0.5 and logNSE > 0.65). The lack of available input data might influence the performance of the long-term simulation.

Table 4
Performance of the hydrological model WaSiM for the Mindel catchment at the gauge Offingen for different evaluation periods.

Period	NSE	logNSE
Calibration (1994–1998)	0.56	0.67
Validation (1998–2003)	0.59	0.72
Long term run (1971–2003)	0.53	0.68

Table 5
Codification of the multiple WaSiM model run results.

Code	Description
M*X*	Member *X* of the RCM
BC0 / BC1	Raw / bias corrected RCM data
OBS	Modeled using observed data (reference data set)
REF	Reference period
FUT	Future period
m / y	Monthly / yearly correction factors

3. Results

3.1. Bias correction results

The hydrological model for the Iller-Lech catchment is driven by the observed data (reference data set) as well as by the raw and corrected data of the dynamical RCMs. The modeled runoff obtained from driving the model with the reference data set forms the basis to assess the influence of the different bias correction methods. For the different hydrological model outputs a codification for the composition of the different RCM and bias correction methods is given in Table 5.

The long term flow regimes for the reference period between 1971 and 2000 shown in Fig. 4 illustrate the more or less distinct differences between the model runs using observed data (OBS) and those using the raw RCM data downscaled to the hydrological model resolution. While the regimes of the RACMO model mainly differ from the reference during the winter months, the results using the CCLM model overestimate the reference by almost 100% throughout all seasons due to significantly higher modeled precipitation. The runoff produced by the REMO model data, however, underestimates the reference entirely. Regarding the weak

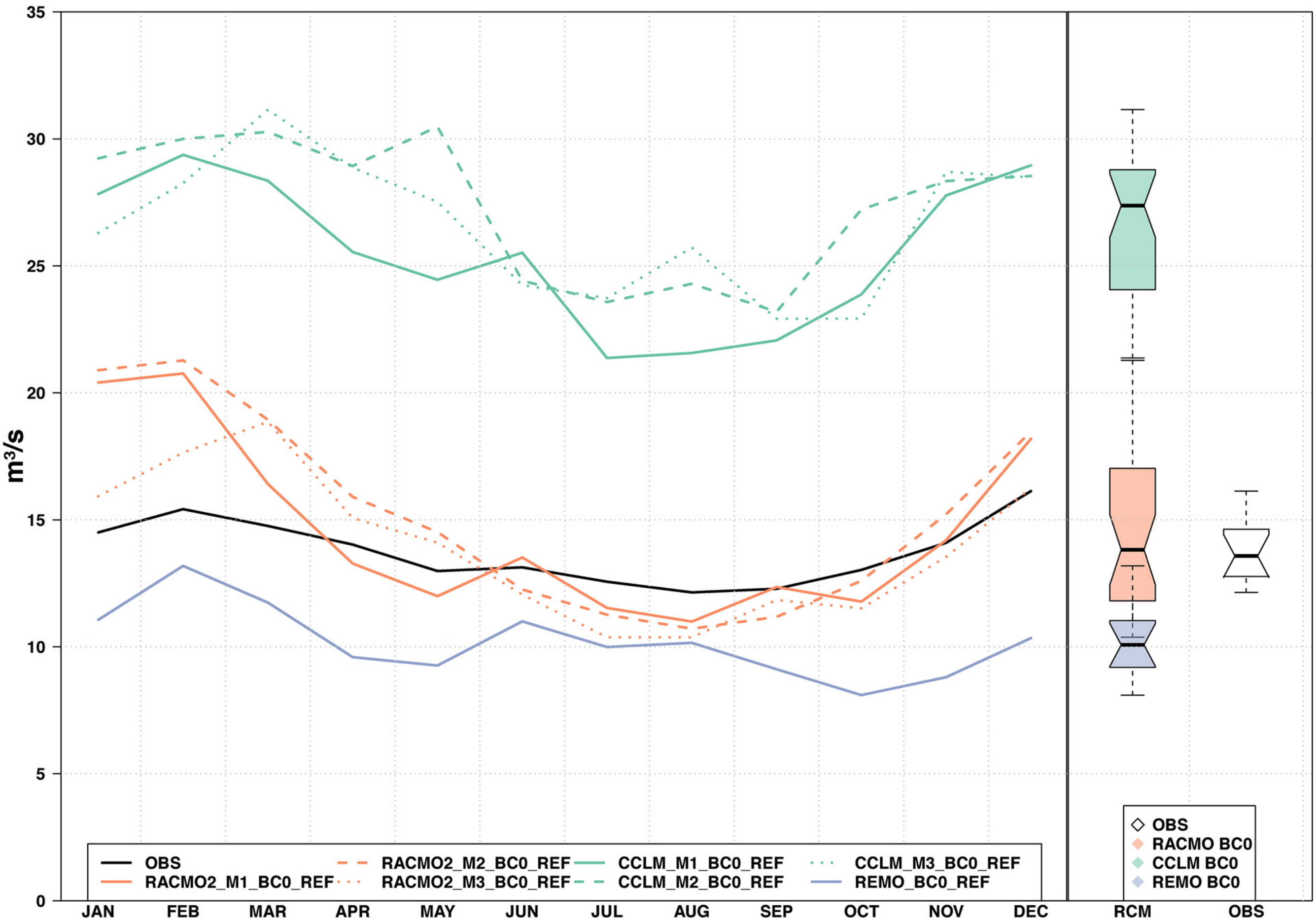


Fig. 4. Long term flow regimes for the reference period (1971–2000) of the river Mindel at the gauge Offingen. The regimes represent results of the hydrological modeling using observed data (reference data set) and raw data of the dynamical RCMs. The regimes using the raw RCM data show more or less significant deviations from the reference regime (REF). Especially the application of the members of the CCLM lead to a significant overestimation. The colored boxplots show the variability of the respective RCM dataset compared to the variability within the model results using observed data.

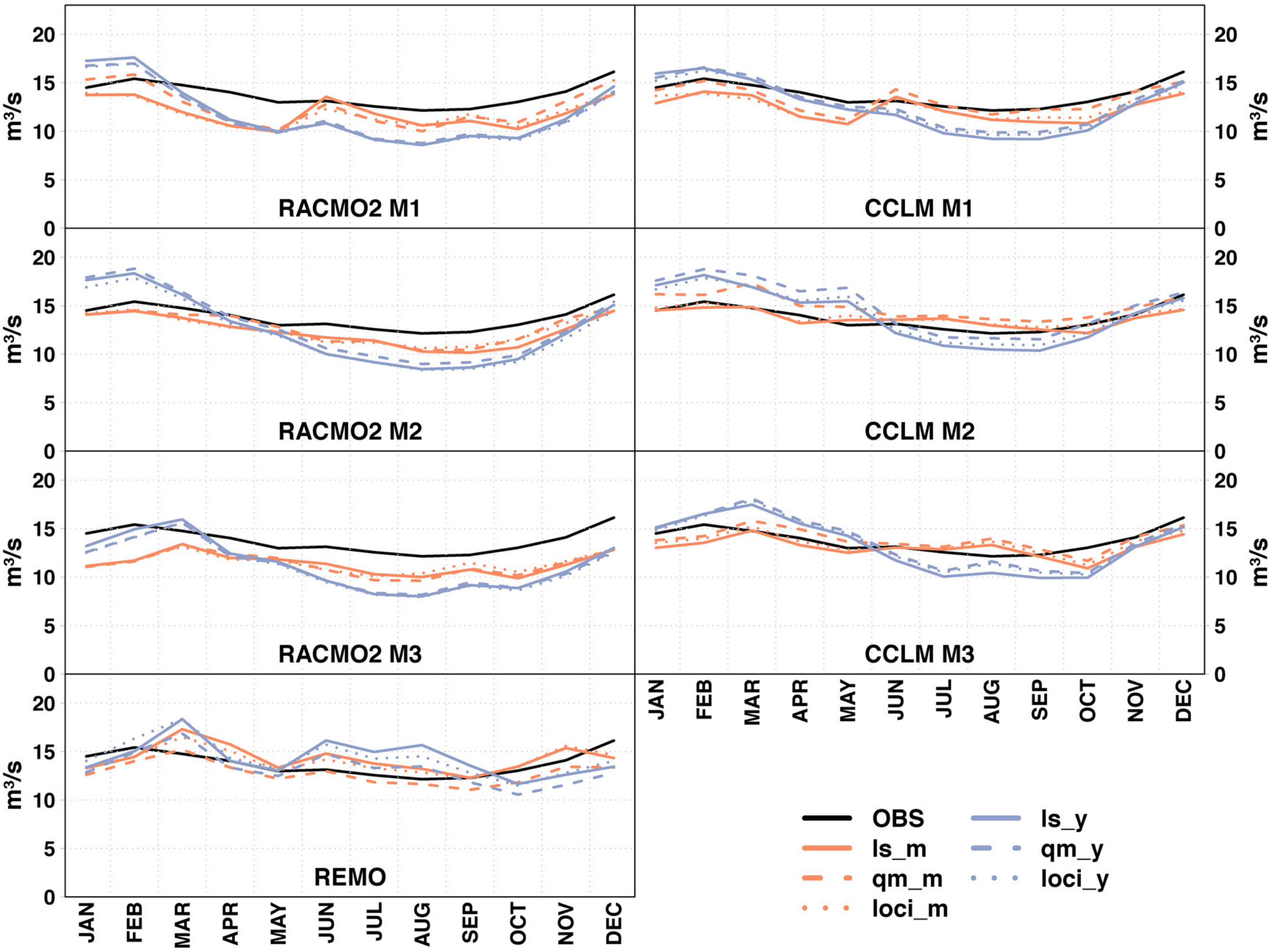


Fig. 5. Long term flow regimes for the reference period (1971–2000) of the river Mindel at the gauge Offingen showing model results using observed data (reference data set) and bias corrected RCM data. In general, approaches using monthly correction factors lead to a better adjustment to the reference. The regimes produced by RCM data using quantile mapping and monthly correction factors show the best results compared to other methods.

seasonal course the modeled regimes of the CCLM and REMO data show higher similarity to the reference. Furthermore, Fig. 4 shows the inter-annual variability of the simulations using raw RCM datasets (colored boxplots) compared to the variability of the model run produced with observed data (transparent box plot). Since there is no overlap between the upper or lower quantiles of simulations results of raw RCM data, these models differ significantly from each other. However, the variability of the RACMO model simulations is similar to the variability of the results using observed data since the notches of both boxplots as well as the median exhibit a good agreement. The variability of modeled results using raw CCLM and REMO RCMs in contrast differs significantly from the reference regime.

Fig. 5 shows the results of the hydrological modeling using the bias corrected RCM data (BC1). For the bias corrected RACMO data using monthly correction factors the results of the hydrological modeling are best for the 2nd member (-7.5% qm_m M2, > 8% for other members and BC methods, see Table 6) showing only minor differences of about 2 m^3/s (> 2 m^3/s for other members and BC methods, e.g. 4m^3/s in spring of M1) during the summer and fall season. In general, while the regimes produced by monthly corrected RACMO data systematically underestimate the observations (-4 m^3/s to -2 m^3/s throughout the year), the results with annual correction coefficients exhibit an overestimation in winter (1 m^3/s to 5 m^3/s) and underestimation in summer (up to -5 m^3/s in August). The flow regimes of the CCLM as well as REMO model show good adjustment, especially for member 2 and 3 of the CCLM with only minor differences regarding those produced with monthly correction factors. Despite the huge deviation of the raw CCLM data the bias correction is able to satisfactorily reproduce the observed runoff.

The results using yearly correction factors depict that the seasonal course of the results produced by raw RCM data is maintained in the corrected data. However, the correction leads to a shift of the respective regime to a slightly lower level. In most cases, this results in a slightly higher deviation from the observed data compared to the BC1 data based on monthly values.

The relative overall differences between the reference regime and those generated by the corrected RCM data given in Table 6 show that best adjustment is gained by the models CCLM member 2 and 3 as well as REMO. This is due to the finer resolution and an already better representation of the seasonal course in the raw data. Furthermore, these values illustrate that in most cases qm leads to the best adjustment. The differences of this approach are in most cases considerably lower (e.g. RACMO M1 qm_m -9.4%, locy_y

Table 6
Relative difference [%] of mean difference in runoff regimes produced by the application of the reference data set of observed data and those produced using the bias corrected RCM data.

BC Method	Overall relative difference [%]						
	RACMO M1	RACMO M2	RACMO M3	CCLM M1	CCLM M2	CCLM M3	REMO
ls_m	−13.4	−10.2	−17.2	−10.3	−0.3	−4.6	4.3
qm_m	−9.4	−7.5	−17.8	−3.3	8.6	1.5	−6.8
loci_m	−13.8	−8.8	−16.8	−9.3	1.6	−2.0	2.3
ls_y	−14.2	−10.1	−18.5	−9.3	1.4	−4.3	4.7
qm_y	−15.3	−6.8	−19.7	−6.4	8.8	−0.4	−3.5
loci_y	−16.1	−12.0	−21.0	−8.6	2.9	−2.1	4.6

Table 7
Flow indicators applied for the BI-KLIM project to determine the effects of bias correction of RCMs on hydrological climate change signals.

Flow indicator	Explanation
LF	Low flow, lowest flow of the entire runoff time series
MLF	Mean low flow, mean of the lowest flows of each model year of the runoff time series
7LF2	7 day duration low flow with a return period of 2 years
MF	Mean flow, over all mean of the entire runoff time series
MHF	Mean high flow, mean of the highest flows of each model year of the runoff time series
HF	High flow, highest flow of the entire runoff time series
HF2	High flow with a 2-year return period
HF100	High flow with a 100-year return period

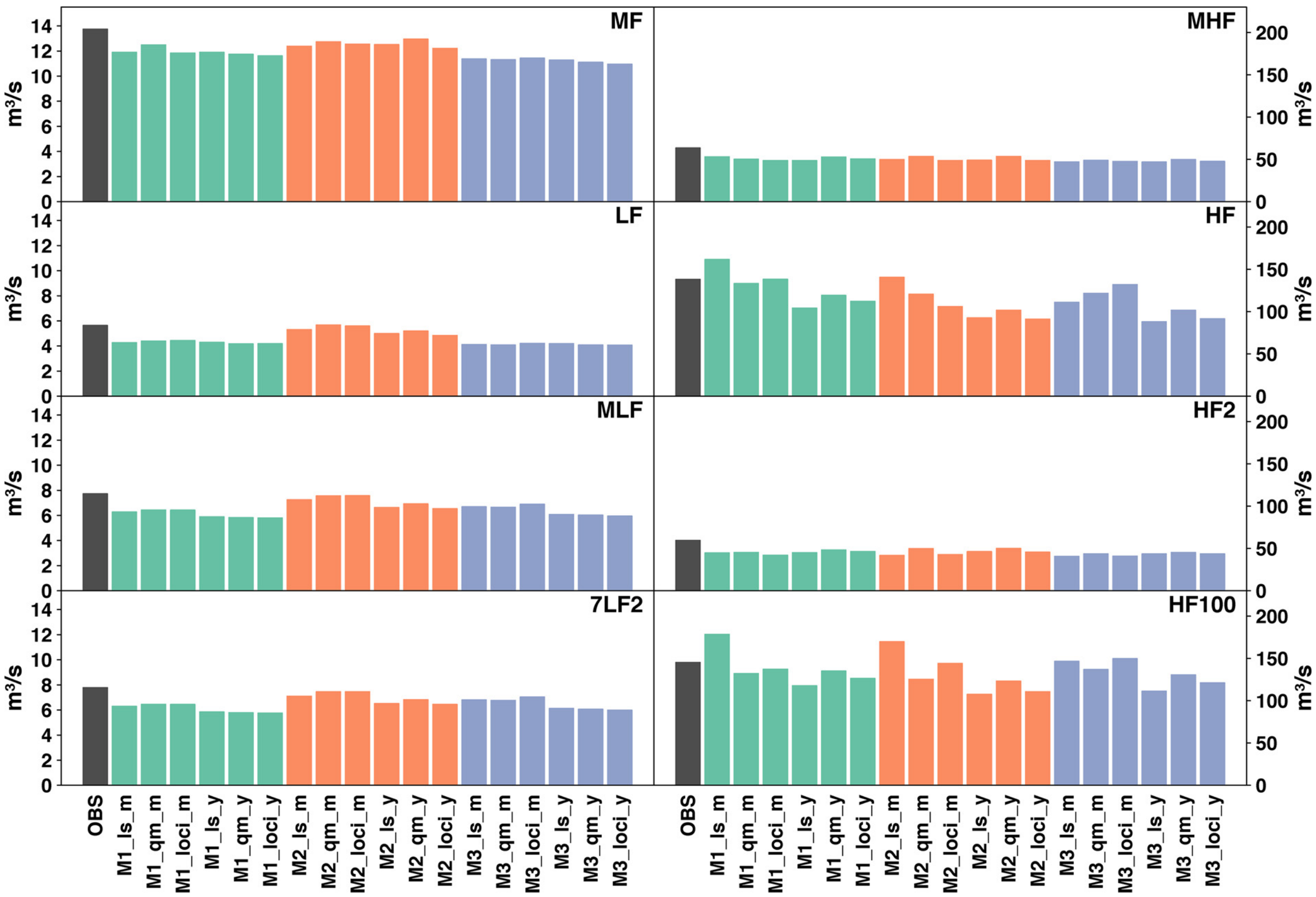


Fig. 6. Flow indicators for the gauge Offingen (Mindel river) for the reference period (1971–2000). The solid blue bar shows the reference values. The shaded bars represent the results of the model runs using bias corrected RCM data of the different members of the RACMO model (red: member 1, blue: member 2, green: member 3).

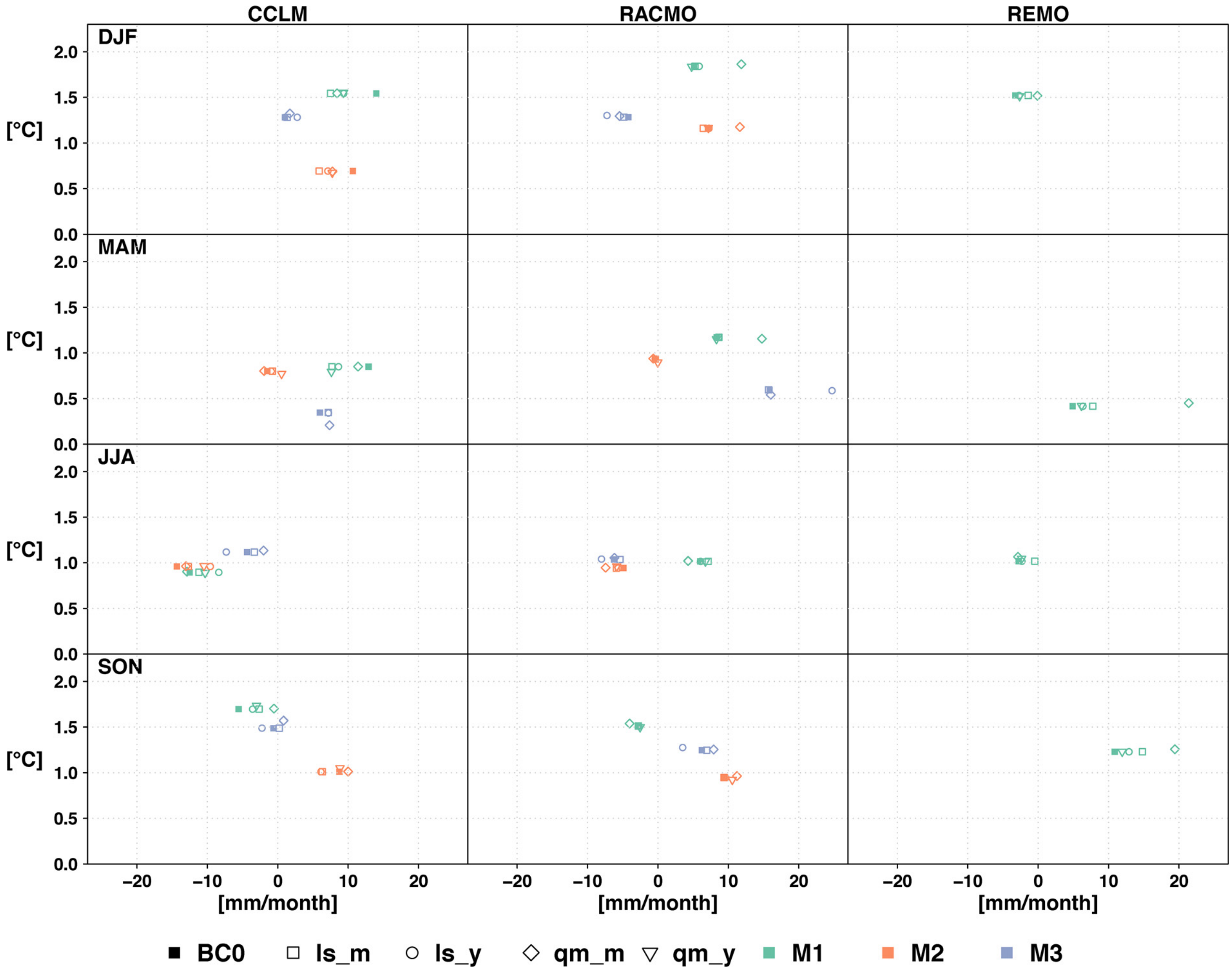


Fig. 7. Change signals of precipitation (x-axes) and temperature (y-axis) between the reference and future period for each season (DJF: winter, MAM: spring, JJA: summer, SON: fall) of raw (BC0) and bias corrected RCM data over the Offingen catchment.

-16.1%). Furthermore, yearly correction factors yield greater deviations for almost all RCMs. Also, the internal variations between the members of the CCLM and RACMO seem to be maintained by the multi member correction approach. The single set of correction factors for the REMO RCM however does not produce better results compared to the multi member approach, since the differences are comparable to those of the CCLM and RACMO (e.g. REMO qm_m -6.8% to RACMO qm_m -7.5%).

3.2. Changes in hydrological signals

The changes in hydrological signals are analyzed for the 'near future' scenario period ranging from 2021 to 2050 since data for the CCLM RCM is only available for this period. To determine the effects of bias correction of RCM data on the hydrological CCS we applied eight different flow indicators as described in Table 7. The extreme value statistics of the low and high flow indicators of a certain return period are based on the Pearson III distribution (DVWK, 1979, 1983).

The flow indicators for the reference period (1971–2000) show good agreement with observations for the hydrological data produced by bias corrected RCM time series in terms of mean flow, low flow and mean high flows compared to the reference (REF). However, the extreme high flow indicators show greater deviations. Fig. 6 exemplarily illustrates this behavior for the RACMO RCM and is representative for the results of the CCLM and REMO RCM as well. Whereas inter-member differences are very small for MF, LF, MLF, 7LF2, MHF, and HF2, the HF and HF100 depict significant variations between the different correction methods. These distinct differences in HF and HF100 for the various BC methods originate from their statistical characteristics. The HF index represents the highest runoff value of a chosen period (e.g. 30 years). Hence, the runoff simulations using BC data might not capture this specific value since the driving BC meteorology is possibly lacking a proper representation of extreme values. The HF100 is a statistical extrapolation based on yearly HF events of a certain period. Thus, the insufficient representation of HF values when using the BC meteorology directly affects the HF100 values. HF2 on the other hand is based on annual HF events. In this case an extrapolation is obsolete since the available runoff time series are sufficiently long (e.g. 30 years).

Fig. 7 shows the CCS for the different seasons (DJF, MAM, JJA, SON) of the raw and corrected RCMs. The different seasons exhibit

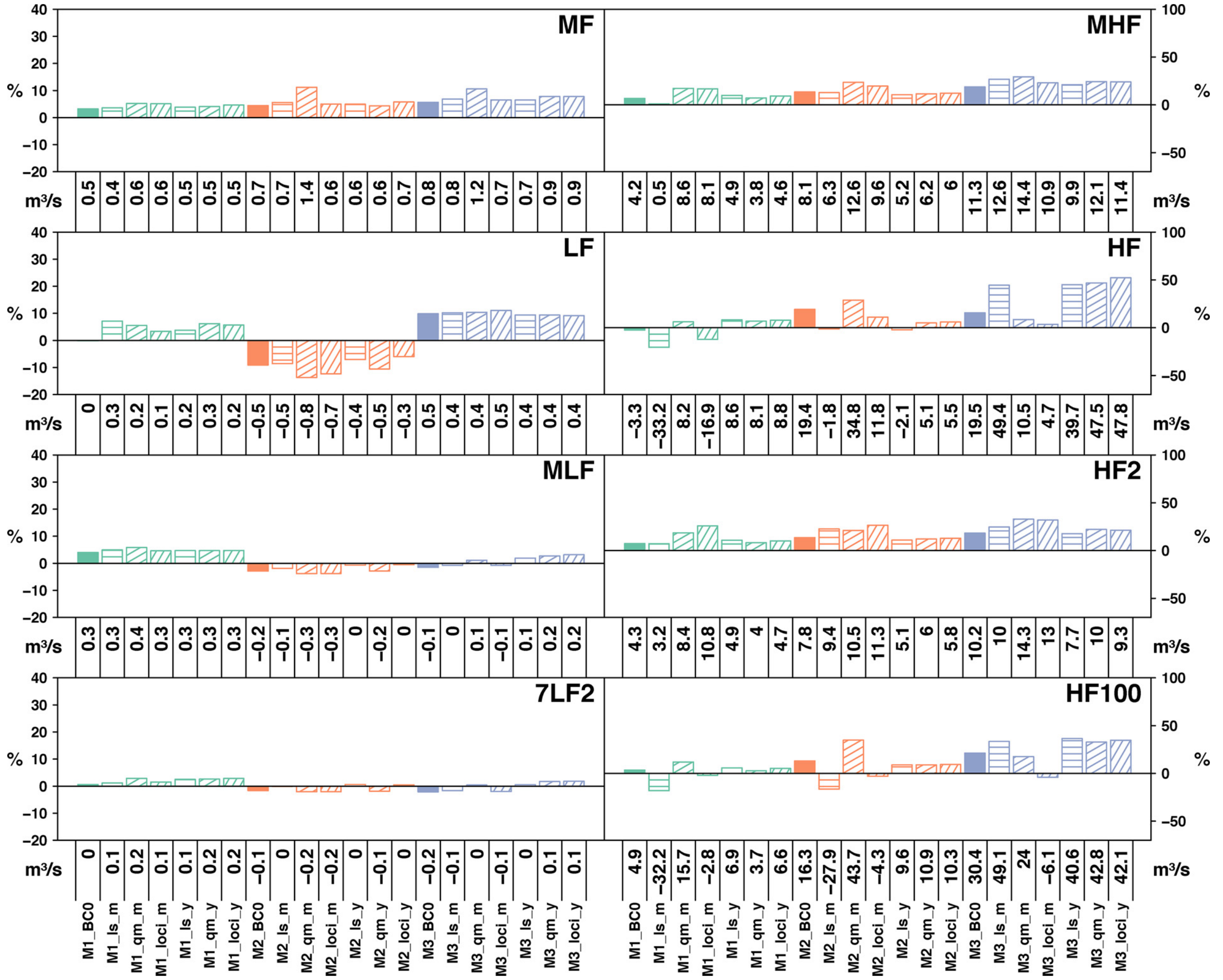


Fig. 8. Relative ([%]; bars) and absolute ([m³/s]; numbers below graphs) hydrological climate change signals for the various flow indicators for the three RACMO RCM members.

various changes in the signal. The internal variability between the members of the CCLM and RACMO RCMs is visible and maintained by the applied correction methods (a warm and moist signal remains warm and moist). Signals in summer (JJA) are rather small for the REMO RCM, while in winter and spring they are larger for all RCMs. Apart from CCLM M1 and M3, and RACMO M1 and M3 for qm_m the CCS of BC1 data for the winter period show little deviations from the BC0 signal. This is also visible for the other seasons and the CCLM and RACMO model for almost all BC methods. Greater deviations in CCS between BC0 and BC1 data are present in the spring and fall season. The shift in precipitation fields, as described earlier, might be accountable for this larger change in signals by the various RCM methods since this shift is adjusted by the correction as well.

The following graphs (Fig. 8–10) show the results of the CCS analysis for streamflow. While the bars illustrate the relative [%] change signal (i.e. relative difference between reference and future scenario on RCM-to-RCM basis) of the flow indicators, the numbers below represent their respective absolute values [m³/s]. It should be mentioned that the relative change signals might indicate a more severe change than the absolute value actually provides for; this is obviously pronounced for the low flow indicators. For the RACMO and CCLM RCMs the change signals of all three members are illustrated (member 1 red, member 2 blue, member 3 green). The solid bar represents the raw RCM (BC0) inherent climate change signal (CCS), the shaded bar the induced changes according to the bias corrected model data.

Fig. 8 illustrates the changes in the CCS of the original RCM and bias corrected model data for RACMO. The mean flow shows little to no difference between the change signals throughout all members regarding the absolute values (except for member 2 and 3 qm_m values, which is ascribed to a strong wet signal in winter, see Fig. 7) as well as the relative signals. Thus, in this case bias correction does not contribute to uncertainty in long-term water balance assessments (changes in mean flows under new climate conditions) since the CCS is not affected by the corrected data. Considering the LF, only member 1 and 2 show differences between the CCS of BC0 and BC1 (relative and absolute). Absolute and relative values of member 3 vary around the same magnitude. Furthermore, the LF depicts that natural variability between the three members is conserved by the multi member bias correction approach, with the 2nd member showing a negative signal whereas the other two members deviate positively. The absolute and relative CCS of the other low

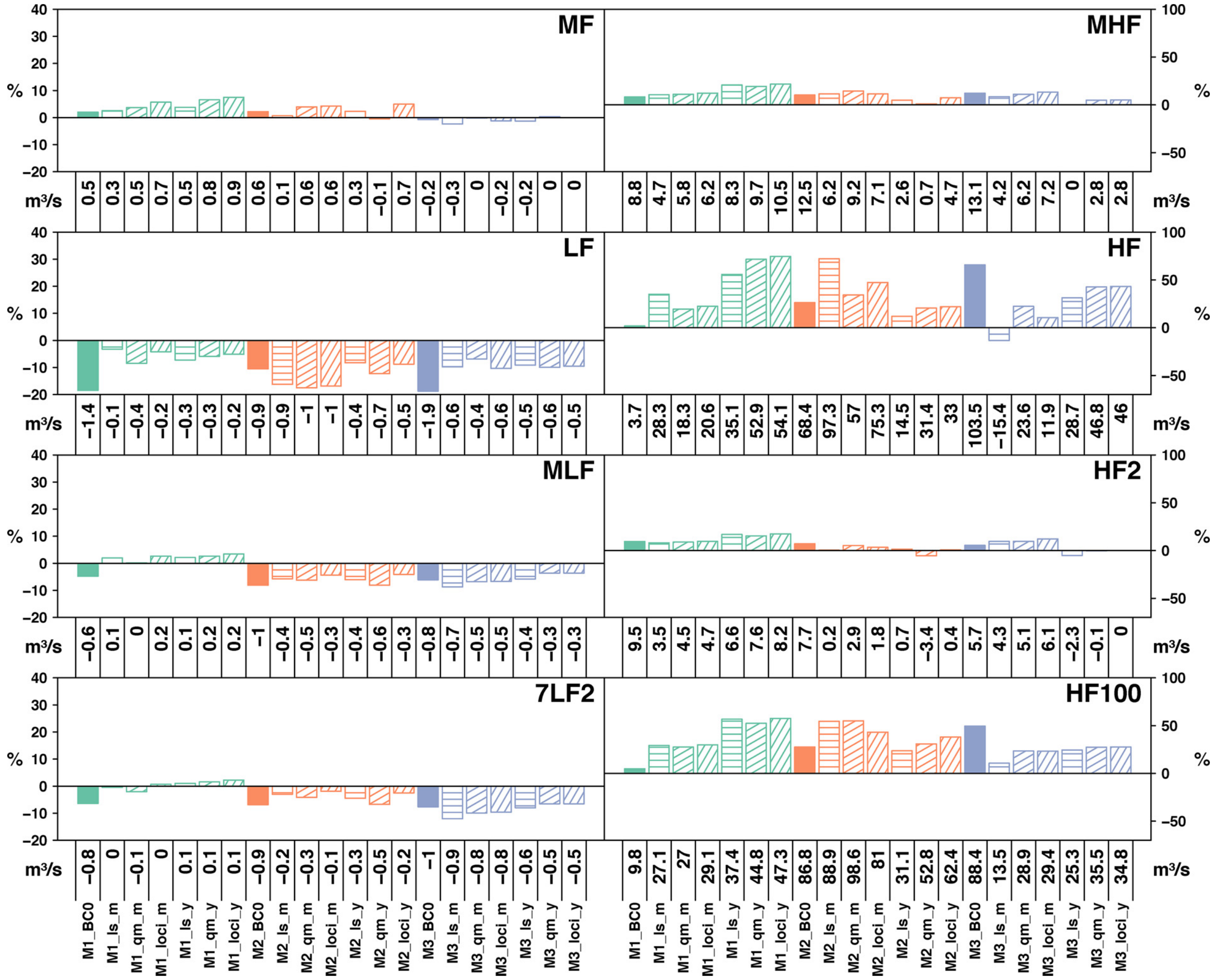


Fig. 9. Relative ([%]; bars) and absolute ([m³/s]; numbers below graphs) hydrological climate change signals for the various flow indicators for the three CCLM RCM members.

flow indicators (MLF, 7LF2) vary just slightly. Since the absolute CCS values are close to zero the shift in direction may be neglected. The high flow indicators however show more distinct differences between the BC0 (raw RCM) CCS and those produced using bias corrected data. The MHF exhibits the least distinctive absolute and relative deviations as well as the HF2. Furthermore, the different methods lead to large differences between BC0 CCS and BC1 CCS which holds especially for the 3rd member of the HF indicator. Here, the relative BC0 CCS is below 20% but the BC1 CCS of the LOCI methods are 5% and greater 50%, respectively.

The absolute value of the loci_y method exceeds the BC0 value by over 100% (19.5 m^3/s BC0 to 47.8 m^3/s loci_y). In contrast to the low flow indicators, the shift in CCS direction within a member is more severe. Regarding the HF100 member 1 ls_m shows a decrease by 32.2 m^3/s whereas the BC0 displays a slight increase of 4.9 m^3/s.

The differences in CCS of the flow indicators are more pronounced for the CCLM model results (Fig. 9). The MF shows varying relative changes of the CCS (which is a direct response to the input data for all seasons, see Fig. 7), but the absolute values differ only very little. Compared to the RACMO model, the absolute differences of the CCS for the low flow indicators between the BC1 and BC0 are higher for CCLM data. Only the 3rd member, using the monthly adjustment factors, shows similarities to the raw RCM. However, the absolute changes are very small. Again, the high flow indicators show major differences in CCS between the BC1 and BC0 values for both, relative and absolute numbers. These changes are less severe for the MHF and in some cases for the HF2. However, the BC1 data in most cases overestimate the CCS of HF and HF100 for the members 1 and 2 by up to 145% (HF: BC1_M1_loci_y 54.1 m^3/s; BC0_M1 3.7 m^3/s) and underestimate the CCS for the 3rd member.

The REMO model data induce the most severe changes in CCS of the flow indicators between the original RCM and the BC1 results (Fig. 10). Here almost every indicator switches from a negative signal in BC0 to a positive in BC1 or vice versa, except for the MHF. While the BC0 data depict a reduction in MF and all low flow indicators, the BC1 data mostly show a slight increase. The CCS of the HF and the HF100 indicator exhibit considerable differences between BC0 and BC1 data. While the raw RCM data for these indicators depict an increase of less than 20%, the bias-corrected data reveal a significant decrease between about 30% (HF100, qm_m) and over 50% (HF, ls_y). As mentioned in Section 2.2 the BC0 REMO inherent spatial offset of precipitation fields might influence the CCS. All

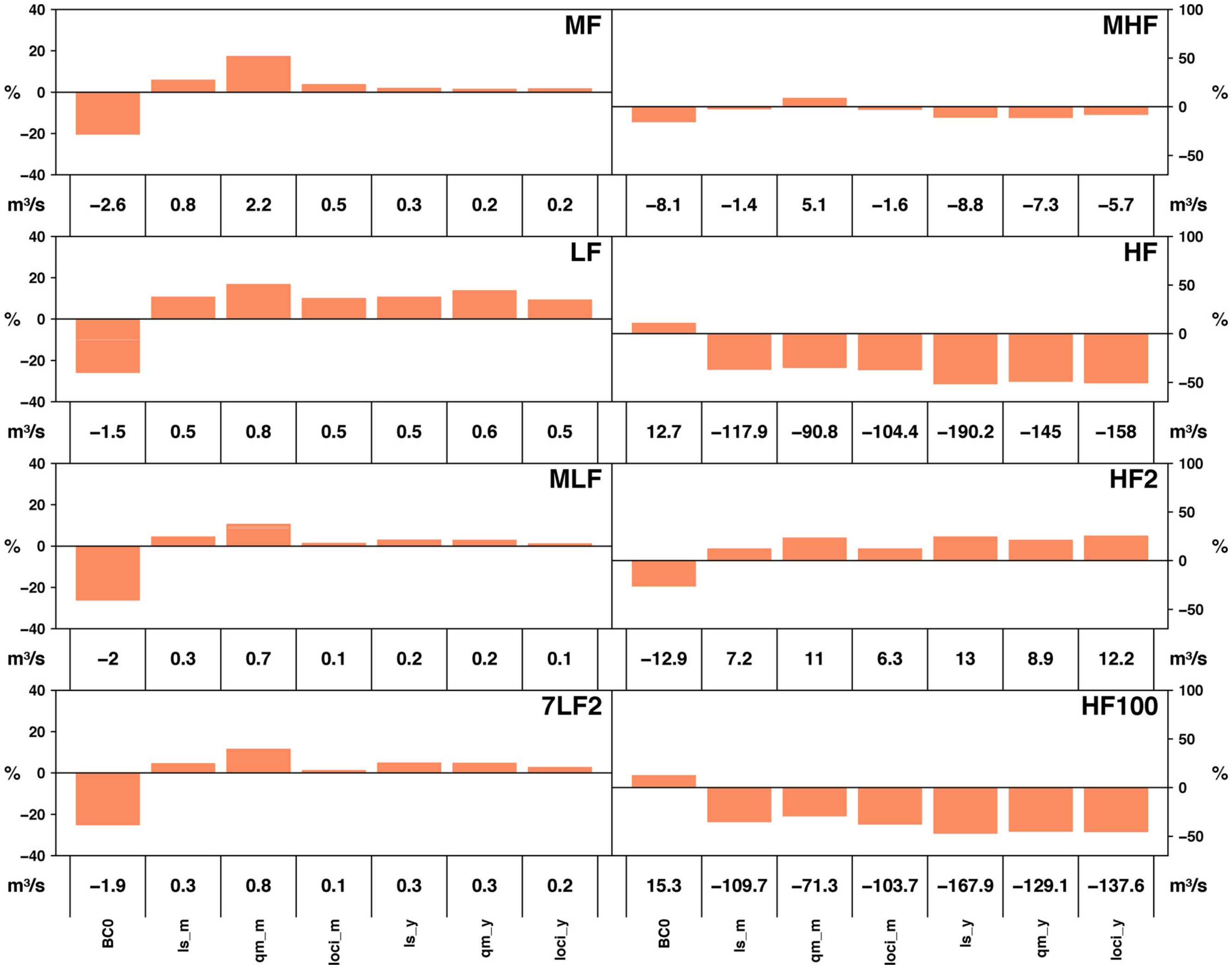


Fig. 10. Relative ([%]; bars) and absolute ([m³/s]; numbers below graphs) hydrological climate change signals for the various flow indicators for the REMO RCM.

bias correction methods remove this shift from the BC0 data.

4. Conclusions and discussion

The results of the hydrological modeling using the BC0 RCM data clarify the indispensable need for bias correction for climate impact studies, if the results significantly differ from observations and data are applied for subsequent hydrological applications (e.g. water management). The long term yearly flow regimes of the CCLM and REMO differ from the reference. However, apart from the winter season, the raw RACMO model shows a good regime representation. For the Mindel catchment, the different correction approaches account for good adjustment of the modeled runoff to the reference of observed data when applied to raw CCLM and REMO data. The regime simulated using the modified RACMO data on the other hand are at least comparable to the results using the raw data. In general, while in northern Bavaria the available models fit their respective reference equally well after correction (exemplarily shown in Fig. 11 for the gauge Kemmern, outlet of catchment 18, lower left map of Fig. 1), the RACMO model shows some greater differences in adjustment after the correction in the southern part of Bavaria.

Since the bias correction is performed on the RCM scale using the spatially aggregated reference data, localized small scale events within aggregation are also averaged and smoothed. Hence, this aggregation to the coarse RCM model resolution of 50 km is considered to be the major source for the partly huge deviations, especially in distinctive topography like the Alps. Considering the uncertainties added by applying bias correction to raw RCM data (e.g. by losing coherence between variables, assumption of temporal stationarity of correction factors, discrepancy in scales between RCM and observations) and the little effect it has on the RACMO data, the raw data might also be useful. However, judging from the indicators, the qm approach using monthly correction factors shows the best results and thus supports earlier suggestions by Themeßl et al. (2011). High flow indicators are an exception, which was also found by Muerth et al. (2012). However, it should be mentioned that the applied hydrological model WaSiM is not specifically calibrated for high flows. Thus, this must influence rare and single extreme high flow events like the HF and HF100. Combined with the BC1 RCMs, a sufficient match with the reference can hardly be achieved and might occur randomly. Such flood extremes are

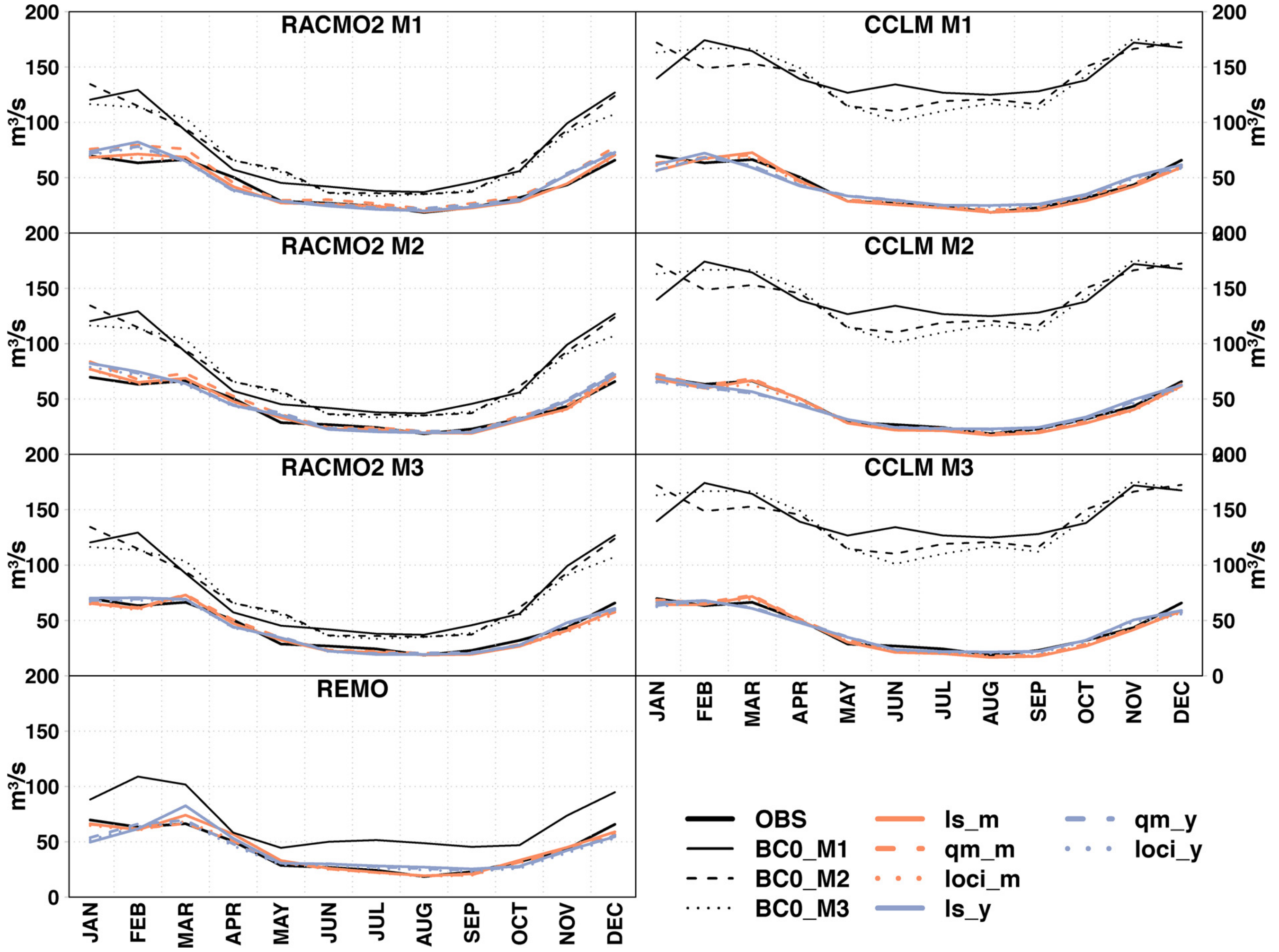


Fig. 11. Flow regimes of modeled runoff using observed (OBS), uncorrected (BC0) and bias corrected RCM data for the gauge Kemmern (Fig. 1, lower left map, outlet catchment 18).

usually triggered by extreme precipitation events at the far end of the cumulative distribution function or the highest percentiles of occurrence, for which the correction factors may only be able to provide insufficient approximations. IN conclusion, this study confirms that the qm approach applied to all meteorological variables results in a better representation of mean streamflow indicators across Bavaria than the other two investigated methods. Regarding extreme flow indicators (HF, HF100), these methods are still not able to reproduce the statistics of the observations at the upper end of the distribution for any Bavarian region. However, if extreme flow indicators are of particular interest (e.g. if a flood detention basin should be designed to store runoff up to a certain HF threshold) and other indicators (e.g. mean flow) are well represented by raw RCM data, Hattermann et al. (2014) suggest to correct discharge values by their return periods using extreme value statistics

The analysis for the gauge Offingen also shows that the bias correction of RCM data affects the CCS of hydrological indicators to an extent that may not be negligible for subsequent applications (e.g. hydrological modelling, water management or the design flood protection infrastructure). Differences in the relative CCS of mean flow indicators between raw and corrected data are small in most cases. The relative signals of BC1 low flow indicators show more severe deviations from the reference signal of the BC0 RCM data. This effect of bias correction on the CCS of mean indicators is also shown by Stagl and Hattermann (2015) and Muerth et al. (2012). Hence, in this case, raw RCM data can be considered useful, unless overall characteristics of these data (absolute values, seasonality) significantly differ from those of the observations. In this case, the RCM data might not be suitable for climate change impact assessment. Absolute values in CCS show less difference and are mostly of the same magnitude. This applies at least for the mean and low flow indicators. The bias correction depicts a stronger impact on the CCS for high flow indicators. Despite regional disparities in absolute quantities, this holds true for other catchments of the hydrological Bavaria that was analyzed for streamflow. Fig. 12 emphasizes this result showing the CCS for the RACMO RCM at the gauge Kemmern.

The REMO RCM shows significant deviations in the CCS across all indicators due to the correction of the inherent spatial drift of precipitation fields. Furthermore, since only a single member was available, a particular extreme event within this realization (e.g. high precipitation event during spring) affects the bias correction as well as the CCS. The member bound derivation of correction factors (i.e. deriving the factors using one member of a RCM only) might result in a misleading adjustment for this specific season. Hence, this may lead indirectly to a fraud of the CCS, which could be avoided by a multi member approach if more realizations would

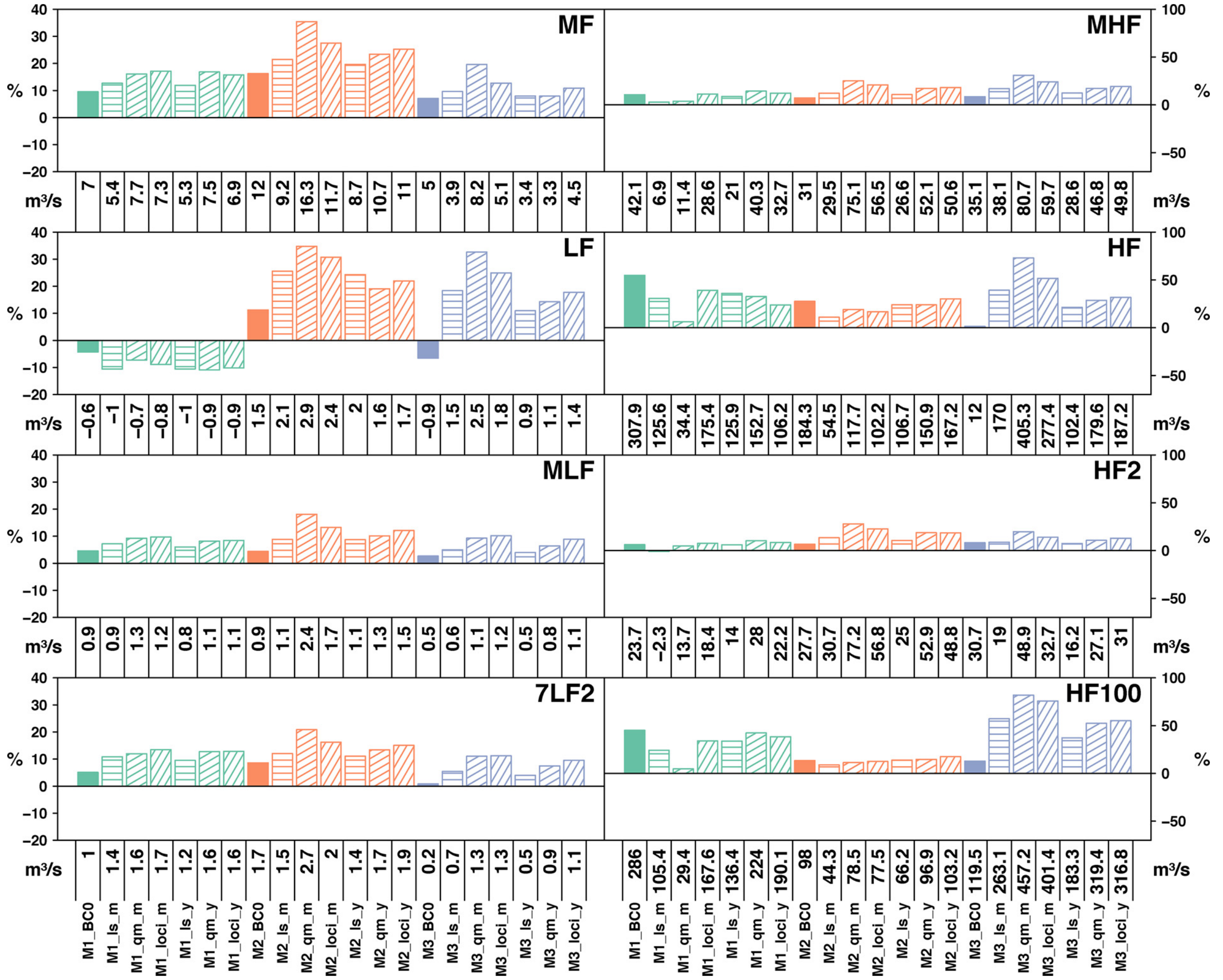


Fig. 12. Relative ([%]; bars) and absolute ([m³/s]; numbers below graphs) hydrological climate change signals for the various flow indicators for the three RACMO RCM members for the gauge Kemmern.

have been available showing different seasonal values. Therefore, this study shows that the application of bias corrected RCM data for hydrological modeling has an impact on the CCS of streamflow indicators derived for catchments situated in southern and northern Bavaria (catchment 5 and 18, Fig. 1, lower left map). Furthermore, the impacts on the CCS of extreme high flow indicators can be severe (up to or greater than 100%). Similar effects have been found by Cloke et al. (2013) over a catchment situated in England. Hence, the applicability of bias correction approaches for extreme values is still questionable and further development should be made to account for extreme value statistics.

Funding

This work was supported by the Bavarian Environment Agency (LfU).

Conflict of interest

I wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

References

Berg, P., Wagner, S., Kunstmann, H., Schädler, G., 2013. High resolution regional climate model simulations for Germany: part I—validation. Clim. Dyn. 40 (1–2), 401–414.
Chen, J., Brissette, F.P., Leconte, R., 2011. Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. J. Hydrol. 401 (3), 190–202.
Christensen, J.H., Boberg, F., Christensen, O.B., Lucas-Picher, P., 2008. On the need for bias correction of regional climate change projections of temperature and precipitation. Geophys. Res. Lett. 35 (20).

Cloke, H.L., Wetterhall, F., He, Y., Freer, J.E., Pappenberger, F., 2013. Modelling climate impact on floods with ensemble climate projections. Q. J. R. Meteorol. Soc. 139 (671), 282–297.
Dai, A., 2001. Global precipitation and thunderstorm frequencies. Part II: diurnal variations. J. Clim. 14 (6), 1112–1128.
Dai, A., 2006. Precipitation characteristics in eighteen coupled climate models. J. Clim. 19 (18), 4605–4630.
DVWK, 1979. Empfehlungen zur Berechnung der Hochwasserwahrscheinlichkeit. DVWK Hamburg, Berlin.
DVWK, 1983. Niedrigwasseranalyse Teil I: Statistische Untersuchung des Niedrigwasser-Abflusses. DVWK Hamburg, Berlin.
Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., Liebert, J., 2012. HESS opinions" Should we apply bias correction to global and regional climate model data?". Hydrol. Earth Syst. Sci. 16 (9), 3391.
Elía, R., Côté, H., 2010. Climate and climate change sensitivity to model configuration in the Canadian RCM over North America. Meteorologische Zeitschrift 19 (4), 325–339.
Foltyn, M., Steinbauer, A., Kopp, B., 2017. Niedrigwasser in Bayern - Grundlagen. Veränderung und Auswirkungen, Augsburg.
Frick, C., Steiner, H., Mazurkiewicz, A., Riediger, U., Rauthe, M., Reich, T., Gratzki, A., 2014. Central European high-resolution gridded daily data sets (HYRAS): mean temperature and relative humidity. Meteorologische Zeitschrift 15–32.
Gandin, L., 1965. Objective Analysis of Meteorological Fields. Israel Program for Scientific Translation, Jerusalem, Jerusalem.
Giorgi, F., Jones, C., Asrar, G.R., 2009. Addressing climate information needs at the regional level: the CORDEX framework. World Meteorol. Organiz. (WMO) Bull. 58 (3), 175.
Göttel, H., 2009. Einfluss der nichthydrostatischen Modellierung und der Niederschlagsverdriftung auf die Ergebnisse regionaler Klimamodellierung.
Hagemann, S., Chen, C., Haerter, J.O., Heinke, J., Gerten, D., Piani, C., 2011. Impact of a statistical bias correction on the projected hydrological changes obtained from three GCMs and two hydrology models. J. Hydrometeorol. 12 (4), 556–578.
Hattermann, F.F., Huang, S., Burghoff, O., Willems, W., Österle, H., Büchner, M., Kundzewicz, Z., 2014. Modelling flood damages under climate change conditions–a case study for Germany. Nat. Hazards Earth Syst. Sci. 14 (12), 3151–3168.
Kjellström, E., Boberg, F., Castro, M., Christensen, J.H., Nikulin, G., Sánchez, E., 2010. Daily and monthly temperature and precipitation statistics as performance indicators for regional climate models. Clim. Res. 44 (2-3), 135–150.
Kleinn, J., Frei, C., Gurtz, J., Lüthi, D., Vidale, P.L., Schär, C., 2005. Hydrologic simulations in the Rhine basin driven by a regional climate model. J. Geophys. Res.: Atmos. 110 (D4).
Kotlarski, S., Keuler, K., Christensen, O.B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., 2014. Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. Geosci. Model. Dev. 7 (4), 1297–1333.
Lenderink, G., Buishand, A., van Deursen, W., 2007. Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach. Hydrol. Earth Syst. Sci. 11 (3), 1145–1159.
Majone, B., Bovolo, C.I., Bellin, A., Blenkinsop, S., Fowler, H.J., 2012. Modeling the impacts of future climate change on water resources for the Gállego river basin (Spain). Water Resour. Res. 48 (1).
Maraun, D., 2016. Bias correcting climate change simulations-a critical review. Curr. Clim. Change Rep. 2 (4), 211–220.
Maraun, D., Wetterhall, F., Ireson, Am, Chandler, R.E., Kendon, E.J., Widmann, M., Brienen, S., Rust, H.W., Sauter, T., Themeßl, M., 2010. Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. Rev. Geophys. 48 (3).
Marke, T., 2008. Development and Application of a Model Interface to Couple Land Surface Models With Regional Climate Models for Climate Change Risk Assessment in the Upper Danube Watershed, Munich.
Mpelasoka, F.S., Chiew, F.H.S., 2009. Influence of rainfall scenario construction methods on runoff projections. J. Hydrometeorol. 10 (5), 1168–1183.
Mudelsee, M., Chirila, D., Deutschländer, T., Döring, C., Haerter, J., Hagemann, S., Hoffmann, H., Jacob, D., Krahe, P., Lohmann, G., 2010. Climate Model Bias Correction und die deutsche Anpassungsstrategie. Mitteilungen Deutsche Meteorologische Gesellschaft, pp. 2–7 03/2010.
Muerth, M., Gauvin St-Denis, B., Ricard, S., Velázquez, J.A., Schmid, J., Minville, M., Caya, D., Chaumont, D., Ludwig, R., Turcotte, R., 2012. On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff. Hydrol. Earth Syst. Sci. Disc. 10205–10243.
Nakicenovic, N., 2000. In: Nakicenovic, Nebojsa, Swart, Robert (Eds.), Special Report on Emissions Scenarios, ISBN 521804930, 612.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—A discussion of principles. J. Hydrol. 10 (3), 282–290.
Pöhler, H., Schultze, B., Stadelbacher, V., Scherzer, J., 2009. Vorhaben KLIWA: Wasserhaushaltsmodellierung für die südlichen Zuflüsse der Donau in Bayern / Kalibrierung und Validierung. UDATA - Umweltschutz und Datenanalyse, Neustadt.
Pöhler, H., Wendel, S., Schultze, B., Karl, S., Scherzer, J., 2010. Vorhaben KLIWA: Klimawandel und Wasserhaushalt: Teilprojekt AdaptAlp: Wasserhaushaltsmodellierung Inn und neue Niederschlagsregionalisierung Südbayern. UDATA - Umweltschutz und Datenanalyse, Neustadt.
Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., Gratzki, A., 2013. A Central European precipitation climatology–Part I: generation and validation of a high-resolution gridded daily data set (HYRAS). Meteorologische Zeitschrift 22 (3), 235–256.
Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kirchner, I., Kornblueh, L., Manzini, E., 2003. The Atmospheric General Circulation Model ECHAM 5. PART I: Model Description.
Rößler, O., Löffler, J., 2010. Potentials and limitations of modelling spatio-temporal patterns of soil moisture in a high mountain catchment using WaSiM-ETH. Hydrol. Process. 24 (15), 2182–2196.
Schmidli, J., Frei, C., Vidale, P.L., 2006. Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods. Int. J. Climatol. 26 (5), 679–689.
Schulla, J., 2012. Model Description WaSiM (Water Balance Simulation Model), Completely Revised Version 2012. Hydrology Software Consulting, Zürich, Switzerland.
Sennikovs, J., Bethers, U. (Eds.), 2009. Statistical Downscaling Method of Regional Climate Model Results for Hydrological Modelling. Citeseer.
Stagl, J.C., Hattermann, F.F., 2015. Impacts of climate change on the hydrological regime of the Danube River and its tributaries using an ensemble of climate scenarios. Water 7 (11), 6139–6172.
Teutschbein, C., Seibert, J., 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods. J. Hydrol. 456, 12–29.
Themeßl, J.M., Gobiet, A., Leuprecht, A., 2011. Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. Int. J. Climatol. 31 (10), 1530–1544.
Themeßl, M.J., Gobiet, A., Heinrich, G., 2012. Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. Clim. Change 112 (2), 449–468.
van der Linden, P., Mitchell, J.E., 2009. ENSEMBLES: Climate Change and Its Impacts-Summary of Research and results from the ENSEMBLES Project.
van Meijgaard, E., van Ulft, L.H., van de Berg, W.J., Bosveld, F.C., van den Hurk, B., Lenderink, G., Siebesma, A.P., 2008. The KNMI Regional Atmospheric Climate Model RACMO Version 2.1. Koninklijk Nederlands Meteorologisch Instituut.
van Vuuren, D.P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G.C., Kram, T., Krey, V., Lamarque, J.-F., 2011. The representative concentration pathways: an overview. Clim. Change 109 (1-2), 5.
Wagner, S., Berg, P., Schädler, G., Kunstmann, H., 2013. High resolution regional climate model simulations for Germany: part II—projected climate changes. Clim. Dyn. 40 (1–2), 415–427.

2.3 Paper III: Assessing the impact of climate change on high return levels of peak flows in Bavaria applying the CRCM5 Large Ensemble

Reference: Willkofer, F., Wood, R.R., and Ludwig, R. (2023). Assessing the impact of climate change on high return levels of peak flows in Bavaria applying the CRCM5 Large Ensemble. Hydr. & Earth Sys. Sci. [submitted]

Transition to paper III: This publication describes the dynamics of 100-year flood events in a changing climate. The approach introduced in this paper employs climate model data from 50 members of the Canadian Regional Climate Model, version 5 Large Ensebmle (CRCM5-LE) - using the RCP8.5 radiative forcing - as a driver for the hydrological model introduced in paper I, resulting in the first process-based hydrological large ensemble for Bavarian rivers. The climate data were bias-corrected using the quantile-mapping method which according to the findings in paper II showed the best adjustment between simulations with observations and RCM data for the investigated hydrological indicators. First, this paper shows the advantage of the new approach using a single model initial condition large ensemble (SMILE) such as the CRCM5-LE for the robust estimation of extreme discharge values. Finally, the impact of climate change on the dynamics of the 100-year flood for Bavarian rivers in terms of frequency and intensity is presented for different discharge regimes according to Poschlod et al. (2020).

Author's contribution: FW developed the concept of this study including methods, investigation, and visualization, performed the formal analysis and validation, and prepared the original draft. FW and RRW were responsible for data curation and software development. RL acquired the funding for the presented research, provided the required resources, was responsible for the project administration, and supervised the presented research.

Status: submitted
Journal: Hydrology and Earth System Sciences (HESS)
Impact factor: 6.3

Assessing the impact of climate change on high return levels of peak flows in Bavaria applying the CRCM5 Large Ensemble

Florian Willkofer[1], Raul R. Wood[1,2,3], Ralf Ludwig[1]

[1] Department of Geography, Ludwig-Maximilians-Universität München, Munich, 80333, Germany
[2] WSL Institute for Snow and Avalanche Research SLF, Davos Dorf, 7260, Switzerland
[3] Climate Change, Extremes and Natural Hazards in Alpine Regions Research Center CERC, Davos Dorf, 7260, Switzerland

Correspondence to: Florian Willkofer (florian.willkofer@campus.lmu.de)

Abstract. Severe floods with extreme return periods of 100 years and beyond have been observed in several large rivers in Bavaria in the last three decades. Flood protection structures are typically designed based on a 100-year event, relying on statistical extrapolations of relatively short observation time series while ignoring potential temporal non-stationarity. However, future precipitation projections indicate an increase in the frequency and intensity of extreme rainfall events, as well as a shift in seasonality. This study aims to examine the impact of climate change on the 100-year flood (HF_{100}) events on 98 hydrometric gauges within the Hydrological Bavaria. A hydrological climate change impact (CCI) modelling chain consisting of a regional single model initial condition large ensemble (SMILE) and a single hydrological model was created. The 50 equally probable members of the CRCM5-LE were used to drive the hydrological model WaSiM to create a hydro-SMILE. As a result, a database of 1,500 model years (50 members x 30 years) per investigated time period was established for extreme value analysis (EVA) to illustrate the benefit of the hydro-SMILE approach for a robust estimation of the HF_{100} based on annual maxima (AM), and to examine the CCI on the frequency and intensity of HF_{100} in different discharge regimes under a strong emission scenario (RCP8.5). The results demonstrate that the hydro-SMILE approach provides a clear advantage for a robust estimation of the HF_{100} using empirical probability on 1,500 AM compared to its estimation using the generalized extreme value (GEV) distribution on 1,000 samples of typically available time series size of 30, 100, and 200 years. Thereby, by applying the hydro-SMILE framework the uncertainty from statistical estimation can be reduced. The CCI on the HF_{100} varies for different flow regimes, with snowmelt-driven catchments experiencing severe increases in frequency and intensity, leading to unseen extremes that impact the distribution. Pluvial regimes show a lower intensification or even decline. The study highlights the added value of using hydrological SMILEs to project future flood return levels.

1 Introduction

The devastating force of floods poses a significant threat to infrastructure, livestock, and human life. In Germany, two of the most severe floods in the last three decades were the 2002 and 2013 flood events (along with other major events in 1999, 2005, and 2016) (Thieken et al., 2016; Blöschl et al., 2013). The 2002 and 2013 events caused a total of about 17 billion Euros in economic damage due to their large spatial extent and high water levels, with the 2013 flood considered the most extreme event in the last sixty years (Thieken et al., 2016). However, different climatic and catchment conditions caused these events, with the 2002 event resulting from intense rainfall leading to flash floods across multiple small catchments, and the 2013 event due to high antecedent soil moisture from long-lasting precipitation followed by more moderate but spatially widespread rainfall (Thieken et al., 2016). In addition to precipitation intensity, other flood drivers such as antecedent soil moisture conditions, snowmelt, as well as flood driving processes determined by catchment and river characteristics contribute to the non-linearity of the hydrological response to extreme precipitation events (Blöschl et al., 2015). Recent studies analyzing European flood events over the last five decades suggest an increase in the intensity and frequency of high flows and flood events depending on the event type and region (Blöschl et al., 2019; Bertola et al., 2020; Blöschl et al., 2015). However, this trend depends on the time frame considered for the analysis, and the evaluation period remains crucial for either the estimation or the development of high return periods (Blöschl et al., 2015; Schulz and Bernhardt, 2016). Precipitation (heavy precipitation and long-lasting rainfall) and snowmelt (in regions with snowmelt-governed regimes) remain the primary natural causes of flooding, with other influences (e.g., catchment characteristics, antecedent catchment conditions, compound events with snow- or glacier melt) and snowmelt becoming less important once a certain threshold of extreme precipitation is exceeded (Brunner et al., 2021b).

According to the sixth Intergovernmental Panel on Climate Change (IPCC) Assessment Report, there is high confidence that a warmer climate will intensify wet weather and climate conditions affecting flooding (IPCC, 2021). Even with a 1.5 °C warming limit under the Paris agreement, heavy precipitation, along with extreme discharge events, is likely to intensify in Europe, with increasing confidence above 2 °C warming (IPCC, 2021).

For most discharge gauges, observational records begin in the 19th century or even later (Blöschl et al., 2015). Although most of these observations offer sufficiently long time series of data for estimating peak flows of moderate return periods, they still hinder a robust statistical estimation of extreme return periods, such as the 100-year flood and above. These types of extreme hydrological events are required for structural flood protection and risk management (Wilhelm et al., 2022; Brunner et al., 2021a; Blöschl et al., 2019). Brunner et al. (2021a) illustrate the challenges in modeling and predicting high flows due to data availability, process representation, human influences, and prediction.

Recently, single model initial condition large ensembles (SMILE) have emerged as a powerful tool to enhance statistical analysis of extremes in climatological behavior (von Trentini et al., 2020; Wood and Ludwig, 2020; Wood et al., 2021; Aalbers et al., 2018; Martel et al., 2020). Unlike other common ensembles of different global or regional climate model (GCM/RCM) combinations, SMILEs comprise multiple equiprobable realizations (members) of a single GCM or GCM/RCM combination that differ only in their initial conditions, representing the chaotic nature of the climate system (Arora et al., 2011; Fyfe et al., 2017; Kirchmeier-Young et al., 2017; Sigmond et al., 2018; Leduc et al., 2019). The actual model structure, physics, parameterization, external forcings are preserved. Thus, SMILEs offer a profound database for analyzing internal (or natural) climate variability (Wood and Ludwig, 2020; Martel et al., 2018), separating natural variability from an actual change signal (Aalbers et al., 2018; Wood and Ludwig, 2020), and extreme events (Wood et al., 2021; Martel et al., 2018). Applying SMILEs for hydrological modelling allows for the creation of a so-called hydro-SMILE, which in turn allows for the exploitation of vast data for the analysis of the hydrological response of catchments to extraordinary precipitation events.

This approach of high spatio-temporal resolution for climate and hydrological modelling is computationally demanding. However, considering spatially refined catchment features (e.g., slopes, soil characteristics, land use), precise values due to higher temporal resolution, and the application of a SMILE for hydrological modelling supports an enhanced representation of extreme values within models. Thus, this study focuses on the major Bavarian river basins (upper Danube, Main, Inn) with all their tributaries.

In this study, a climatological SMILE is employed to drive a physically based hydrological model with high spatio-temporal resolution for the major Bavarian river catchments. The resulting hydro-SMILE is used to answer the following questions:

a) Is there a benefit applying a SMILE for hydrological impact modelling regarding the estimation of high flows of large return periods?
b) How might climate change affect the dynamics in frequency and intensity of extreme discharges?

In this study, we focus on the 100-year flood event to answer both questions. Therefore, the study area is first introduced, followed by an overview of the climatological SMILE post-processing to meet the requirements for the hydrological modelling. The hydrological model setup used to produce the hydro-SMILE along an evaluation of its performance are then presented. The subsequent section describes the approaches taken to illustrate the benefit of a hydro-SMILE for the estimation of peak flow with high return periods and to assess the influence of climate change on their intensity and frequency. Finally, the results of the analysis are then presented and discussed, followed by concluding remarks.

2 Study Area, Data, and Methods

2.1 Study Area

This study focuses on the major Bavarian rivers, including the upper Danube upstream of Achleiten, Main, Inn, and upstream tributaries of the Elbe, as well as their smaller and larger tributaries originating from adjacent states (Bade-Württemberg, Hessen, Thuringia) and countries (Austria, Switzerland, Italy, Czech Republic). As a result, the catchments of these rivers extend beyond the political borders of Bavaria (Figure 1). The entirety of these catchments is referred to as the Hydrological Bavaria in this study.

The Hydrological Bavaria covers approximately 100,000 km² and features a diverse landscape ranging from the Alps (with the highest point being Piz Bernina at 4049 meters above sea level; m.a.s.l) and the alpine foreland in the south to the southern German escarpment in the north of the study area (with the lowest point being 90 m.a.s.l at Frankfurt-Osthafen) and the eastern mountain ranges to the east (Willkofer et al., 2020; Poschlod et al., 2020). The complexity of these landscapes and different climatological conditions (up to 1100 mm precipitation sums in the north, 2500 mm in the south; an average temperature of 10 °C in the north, down to 5 °C (-8 °C on alpine summits) in the south results in a variety of runoff regimes (Poschlod et al., 2020).

The discharge of many rivers within the Hydrological Bavaria is influenced by artificial retention structures (i.e., dams, retention basins), naturally formed lakes, or transfer systems (drinking water supply, low flow elevation) (Willkofer et al., 2020). The major river catchments were divided into a total of 98 smaller sub-catchments based on a common interest in flood protection and a more detailed variation in catchment characteristics, using a selection of gauges (Willkofer et al., 2020).

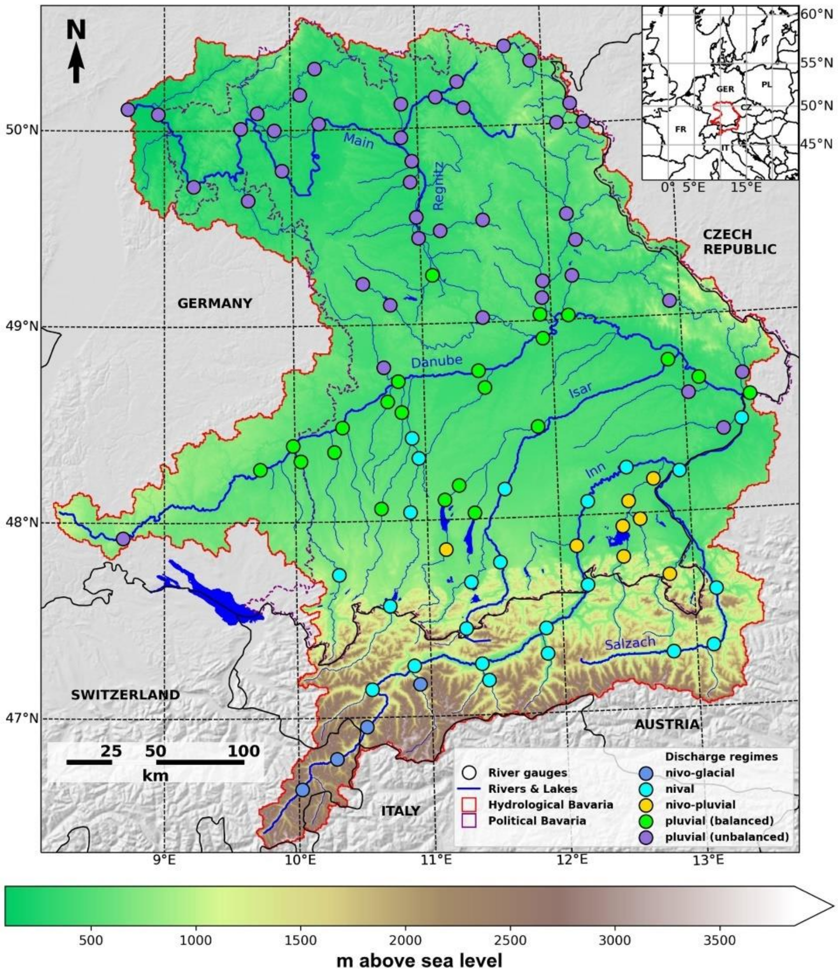


Figure 1: Map showing the elevation of the Hydrological Bavaria (red line) which comprises the political Bavaria (dashed purple line) and the 98 hydrometric gauges used in this study as well as their respective discharge regime type (colored dots) at their respective rivers (blue lines).

2.2 Data and Methods

To assess the impact of climate change on extreme return periods of peak flows, the hydroclimatic modeling chain illustrated in Figure 2 was introduced within the scope of the ClimEx project (Climate Change and Hydrological Extreme Events, www.climex-project.org). This common chain is divided into a climate and a hydrological impact section and covers three spatial scales (GCM scale, RCM scale, hydrological model scale) with increasing resolution along the chain.

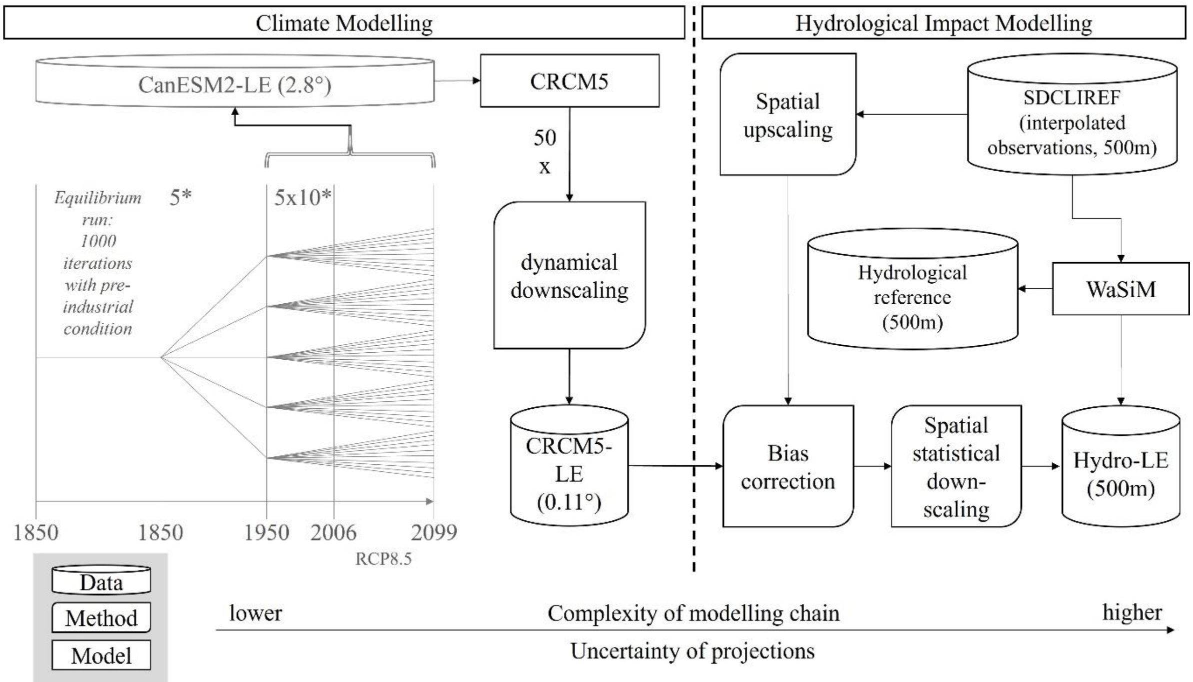


Figure 2: The ClimEx modelling chain uses the CanESM2 large ensemble (LE, gray, not created within the ClimEx project) to generate the CRCM5-LE. The CRCM5-LE is then used to explore the impacts of climate change on the hydrology of the Hydrological Bavaria through a hydrological large ensemble (Hydro-LE). The CRCM5-LE represents a SMILE, consisting of a single model that downscales output from the employed ESM using slight differences in the initialization.

Since the introduced model chain requires a vast number of computational resources, the ClimEx project employed the high-performance computing systems of the Leibniz Supercomputing Centre (LRZ) as well as its technical and consultative support to migrate and adapt software and data to its systems, facilitate calculations, and provide an extensive amount of storage to archive the data and make them available to the scientific community (data available at https://www.climex-project.org).

2.2.1 Climate data

A SMILE composed of 50 independent members of the Canadian Earth System Model, version 2 (CanESM2) large ensemble (LE) was used as a base for all further analysis. The CanESM2-LE was produced by the Canadian Centre for Climate Modelling and Analysis (CCCma) and described in previous publications (Fyfe et al., 2017; Kirchmeier-Young et al., 2017; Arora et al., 2011; Leduc et al., 2019). All members of the CanESM2-LE used natural and anthropogenic forcings for the historical period from 1950 to 2005 and the representative concentration pathway 8.5 (RCP8.5; van Vuuren et al., 2011) emission scenario from 2006 to 2099 (Kirchmeier-Young et al., 2017; Leduc et al., 2019; Fyfe et al., 2017; Sigmond et al., 2018). The individual members differ only in their initial conditions rather than changes in model structure, physics, or parameters, and therefore offer a range of internal or natural variability of the climate system at a global scale.

These 50 members were dynamically downscaled from ~2.85° (T63; ≈ 310 km) to 0.11° (≈ 12 km) using the Canadian Regional Climate Model, version 5 (CRCM5; Martynov et al., 2013; Šeparović et al., 2013) over two spatial domains, the European and the northeastern North American domains (Leduc et al., 2019). As with the CanESM2-LE, variations between the individual members were obtained by unique initial conditions for each member, thus providing a range of internal or natural variability on a regional scale. The resulting CRCM5 large ensemble (CRCM5-LE; Leduc et al., 2019) of 50 transient members provides the basis for assessing the impact of climate change on hydro-meteorological extreme events for the Hydrological Bavaria. Furthermore, the individual members of the CRCM5-LE are considered independent for the hydrological evaluation period from 1981 to 2099, as the analysis of variations in temperature and precipitation over land and ocean shows (Leduc et al., 2019). A comparison between the CRCM5-LE and the E-OBS observational gridded dataset (Haylock et al., 2008) at the CRCM5 grid revealed biases for a historical period between 1980 and 2012, showing regional and seasonal variations in magnitude over Europe (Leduc et al., 2019).

Since this bias was considered to affect the behavior of the outputs of the hydrological model due to shifts in seasonality and intensity, a bias correction was applied. The required meteorological data of precipitation, air temperature, relative air humidity, incoming shortwave radiation, and wind speed were adjusted to a meteorological reference of interpolated 3-hourly station data (Sub-Daily Climate Reference, SDCLIREF; Ludwig et al., 2019) on the RCM scale using an adaptation of the quantile-mapping approach after Mpelasoka and Chiew (2009). This approach as described in Willkofer et al. (2018) involved using multiplicative or additive correction factors, and was further adapted for using 3-hourly correction factors for every quantile and month. To preserve an internal spread between the members, a single set of factors was deduced from a combination of all 50 members. Despite bias correction being often considered inevitable for climate change impact studies (Gampe et al., 2019), numerous studies argue about the benefits (increasing reliability of climate change projections of the hydrological impact model, reducing bias in mean annual discharge) and shortcomings (disrupting feedbacks between fluxes, modification of change signals, assumption of a stationary bias) of its application for hydrological climate change impact studies (e.g., Teutschbein and Seibert, 2012; Maraun, 2016; Ehret et al., 2012; Dettinger et al., 2004; Chen et al., 2021; Huang et al., 2014).

Subsequently, the bias corrected data were statistically downscaled to the hydrological model scale (500 x 500 m^2) using a mass preserving approach. This approach involved the spatial interpolation (inverse distance weighting) of anomalies for each time step from the monthly mean state (1981-2010) for each of the CRCM5-LE cell center points to the hydrological model scale (Brunner et al., 2021b). The interpolation result was then applied to the SDCLIREF reference fields (Brunner et al., 2021b).

For further details, readers are referred to a comprehensive summary in the Supplementary Materials for the CanESM2-LE (S1), the CRCM5-LE (S2), the bias correction (S3), and spatial downscaling method (S4).

2.2.2 Hydrological Model WaSiM

The Water balance simulation Model (WaSiM; Schulla, 2021) was employed to perform the hydrological simulations driven by the CRCM5-LE resulting in a hydro-SMILE (the WaSiM-LE). WaSiM is a distributed, mostly physically-based, and deterministic model for simulations on various spatial (1 m to 10 km) and temporal (minute to daily) scales with a constant time step. It includes routines for evapotranspiration, snow accumulation and melt, glaciers, soil water transfer, groundwater, discharge generation and routing (Schulla, 2021). The model is frequently used for hydrological climate change impact studies for small-scale to mesoscale catchments on various topics, such as glaciers, groundwater, and discharge (Iacob et al., 2017; Neukum and Azzam, 2012; JÓNSDÓTTIR, 2008).

The model was set up in high spatio-temporal resolution (500 m and 3 h) for 98 catchments of the Hydrological Bavaria with a focus on high flow representation using distributed data derived from the European DEM (EU-DEM; European Environment Agency, 2013b), land use data provided by the CORINE land cover dataset (European Environment Agency, 2013a), distributed soil information from the European Soil Database (ESDBv2.0; European Environment Agency, 2013a), as well as groundwater information provided by the Hydrogeologische Übersichtskarte (HÜK; Dörhöfer et al., 2001) and IMHE (IHME; BGR, 2014). A single set of parameters for distributed parameters (i.e., evapotranspiration, soil properties) was defined globally for the entire modeling domain (Willkofer et al., 2020). Local parameters for discharge storage components (i.e., interflow, direct flow) were calibrated using an automated algorithm (dynamically dimensioned search (Tolson and Shoemaker, 2007) and simulated annealing with progressing iterations (Černý, 1985; Kirkpatrick et al., 1983)) minimizing a weighted combination of performance metrics, including Nash and Sutcliff efficiency (NSE; Nash and Sutcliffe, 1970), Kling-Gupta efficiency (KGE; Gupta et al., 2009), the logarithmic NSE and the ratio of root mean squared error to standard deviation (RSR; Moriasi et al. (2007)) (Eq. (1)) (Willkofer et al., 2020). Due to the focus on high flow representation more emphasis was placed on the respective measures (i.e., NSE and KGE). For further details about the model setup the reader is referred to Willkofer et al. (2020). The overall metric (OM) is defined as follows:

$$OM = 0.5 \times (1 - NSE) + 0.25 \times (1 - KGE) + 0.15 \times (1 - logNSE) + 0.1 \times RSR \quad (1)$$

The simulations of a single parameter set for various catchments within a heterogeneous landscape revealed satisfactory to very good results for most of the 98 gauges during the 30-year reference period of 1981 to 2010. However, for a few gauges (NSE: 16; KGE: 5), the model was not able to reproduce the observed discharge satisfactorily (values below 0.5) (Willkofer et al., 2020; Poschlod et al., 2020). Furthermore, the simulations reproduced the mean high flow sufficiently well, with over 60% of the gauges showing absolute deviations from observed values below 20%. Nonetheless, gauges in alpine or pre-alpine catchments exhibited a deficit in mean high flow values due to the lack of observed precipitation resulting from an undercatch of precipitation for that region (Poschlod et al., 2020). Consequently, the level of trust (LOT) for peak flows of return periods of 5, 10, and 20 years flood events, introduced in Willkofer et al. (2020) showed a moderate to high confindence for most catchments, with gauges of poor simulated performance yielding a lower LOT with increasing return levels. LOT were not provided for extreme flood events (i.e., 100-year flood events) since they are subject to significant epistemic uncertainty due to the restricted availability of simulated data (30 years).

The resulting hydro-SMILE comprises 50 members of transient simulated data from 1961 to 2099, providing a total of 6,950 model years to be exploited to analyze extreme values. The entire modelling period is shortened by ten years to account for the time span it takes the RCM to produce fully independent realizations due to the inertia of the ocean model (Leduc et al., 2019).

2.2.3 Benefit of a hydro-SMILE for the estimation of extreme peak flows

This study used the simulated discharge for the reference period of 1981 to 2010 out of the entire dataset to assess the benefits of the hydro-SMILE in estimating return levels. Similar to the individual members of the CRCM5-LE, the members of the WaSiM-LE are equally probable and, therefore, provide a comprehensive database to facilitate the analysis of extreme values.

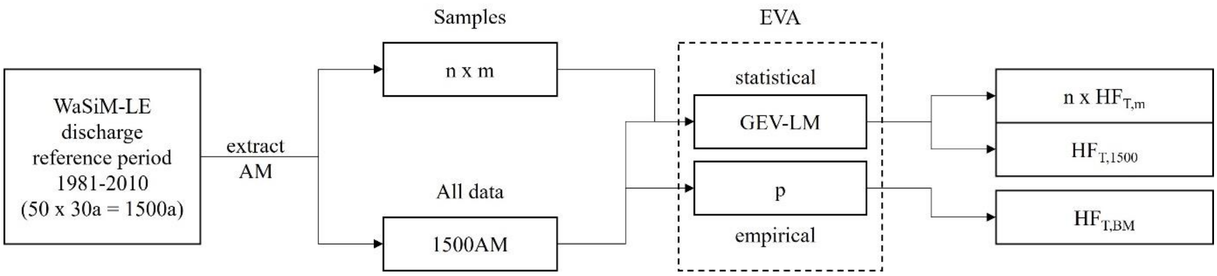


Figure 3: Process chain illustrating the benefit of a hydro-SMILE for climate change impact studies on peak flows of extreme return periods. The process includes extreme value analysis (EVA) based on annual maximum (AM), with bootstrapping resampling to create n different samples of sample size (m). The probability of non-exceedance (p) and the Generalized Extreme Value (GEV) distribution with the L-Moments (LM) estimators are used to derive estimates for high flow values of the return period T (HF_T). The statistical analysis was performed using the extRemes package (v2.0) for R (Gilleland and Katz, 2016).

Figure 3 illustrates the approach taken to emphasize the benefits of the hydro-SMILE in analyzing peak flows of high return periods for the reference period. The 30-year reference period (ref) was selected for all 50 members, resulting in 1,500 model years (50 members x 30 years) of discharge data for each of the 98 gauges. First, the annual maximum of each model year (hydrological year) was extracted for the analysis. Since the database consists of 1,500 model years, this number is considered sufficient to employ empirical non-exceedance probabilities (Martel et al., 2020). However, to demonstrate the benefit of the hydro-SMILE database a statistical analysis using the stationary Generalize Extreme Value (GEV) distribution was also conducted for comparison purposes. A bootstrapping approach with resampling was used to create 1,000 samples (*n*) of different sizes *m* (30, 100, 200) years (each sample without replacement). Using 1,000 samples ensures that one value of the 1,500 AM has a chance of >99% of being selected by chance for $m = 30$. The GEV was employed to estimate the return periods and corresponding confidence intervals. The parameters of the GEV distribution were estimated using L-Moments. The GEV distribution was selected as it is among the better performing methods relying on AM (Bezak et al., 2014) and is the recommended choice for German gauges (Salinas et al., 2014; Fischer and Schumann, 2016).

Although the sample size of 30 and 100 AM may be small for estimating peak flows of high return periods, they were selected along with a size of 200 AM as they represent an average (30 years) to rare (100 & 200 years) data availability of observed discharge values at different gauges (GRDC, 2021). The resulting 1,000 estimates for return levels of peak flows offer a comprehensive database to demonstrate the benefit of the hydro-SMILE. Additionally, the GEV was calculated using the entire 1,500 AM database for each gauge to allow for a comparison with a benchmark value. This benchmark for the return levels of peak discharge was deduced by applying the quantile based on the empirical probability of non-exceedance *p* (Eq. (2)) to all 1,500 AM values for each gauge, and it is considered to represent a robust estimate. This analysis focused on the 100-year flood, which is an event of a 100-year return period T (HF_{100}; T = 100) and the corresponding 99th percentile p of the distribution of the 1,500 AM values as a benchmark.

$$p = 1 - \frac{1}{T} \tag{2}$$

Values for the benchmark derived by the empirical probability as well as the HF_{100} values estimated using the GEV are further normalized to the benchmark to allow for a better comparison.

2.2.4 Projection of changes in frequency and intensity

This study further investigates the dynamics of intensity and frequency of the HF_{100} for three future periods (near future: 2020-2049; mid future: 2040-2069; far future: 2070-2099). Therefore, the robust estimates of extreme return levels of peak flows derived by the empirical probabilities are used for the assessment of climate change impacts on their intensity (C_I, Eq. (3)) and frequency (C_F, Eq. (4a to c)) in the three future periods.

$$C_I = \left(\frac{HF_{T_{fut}} - HF_{T_{ref}}}{HF_{T_{ref}}}\right) \cdot 100\ \% \tag{3}$$

$$C_F = \frac{1}{1 - f\left(HF_{T_{ref}}\right)} \tag{4a}$$

$$f = F\left(HF_{T_{fut}}\right) \tag{4b}$$

$$F(x) = \sum_{i=1}^{j} h_i = \sum_{i=1}^{j} \frac{h(x_i)}{n} \tag{4c}$$

The change in intensity is given as the difference between the future (HF_{Tfut}) and reference value (HF_{Tref}) relative to the reference value in percent. The change in frequency is expressed as the return period value T and is calculated by applying the empirical cumulative distribution function F (ECDF with frequency for an event h_i described as the ratio between the frequency for the specific event $h_{(xi)}$ and the number of all values n, Eq. (4c)) for the respective future period (f, Eq. (4b)) to the value of the 100-year flood of the reference period (Eq. (4a)). The percentile value of f for the reference 100-year flood value is then used to deduce the future return period by solving the empirical probability of non-exceedance for the return period T (Eq. (4a)). The change signals are calculated for each of the above mentioned 30-year future periods. However, this analysis requires stationarity for the underlying data. Since we use the entire 1,500 model years provided by the 50 members, we determine stationarity if less than 5 % of the members exhibit a significant trend for each individual gauge. A Mann-Kendall (MK) test for stationarity conducted on each individual member and gauge revealed no significant trend for the reference period (with significance level $\alpha = 0.01$) for more than 95 % of the members along all gauges. However, for the future periods the MK test exhibits significant trends for more than 5 % of the members in 6 of the 98 gauges. Limiting the evaluation periods to 20 years instead of 30 years lead to similar results for the MK test showing no apparent trend for all gauges in the reference period, but showing for at least one gauge a significant trend (more than 5 % of members with a trend) in the future periods. Studies by Poschlod et al. (2020) and Brunner et al. (2021b) conducted their analysis on the same database using time slices of at least 30 years as well. Thus, we choose to use 30-year periods since stationarity criteria are met in most catchments and opt for the larger database, as well as maintaining consistency with these studies.

3 Results

3.1 Benefits of hydro-SMILEs for the estimation of extreme return periods of peak flows

Large ensembles provide a vast amount of data, therefore they are considered to be beneficial for extreme value analysis (Kendon et al., 2008; Kjellström et al., 2013; Wood and Ludwig, 2020). The benefit of a hydro-SMILE to determine robust extreme hydrological discharge values for Hydrological Bavaria are analyzed, specifically for the 100-year flood. The robust values for the discharge gauges, derived using the empirical probability of non-exceedance for a 100-year event, serve as a benchmark for comparison with values derived by the GEV distribution using three different sample sizes (30, 100, 200) of AM values (Figure 4).

The results shown in Figure 4 (a, b, and c) illustrate that the estimates of HF_{100} are more robust with an increasing number of AM values used for the GEV, as indicated by the spread of the blue markers around the black benchmark line. Table 1 summarizes the statistical characteristics of the deviation of the estimates from the benchmark across all 98 gauges. While the range of the relative deviation of the 1,000 samples of HF_{100} estimates from the benchmark is between 0.33 and 2.71 when calculated with a sample size of 30 AM values (panel a), this range diminishes to 0.49 and 1.91 for 100 AM values (panel c) and 0.56 and 1.60 for 200 AM values (panel e). Therefore, the range of the 1,000 estimates diminishes with an increase in sample size and the values cluster more densely around the benchmark. However, despite the remaining non-negligible range of deviations from the benchmark, the mean (1.01) as well as the median (0.98 to 1.0) across all values for all gauges are close to the benchmark value for different sample sizes. The inner 50 % of the 1000 samples across all 98 gauges exhibit the largest deviation with a sample size of 30 AM (between 0.84 and 1.15) and the lowest for 200 AM (0.94 to 1.07). Therefore, only 25 percent of the samples show underestimations below 0.84 (0.92, 0.94) and only 75 percent exhibit larger overestimations than 1.15 (1.08, 1.07) with a sample size of 30 AM (100 AM, 200 AM). Thus, with deviations larger than 15 % for 50 percent of the estimates calculated using a sample size of 30 AM, only half of the estimated HF_{100} values are within an acceptable range (±15 %, considering model parameter uncertainty and errors in observations affecting the model quality regarding high flows) compared to the benchmark. This number increases with a larger sample size.

Table 1: Summary of overall statistics of the relative deviation of the HF_{100} estimates from the benchmark value across all gauges. The table incldues the number of sample (n), sample size (m) given in annual maximum (AM) values and the 0.25/0.75 quantile (Q25, Q75) values.

n	m	minimum	Q25	mean	median	Q75	maximum
1000	30	0.33	0.84	1.01	0.98	1.15	2.71
1000	100	0.49	0.92	1.01	1.00	1.08	1.91
1000	200	0.56	0.94	1.01	1.00	1.07	1.60
1	1500	0.98	1.00	1.02	1.01	1.03	1.09

While the majority of gauges show estimates that are evenly distributed around the benchmark, some gauges exhibit a tendency towards over- or underestimation of the HF_{100} estimates with more values falling above or below the benchmark line. This behavior may be different when using more than 1000 samples to conduct the analysis. The difference between the benchmark value obtained from empirical probability and the estimates obtained from the GEV distribution can vary greatly depending on the samples selected from 1,500 AM values.

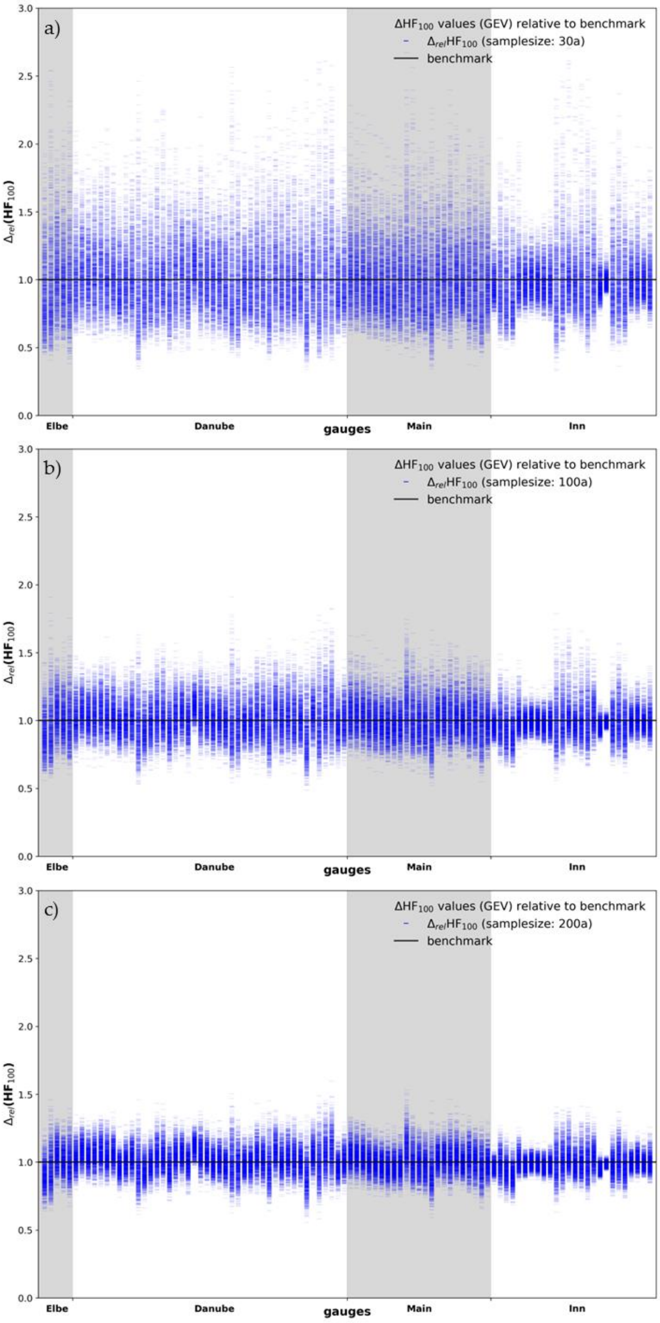


Figure 4: Comparison of HF_{100} estimates calculated using the GEV distribution with 1000 AM samples of a) 30, b) 100, and c) 200 years per gauge (blue markers) with the respective benchmark value (solid black line) for 98 gauges.

In Figure 5, a comparison is made between the HF_{100} estimates derived using the empirical probability of non-exceedance and those obtained using the GEV distribution (and associated 95 % confidence intervals) for the entire ensemble of 1,500 AM values. The robust values obtained from the empirical probability are used as the benchmark for this comparison.

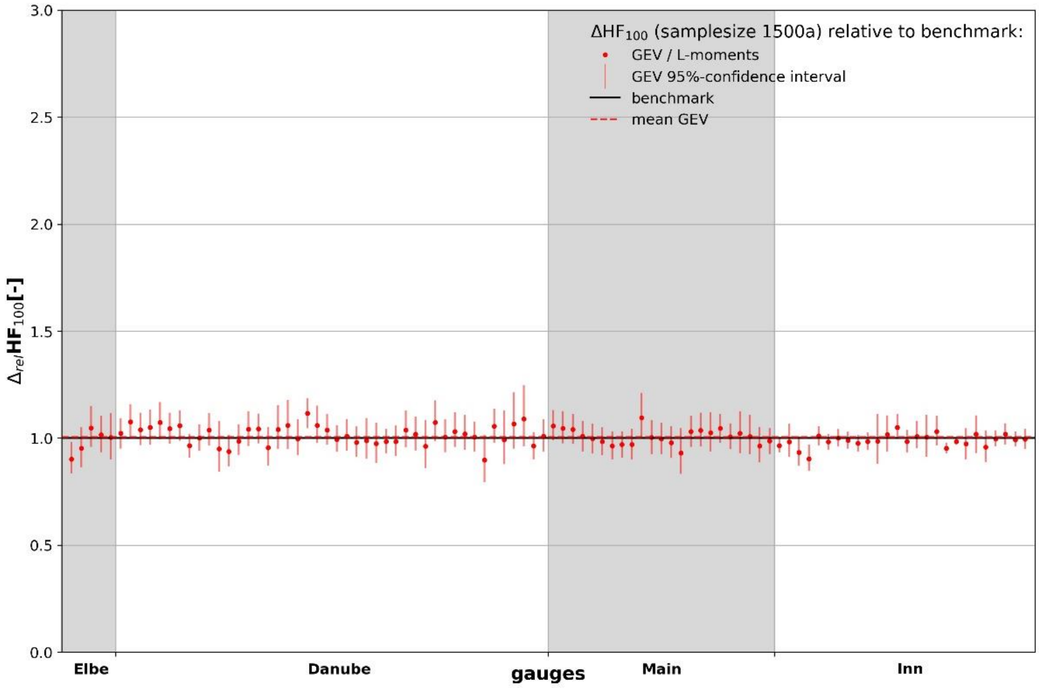


Figure 5: Relative difference of the HF_{100} values (red dots) and respective 95^{th} confidence intervals (vertical red lines) calculated with the GEV distribution from the benchmark HF_{100} value (black solid line) derived from the probability of non-exceedance p = 0.01. The results are based on the entire reference database of 1500 AM values for each of the 98 gauges. The horizontal red dashed line illustrates the mean of relative difference the HF_{100} calculated with the GEV distribution for the 98 gauges.

The estimates obtained using the GEV show differences from the benchmark, with varying magnitudes across the gauges. However, the mean difference across all HF_{100} estimates is small and marginally different from the benchmark. For most gauges, the individual differences from the benchmark are also small, with only two gauges showing values exceeding ±10%. However, for 7 of 98 gauges the 95 % confidence interval does not overlap with the benchmark, indicating that in this case the empirical approach yields a more robust value compared to the GEV estimates even with the enhanced robustness gained by employing 1,500 AM values as a sample. However, if the unknown population could be represented by the GEV, the distribution would yield a better fit. Therefore, the determination of peak flows with extreme return periods employing the empirical probability on the vast data base of a hydro-SMILE allows for a more precise estimation. Hence, also the estimation of future return periods is more robust allowing for a better quantification of the changing dynamics in the frequency and intensity of high return periods in future projections due to changes in the climate.

3.2 Changing dynamics of the 100-year peak flows in future projections

The changes in 100-year peak flows (HF_{100}) for the investigated gauges in Hydrological Bavaria in the 21st century are summarized for five distinct discharge regimes (defined by the Pardé coefficient) which were adapted from Poschlod et al. (2020) (Figure 1). One gauge that was originally assigned to its own regime has been re-allocated to the pluvial (unbalanced) regime, as it exhibits a similar mean discharge behavior. The regimes comprise the glacio-nival regime of four high Alpine catchments, a nival regime of mostly Alpine to pre-Alpine catchments, a nivo-pluvial regime of pre-Alpine catchments, a balanced pluvial regime along the Danube and its tributaries in the Alpine foreland, and the unbalanced pluvial regime.

Within the study area, the flood protection structures are typically desinged based on a stipulated estimation of HF_{100} from observations, which represent a stationary condition in the past. Any future increase in the intensity and frequency of these extreme values poses a threat to these structures. Therefore, the following graphs highlight the changes of the HF_{100} events for the three future periods.

Figure 6 displays violin plots that illustrate the range of changes in the intensity of HF_{100} events for the different discharge regimes as well as the distribution of changes across the respective clusters of gauges for the near (horizon 2035), mid (horizon 2055), and far future (horizon 2085) periods. Overall, 78 % of all gauges (76/98) show an increase in intensity for the 2035 horizon, 76 % (74/98) for the 2055 horizon, and 89 % (87/98) for the 2085 horizon.

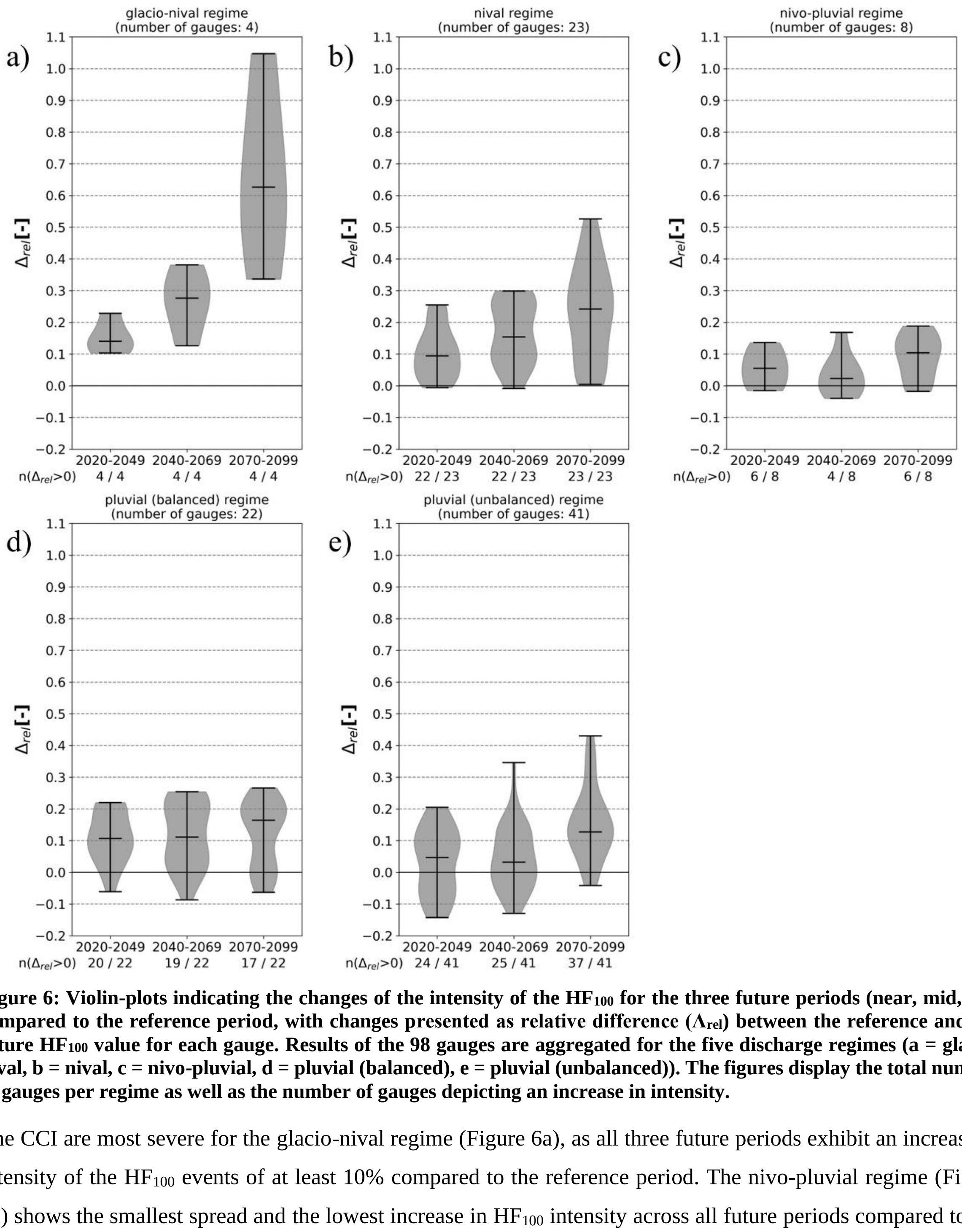


Figure 6: Violin-plots indicating the changes of the intensity of the HF_{100} for the three future periods (near, mid, far) compared to the reference period, with changes presented as relative difference (Λ_{rel}) between the reference and the future HF_{100} value for each gauge. Results of the 98 gauges are aggregated for the five discharge regimes (a = glacio-nival, b = nival, c = nivo-pluvial, d = pluvial (balanced), e = pluvial (unbalanced)). The figures display the total number of gauges per regime as well as the number of gauges depicting an increase in intensity.

The CCI are most severe for the glacio-nival regime (Figure 6a), as all three future periods exhibit an increase in intensity of the HF_{100} events of at least 10% compared to the reference period. The nivo-pluvial regime (Figure 6c) shows the smallest spread and the lowest increase in HF_{100} intensity across all future periods compared to the reference period. As the distance from the Alps increases and the discharge regimes shift from snowmelt influenced to more precipitation driven, the number of gauges projecting a decrease in HF_{100} intensities increases. However, the majority of gauges still exhibit an increase in intensities, with up to 18.8% for the nivo-pluvial Figure 6c), 26.6% for the balanced pluvial (Figure 6d), and 43 % for the unbalanced pluvial regime (Figure 6e) in the far future. The gradient of an increase in intensity over all three projection periods is small for the nivo-pluvial and balanced pluvial regimes, which show the least intensification of HF_{100} values for the respective periods. However, the gradient of increase is more distinct for the remaining regimes, with the largest increase in the glacio-nival regime (Figure 6a). The gauges in this regime depict the strongest increase in HF_{100} intensities for the 2085 horizon, with an increase of 36.6 % to 104.7 %.

Based on the future projections of the hydro-SMILE, the discharge values of the HF_{100} are likely to increase for most of the gauges of Hydrological Bavaria. Consequently, the frequency of the HF_{100} discharge for the reference period also increases. Figure 7 shows the change in frequency between the future and the reference period for the different regimes.

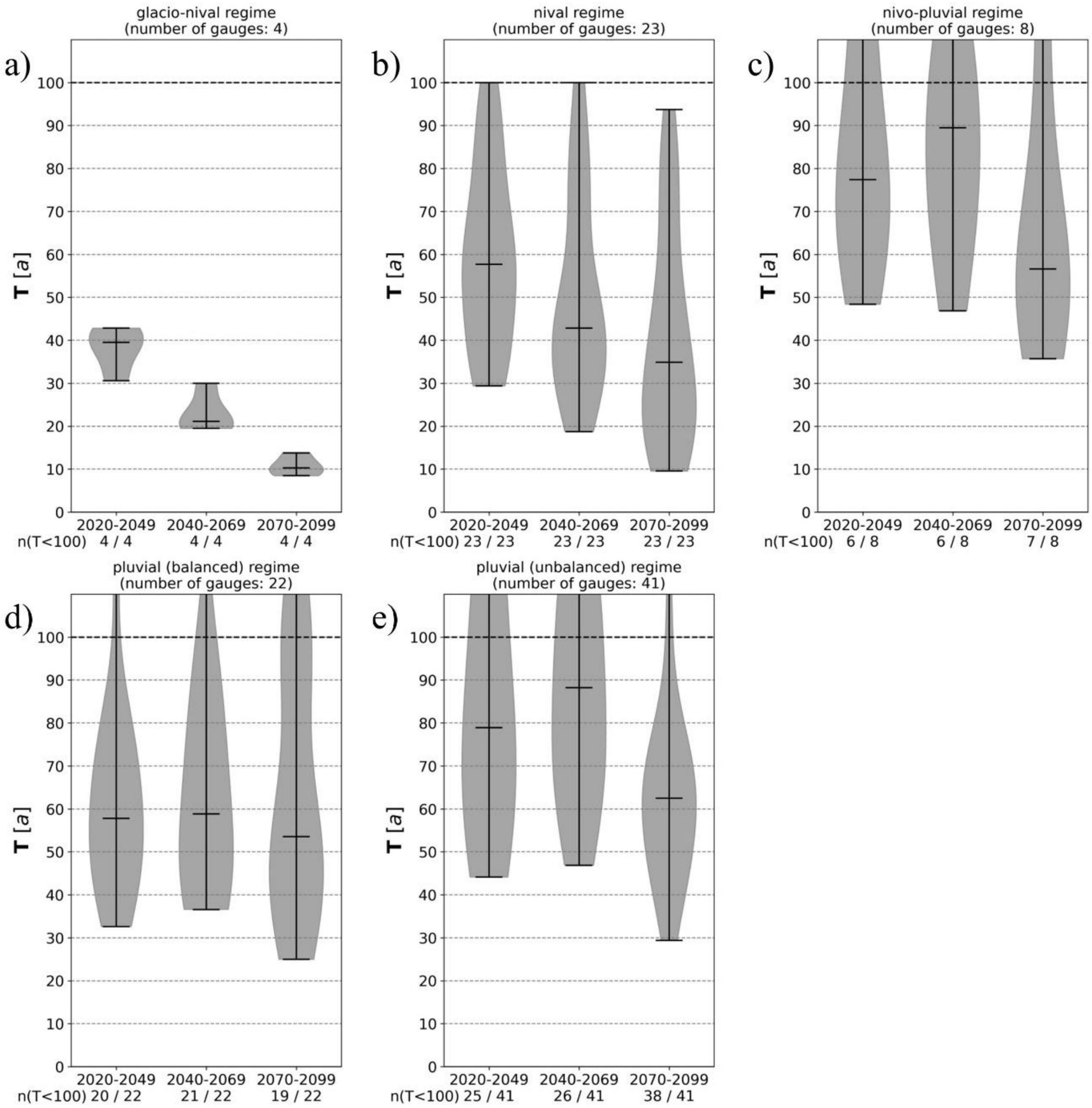


Figure 7: Violin-plots indicating the changes of the frequency of the HF_{100} for the three future periods (near, mid, far) compared to the reference period, with changes presented as absolute values of return periods (T[a]) of the respective future period compared to the 100-year return period for each gauge. Results of the 98 gauges are aggregated for the five discharge regime (a = nivo-glaical, b = nival, c = nivo-pluvial, d = pluvial (balanced), e = pluvial (unbalanced)). The figures display the total number of gauges per regime as well as the number of gauges depicting an increased frequency.

Values indicate the new return period associated with the HF_{100} discharge from the reference period. This means values below 100 indicate an increase in frequency. The glacio-nival regime (Figure 7a) also exhibits the strongest increase in frequency among all regimes with the HF_{100} of the past becoming equivalent to a 31- to 43-year event in the near future, thus becoming roughly two to three times more frequent. For the 2085 horizon the same HF_{100} event becomes an 8- to 14-year event showing a seven to twelve-fold increase in frequency. A similar development is visible for some gauges in the nival regime (Figure 7b). While the violin plot for this regime indicates that the reference 100-year event will become a 70-year event for more than 50% of gauges, some gauges show no or only a minor increase in frequency as well. The changes for the remaining regimes are less severe, but still indicate an increase in frequency for up to 50 % of the respective gauges until the middle of the century and more than 50 % in the far future. The changes for the nivo-pluvial regime (Figure 7c) and the unbalanced pluvial (Figure 7d) regime show that the frequency declines for less than 50 % of the gauges in the near and mid future period. Therefore, the 100-year event becomes more frequent for more than 50 % of the gauges with varying extent. While the magnitude of changes is similarly moderate (except for the far future) for Figure 7c and Figure 7e, projected future return periods for the HF_{100} event for Figure 7d depict stronger change signals towards higher frequencies with more than 50 % of gauges showing values smaller than 60 years. Furthermore, the nivo-pluvial as well as the balanced and unbalanced pluvial regimes exhibit a slight decrease in frequency in the mid future compared to the remaining projection periods while the intensity does not show this behavior. However, this circumstance may be explained by the change in driving agent from snowmelt driven events in the near future to rainfall induced events at the end of the century. Thus, at the 2055 horizon the shift of the ratio of both event types contributes to this slight decline in frequency.

Some gauges within the nivo-pluvial and both pluvial regimes depict a decrease in frequency and/or intensity. These gauges usually exhibit natural or artificial influences, such as the retention effect of natural lakes, reservoirs, or diversions or gauges of small catchments which might experience less dynamics in changes of flood drivers or even a reduction.

Overall, the changes in frequency and intensity due to the projected changes in climate according to the CRCM5-LE become less severe with increasing distance from the Alps. Furthermore, the increase in frequency and intensity for alpine catchments is seemingly high, but in line with the results of Hattermann et al. (2018), which showed comparable results for the near future period (100-year event frequency between 20 and 40 years). The influencing factors for these in parts severe changes are manifold. However, Brunner et al. (2021b) analyzed the relation between the extremeness of precipitation and discharge for 78 out of the 98 gauges within Hydrological Bavaria and concluded that an increase in extreme precipitation intensity is of higher importance for extreme return levels of discharge than land surface processes, such as antecedent soil moisture or changes in snowpack due to warmer temperatures. If precipitation volumes are sufficiently large, they quickly saturate the soil or yield an excessive amount of direct runoff due to infiltration excess (Brunner et al., 2021b).

The mean magnitude of the annual maximum precipitation is projected to change for different temporal aggregation levels (3-hourly to 5-daily) in the CRCM5-LE (Wood and Ludwig, 2020), as well as the magnitude of 100-year return period rainfall increases by 10-20% and the frequency increases by 2 to 4 times (Martel et al., 2020) for Hydrological Bavaria. The changes are associated with seasonal shifts from summer to winter events and are particularly pronounced in the Alpine region (Martel et al., 2020; Wood and Ludwig, 2020). Severe floods that occur simultaneously in different catchments of the study area are usually associated with a cutoff low Vb cyclone that results in prolonged precipitation events lasting up to 15 days over the same region (Stahl and Hofstätter, 2018; Mittermeier et al., 2019). Under changing climate conditions projected by the CRCM5-LE by the end of the 21st century employing the RCP8.5 scenario, these events are likely to intensify in volume and frequency during winter and spring and occur less frequently during the summer months but with an increased precipitation volume (Mittermeier et al., 2019).

4 Discussion

The variability of statistical characteristics within a time series can affect the estimation of extreme values due to extraordinary events (Fischer and Schumann, 2016). The results of this study emphasize the benefit of using data provided by a climatological SMILE for hydrological impact studies as it provides a profound basis for extreme value statistics and allows for more accurate estimation of extreme values, as also shown by other studies (Champagne et al., 2020; Ehmele et al., 2020; Maher et al., 2021). However, the in parts large deviations between the benchmark (robust estimate derived from the empirical probability for a 100-year flood event using 1,500 AM values) and the estimates derived using a GEV based on different sample sizes (30, 100, 200) might be reduced when using an EVD which is better suited for the respective sample when enough data is available (as is the case with hydro-SMILE used here). In some cases the GEV might not the best distribution for the samples of the respective gauge which might affect the differences from the benchmark since higher quantiles heavily depend on the distribution (Schulz and Bernhardt, 2016). However, the approach presented in this study illustrates the benefit of a hydro-SMILE as it provides a more robust estimate by employing empirical probabilities for the deduction of extreme values. Therefore, these robust estimates allow for a more robust assessment of future dynamics of extreme high flows.

The results of this study are subject to uncertainties (parameter, process description) as they are produced by data created at the end of a cascade of modeling steps usually applied for climate change impact studies as displayed in Figure 2. Different components (e.g., climate model, hydrological model, bias correction) affect different discharge characteristics or indicators (e.g., extreme indicators, mean discharge) (Gampe et al., 2019; Muerth et al., 2012; Muerth et al., 2013; Velázquez et al., 2013; Willkofer et al., 2018). A thorough assessment of the contribution of the chain compartments to the overall uncertainty would require anensemble of multiple climate and hydrological models.

The overall strong increase in frequency and intensity of the HF_{100} in the future may be driven by deficiencies of the employed hydrological model, such as generalized glacier model among affected catchments, or a single snow melt approach for the entire Hydrological Bavaria (as described in Willkofer et al., 2020). However, as stated in the previous section, this scale of change was also found by Hattermann et al. (2018) for the upper Danube basin using the same emission scenario projections, but a different hydrological and climate model, which might indicate that the change signals are likely independent of the chosen hydrological or climate model.

The results of the CCI on the frequency and intensity also depend on the performance of the hydrological model. Since it relies on observations for parameter calibration, the quality of this data is crucial, especially for extreme values. For the most extreme events (e.g., HF_{100} and above) the river may inundate the surrounding area and the water level / discharge relationship at the gauging station used to determine discharge values may not be valid anymore and is likely to underestimate the peak discharge. Therefore, the actual observed discharge – and thus, the calibrated model – is prone to these measurement uncertainties. Furthermore, the discharge of rivers within Hydrological Bavaria is heavily impacted by management structures for flood protection or hydro power generation, especially the southern tributaries of the Danube in the Alpine foreland and within the Alps are heavily regulated. Since the management follows somewhat fuzzy rules and actual data is restricted by private companies in most cases, the management rules for these structures have to be deduced from publicly available data and implemented in the hydrological model. These rules are susceptible to extreme conditions as they do not allow for adaptations during model runtime (e.g., flushing a reservoir prior to an anticipated heavy precipitation event).

The projected future changes in extreme discharges may be attributed in part, to the climatological reference dataset, as it affects the performance of the hydrological model as well as the CCS through bias adjustment (Gampe et al., 2019; Meyer et al., 2019; Willkofer et al., 2018). Precipitation in high altitudes (e.g., the Alps) may be under-captured (Westra et al., 2014; Poschlod, 2021; Prein and Gobiet, 2017; Rauthe et al., 2013; Poschlod et al., 2020; Willkofer et al., 2020) resulting in an underestimation of observed precipitation in these regions, especially of extreme values. Assuming a temporally stationary bias, changes in the extremes might be overestimated due to an over-adjustment of the distribution of the reference period towards underestimated observations compared to the future periods. Furthermore, the variables are adjusted individually and thus, physical coherency as for a multivariate approach proposed by Meyer et al. (2019) is not guaranteed. This specifically affects discharges governed by snow or glacier melt of higher elevation within the Alps (Meyer et al., 2019).

Since the presented modelling approach only comprises one GCM-RCM combination forced by the more extreme RCP8.5 emission scenario as well as one hydrological model, the significance of the findings regarding the variance of change effects in the future on the development of extreme peak flows is limited. Furthermore, the projected climate change signals of the CRCM5-LE were found to depict a stronger warming and drying compared to other large ensembles (von Trentini et al., 2020) which might result in these part extreme increase in frequency and intensity of the HF_{100} values among many gauges of Hydrological Bavaria.

Projected discharge extremes at the upper end of the distribution that have not been observed to date might be created by unrealistic compound events due to flaws in the bias correction approach (Kelder et al., 2022). Thus, these events directly influence the EVD, producing higher return values, and consequently, a larger change signal. However, as extreme precipitation events of various durations are expected to intensify within the studied region, the probability for yet unseen floods due to compounding events may also increase in the future.

5 Conclusion

This study emphasizes the benefit of employing a climatological SMILE with a hydrological model to create a hydro-SMILE to foster extreme value statistics and analyze the impacts of climate change on hydrological extreme values such as the HF_{100} due to the provision of a very large database. This database allows for the application of empirical exceedance probabilities to estimate robust discharge values of high return periods rather than statistical extrapolation based on extreme value distributions. The results show that the performance of statistical estimates largely depends on the available length of the time series as well as its values when compared to the empirical benchmark. However, even with a length of 200 AM, the variance of the scatter of HF_{100} estimates of the 1,000 samples was rather large.

As mentioned by Willkofer et al. (2020) the performance of the hydrological model allows for CCI studies - in this case using the CRCM5-LE to elaborate on the effects of climate change on the development of the HF_{100}. The projections reveal a strong increase in the intensity and frequency of HF_{100} events for Alpine and pre-Alpine catchments exhibiting a snowmelt driven discharge regime within the reference period. This strong increase in the intensity and frequency is considerably smaller for catchments north of the Alps and of a more pluvial discharge regime. The in parts tremendous changes of HF_{100} intensities and frequencies may be ascribed to the emission scenario (RCP8.5). Thus, the addition of different SMILEs and hydrological models may foster the significance of the findings due to different climate projections and simulated climatological and hydrological processes along the model chain. However, the establishment of such extensive model chains requires vast computational resources. Nevertheless, this effort should be considered regarding the benefits this profound database offers for

extreme value statistics, fostering the knowledge about the propagation of natural variability of the climate system to the hydrological response (Brunner et al., 2021b), or allowing to distinguish climate change signals (or forced response) from natural variability for extreme values (Wood and Ludwig, 2020; Aalbers et al., 2018).
Furthermore, the results highlight the need to incorporate climate projections in the design of new flood protection infrastructure or adapting existing structures to reduce future flood risk, not only in Hydrological Bavaria, but everywhere in general.

Contributions

FW developed the concept of this study including methods, investigation, and visualization, performed the formal analysis and validation, and prepared the original draft. FW and RRW were responsible for data curation and software development. RL acquired the funding for the presented research, provided the required resources, was responsible for the project administration, and supervised the presented research.

Competing interests:

The authors declare that they have no conflict of interest.

Acknowledgements:

The authors acknowledge the colleagues Gilbert Brietzke, André Kurzmann, and Jens Weismüller from Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities for their technical support during the development of the hydrological large ensemble. We also thank the Leibniz Supercomputing Centre for providing the HPC infrastructure and computation time.
F.W. was supported by the ClimEx project, funded by the Bavarian Ministry for the Environment and Consumer Protection.
The CRCM5-LE was created within the ClimEx Project, which was funded by the Bavarian Ministry for the Environment and Consumer Protection. Computations of the CRCM5-LE were performed on the SuperMUC HPC system of the Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities. We acknowledge Environment and Climate Change Canada for providing the CanESM2-LE driving data.

References

Aalbers, E. E., Lenderink, G., van Meijgaard, E., and van den Hurk, B. J. J. M.: Local-scale changes in mean and heavy precipitation in Western Europe, climate change or internal variability?, Clim Dyn, 50, 4745–4766, doi:10.1007/s00382-017-3901-9, 2018.

Arora, V. K., Scinocca, J. F., Boer, G. J., Christian, J. R., Denman, K. L., Flato, G. M., Kharin, V. V., Lee, W. G., and Merryfield, W. J.: Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases, Geophys. Res. Lett., 38, L05805, doi:10.1029/2010GL046270, 2011.

Bertola, M., Viglione, A., Lun, D., Hall, J., and Blöschl, G.: Flood trends in Europe: are changes in small and big floods different?, Hydrol. Earth Syst. Sci., 24, 1805–1822, doi:10.5194/hess-24-1805-2020, 2020.

Bezak, N., Brilly, M., and Šraj, M.: Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis, Hydrological Sciences Journal, 59, 959–977, doi:10.1080/02626667.2013.831174, 2014.

BGR: International Hydrogeological Map of Europe 1:1,500,000 (IHME1500 v1.1), Bundesanstalt für Geowissenschaften und Rohstoffe (BGR), Hannover, Germany, Paris, France, 2014.

Blöschl, G., Gaál, L., Hall, J., Kiss, A., Komma, J., Nester, T., Parajka, J., Perdigão, R. A. P., Plavcová, L., Rogger, M., Salinas, J. L., and Viglione, A.: Increasing river floods: fiction or reality?, WIREs Water, 2, 329–344, doi:10.1002/wat2.1079, 2015.

Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A. P., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G. T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Frolova, N., Ganora, D., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Volpi, E., Wilson, D., Zaimi, K., and Živković, N.: Changing climate both increases and decreases European river floods, Nature, 573, 108–111, doi:10.1038/s41586-019-1495-6, 2019.

Blöschl, G., Nester, T., Komma, J., Parajka, J., and Perdigão, R. A. P.: The June 2013 flood in the Upper Danube Basin, and comparisons with the 2002, 1954 and 1899 floods, Hydrol. Earth Syst. Sci., 17, 5197–5212, doi:10.5194/hess-17-5197-2013, 2013.

Brunner, M. I., Slater, L., Tallaksen, L. M., and Clark, M.: Challenges in modeling and predicting floods and droughts: A review, WIREs Water, 8, doi:10.1002/wat2.1520, 2021a.

Brunner, M. I., Swain, D. L., Wood, R. R., Willkofer, F., Done, J. M., Gilleland, E., and Ludwig, R.: An extremeness threshold determines the regional response of floods to changes in rainfall extremes, Commun Earth Environ, 2, doi:10.1038/s43247-021-00248-x, 2021b.

Černý, V.: Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm, J Optim Theory Appl, 45, 41–51, doi:10.1007/BF00940812, 1985.

Champagne, O., Leduc, M., Coulibaly, P., and Arain, M. A.: Winter hydrometeorological extreme events modulated by large-scale atmospheric circulation in southern Ontario, Earth Syst. Dynam., 11, 301–318, doi:10.5194/esd-11-301-2020, 2020.

Chen, J., Arsenault, R., Brissette, F. P., and Zhang, S.: Climate Change Impact Studies: Should We Bias Correct Climate Model Outputs or Post-Process Impact Model Outputs?, Water Resour. Res., 57, doi:10.1029/2020WR028638, 2021.

Dettinger, M. D., Cayan, D. R., Meyer, M. K., and Jeton, A. E.: Simulated Hydrologic Responses to Climate Variations and Change in the Merced, Carson, and American River Basins, Sierra Nevada, California, 1900–2099, Climatic Change, 62, 283–317, doi:10.1023/B:CLIM.0000013683.13346.4f, 2004.

Dörhöfer, G., Hannappel, S., Reutter, E., and Voigt, H.-J.: Die Hydrogeologische Übersichtskarte von Deutschland HÜK200, Z. Für Angew. Geol., 153–159, 2001.

Ehmele, F., Kautz, L.-A., Feldmann, H., and Pinto, J. G.: Long-term variance of heavy precipitation across central Europe using a large ensemble of regional climate model simulations, Earth Syst. Dynam., 11, 469–490, doi:10.5194/esd-11-469-2020, 2020.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions "Should we apply bias correction to global and regional climate model data?", Hydrol. Earth Syst. Sci., 16, 3391–3404, doi:10.5194/hess-16-3391-2012, 2012.

European Environment Agency: Corine Land Cover 2006 v17, https://www.eea.europa.eu/data-and-maps/data/clc-2006-raster-3, 2013a.

European Environment Agency: Digital Elevation Model over Europe (EU-DEM), https://www.eea.europa.eu/data-and-maps/data/eu-dem, 2013b.

Fischer, S. and Schumann, A.: Robust flood statistics: comparison of peak over threshold approaches based on monthly maxima and TL-moments, Hydrological Sciences Journal, 61, 457–470, doi:10.1080/02626667.2015.1054391, 2016.

Fyfe, J. C., Derksen, C., Mudryk, L., Flato, G. M., Santer, B. D., Swart, N. C., Molotch, N. P., Zhang, X., Wan, H., Arora, V. K., Scinocca, J., and Jiao, Y.: Large near-term projected snowpack loss over the western United States, Nature communications, 8, 14996, doi:10.1038/ncomms14996, 2017.

Gampe, D., Schmid, J., and Ludwig, R.: Impact of Reference Dataset Selection on RCM Evaluation, Bias Correction, and Resulting Climate Change Signals of Precipitation, Journal of Hydrometeorology, 20, 1813–1828, doi:10.1175/JHM-D-18-0108.1, 2019.

Gilleland, E. and Katz, R. W.: extRemes 2.0: An Extreme Value Analysis Package in R, J. Stat. Soft., 72, doi:10.18637/jss.v072.i08, 2016.

GRDC: Summary Statistics by Country, https://www.bafg.de/SharedDocs/ExterneLinks/GRDC/summary_stat_cc_pdf.pdf?__blob=publicationFile, 2021.

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

Hattermann, F. F., Wortmann, M., Liersch, S., Toumi, R., Sparks, N., Genillard, C., Schröter, K., Steinhausen, M., Gyalai-Korpos, M., Máté, K., Hayes, B., del Rocío Rivas López, M., Rácz, T., Nielsen, M. R., Kaspersen, P. S., and Drews, M.: Simulation of flood hazard and risk in the Danube basin with the Future Danube Model, Climate Services, 12, 14–26, doi:10.1016/j.cliser.2018.07.001, 2018.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, J. Geophys. Res., 113, doi:10.1029/2008JD010201, 2008.

Huang, S., Krysanova, V., and Hattermann, F. F.: Does bias correction increase reliability of flood projections under climate change? A case study of large rivers in Germany, Int. J. Climatol., 34, 3780–3800, doi:10.1002/joc.3945, 2014.

Iacob, O., Brown, I., and Rowan, J.: Natural flood management, land use and climate change trade-offs: the case of Tarland catchment, Scotland, Hydrological Sciences Journal, 62, 1931–1948, doi:10.1080/02626667.2017.1366657, 2017.

IPCC: Summary for Policymakers, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B. (Eds.), In Press., 2021.

JÓNSDÓTTIR, J. F.: A runoff map based on numerically simulated precipitation and a projection of future runoff in Iceland / Une carte d'écoulement basée sur la précipitation numériquement simulée et un scénario du futur écoulement en Islande, Hydrological Sciences Journal, 53, 100–111, doi:10.1623/hysj.53.1.100, 2008.

Kelder, T., Wanders, N., van der Wiel, K., Marjoribanks, T. I., Slater, L. J., Wilby, R. l., and Prudhomme, C.: Interpreting extreme climate impacts from large ensemble simulations—are they unseen or unrealistic?, Environ. Res. Lett., 17, 44052, doi:10.1088/1748-9326/ac5cf4, 2022.

Kendon, E. J., Rowell, D. P., Jones, R. G., and Buonomo, E.: Robustness of Future Changes in Local Precipitation Extremes, J. Climate, 21, 4280–4297, doi:10.1175/2008JCLI2082.1, 2008.

Kirchmeier-Young, M. C., Zwiers, F. W., and Gillett, N. P.: Attribution of Extreme Events in Arctic Sea Ice Extent, J. Climate, 30, 553–571, doi:10.1175/JCLI-D-16-0412.1, 2017.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P.: Optimization by simulated annealing, Science (New York, N.Y.), 220, 671–680, doi:10.1126/science.220.4598.671, 1983.

Kjellström, E., Thejll, P., Rummukainen, M., Christensen, J. H., Boberg, F., Christensen, O. B., and Fox Maule, C.: Emerging regional climate change signals for Europe under varying large-scale circulation conditions, Clim. Res., 56, 103–119, doi:10.3354/cr01146, 2013.

Leduc, M., Mailhot, A., Frigon, A., Martel, J.-L., Ludwig, R., Brietzke, G. B., Giguère, M., Brissette, F., Turcotte, R., Braun, M., and Scinocca, J.: The ClimEx Project: A 50-Member Ensemble of Climate Change Projections at 12-km Resolution over Europe and Northeastern North America with the Canadian Regional Climate Model (CRCM5), Journal of Applied Meteorology and Climatology, 58, 663–693, doi:10.1175/JAMC-D-18-0021.1, 2019.

Ludwig, R., Wood, Raul, R., Willkofer, F., von Trentini, F., Mittermeier, M., Böhnisch, A., and Poschlod, B.: ClimEx - Klimawandel und Extremereignisse: Risiken und Perspektiven für die bayerische Wasserwirtschaft, Abschlussbericht, Ludwig-Maximilians-Universität, 190 pp., 2019.

Maher, N., Milinski, S., and Ludwig, R.: Large ensemble climate model simulations: introduction, overview, and future prospects for utilising multiple types of large ensemble, Earth Syst. Dynam., 12, 401–418, doi:10.5194/esd-12-401-2021, 2021.

Maraun, D.: Bias Correcting Climate Change Simulations - a Critical Review, Curr Clim Change Rep, 2, 211–220, doi:10.1007/s40641-016-0050-x, 2016.

Martel, J.-L., Mailhot, A., and Brissette, F.: Global and Regional Projected Changes in 100-yr Subdaily, Daily, and Multiday Precipitation Extremes Estimated from Three Large Ensembles of Climate Simulations, J. Climate, 33, 1089–1103, doi:10.1175/JCLI-D-18-0764.1, 2020.

Martel, J.-L., Mailhot, A., Brissette, F., and Caya, D.: Role of Natural Climate Variability in the Detection of Anthropogenic Climate Change Signal for Mean and Extreme Precipitation at Local and Regional Scales, J. Climate, 31, 4241–4263, doi:10.1175/JCLI-D-17-0282.1, 2018.

Martynov, A., Laprise, R., Sushama, L., Winger, K., Šeparović, L., and Dugas, B.: Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian Regional Climate Model, version 5: model performance evaluation, Clim Dyn, 41, 2973–3005, doi:10.1007/s00382-013-1778-9, 2013.

Meyer, J., Kohn, I., Stahl, K., Hakala, K., Seibert, J., and Cannon, A. J.: Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments, Hydrol. Earth Syst. Sci., 23, 1339–1354, doi:10.5194/hess-23-1339-2019, 2019.

Mittermeier, M., Braun, M., Hofstätter, M., Wang, Y., and Ludwig, R.: Detecting Climate Change Effects on Vb Cyclones in a 50-Member Single-Model Ensemble Using Machine Learning, Geophys. Res. Lett., 46, 14653–14661, doi:10.1029/2019GL084969, 2019.

Moriasi, D. N., Arnold, J. G., van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations, Transactions of the ASABE, 50, 885–900, doi:10.13031/2013.23153, 2007.

Mpelasoka, F. S. and Chiew, F. H. S.: Influence of Rainfall Scenario Construction Methods on Runoff Projections, Journal of Hydrometeorology, 10, 1168–1183, doi:10.1175/2009JHM1045.1, 2009.

Muerth, M., Gauvin St-Denis, B., Ludwig, R., and Caya, D.: Evaluation of different sources of uncertainty in climate change impact research using a hydro-climatic model ensemble, International Congress on Environmental Modelling and Software, doi:10.5282/ubm/epub.14094, 2012.

Muerth, M. J., Gauvin St-Denis, B., Ricard, S., Velázquez, J. A., Schmid, J., Minville, M., Caya, D., Chaumont, D., Ludwig, R., and Turcotte, R.: On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff, Hydrol. Earth Syst. Sci., 17, 1189–1204, doi:10.5194/hess-17-1189-2013, 2013.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10, 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

Neukum, C. and Azzam, R.: Impact of climate change on groundwater recharge in a small catchment in the Black Forest, Germany, Hydrogeol J, 20, 547–560, doi:10.1007/s10040-011-0827-x, 2012.

Poschlod, B.: Using high-resolution regional climate models to estimate return levels of daily extreme precipitation over Bavaria, [preprint], Nat. Hazards Earth Syst. Sci. Discuss, doi:10.5194/nhess-2021-66, 2021.

Poschlod, B., Willkofer, F., and Ludwig, R.: Impact of Climate Change on the Hydrological Regimes in Bavaria, Water, 12, 1599, doi:10.3390/w12061599, 2020.

Prein, A. F. and Gobiet, A.: Impacts of uncertainties in European gridded precipitation observations on regional climate analysis, Int. J. Climatol., 37, 305–327, doi:10.1002/joc.4706, 2017.

Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., and Gratzki, A.: A Central European precipitation climatology – Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS), metz, 22, 235–256, doi:10.1127/0941-2948/2013/0436, 2013.

Salinas, J. L., Castellarin, A., Viglione, A., Kohnová, S., and Kjeldsen, T. R.: Regional parent flood frequency distributions in Europe – Part 1: Is the GEV model suitable as a pan-European parent?, Hydrol. Earth Syst. Sci., 18, 4381–4389, doi:10.5194/hess-18-4381-2014, 2014.

Schulla, J.: Model Description WaSiM (Water balance Simulation Model), Zurich, 2021.

Schulz, K. and Bernhardt, M.: The end of trend estimation for extreme floods under climate change?, Hydrol. Process., 30, 1804–1808, doi:10.1002/hyp.10816, 2016.

Šeparović, L., Alexandru, A., Laprise, R., Martynov, A., Sushama, L., Winger, K., Tete, K., and Valin, M.: Present climate and climate change over North America as simulated by the fifth-generation Canadian regional climate model, Clim Dyn, 41, 3167–3201, doi:10.1007/s00382-013-1737-5, 2013.

Sigmond, M., Fyfe, J. C., and Swart, N. C.: Ice-free Arctic projections under the Paris Agreement, Nature Clim Change, 8, 404–408, doi:10.1038/s41558-018-0124-y, 2018.

Stahl, N. and Hofstätter, M.: Vb-Zugbahnen und deren Auftreten als Serie mit Bezug zu den resultierenden Hochwassern in Bayern und Auswirkungen auf Rückhalteräume im Isareinzugsgebiet, Hydrologie und Wasserbewirtschaftung, 62, 77–97, doi:10.5675/HyWa_2018,2_2, 2018.

Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, Journal of Hydrology, 456-457, 12–29, doi:10.1016/j.jhydrol.2012.05.052, 2012.

Thieken, A. H., Kienzler, S., Kreibich, H., Kuhlicke, C., Kunz, M., Mühr, B., Müller, M., Otto, A., Petrow, T., Pisi, S., and Schröter, K.: Review of the flood risk management system in Germany after the major flood in 2013, E&S, 21, doi:10.5751/ES-08547-210251, 2016.

Tolson, B. A. and Shoemaker, C. A.: Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, Water Resour. Res., 43, doi:10.1029/2005WR004723, 2007.

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J., and Rose, S. K.: The representative concentration pathways: an overview, Climatic Change, 109, 5–31, doi:10.1007/s10584-011-0148-z, 2011.

Velázquez, J. A., Schmid, J., Ricard, S., Muerth, M. J., Gauvin St-Denis, B., Minville, M., Chaumont, D., Caya, D., Ludwig, R., and Turcotte, R.: An ensemble approach to assess hydrological models' contribution to uncertainties in the analysis of climate change impact on water resources, Hydrol. Earth Syst. Sci., 17, 565–578, doi:10.5194/hess-17-565-2013, 2013.

von Trentini, F., Aalbers, E. E., Fischer, E. M., and Ludwig, R.: Comparing interannual variability in three regional single-model initial-condition large ensembles (SMILEs) over Europe, Earth Syst. Dynam., 11, 1013–1031, doi:10.5194/esd-11-1013-2020, 2020.

Westra, S., Fowler, H. J., Evans, J. P., Alexander, L. V., Berg, P., Johnson, F., Kendon, E. J., Lenderink, G., and Roberts, N. M.: Future changes to the intensity and frequency of short-duration extreme rainfall, Rev. Geophys., 52, 522–555, doi:10.1002/2014RG000464, 2014.

Wilhelm, B., Rapuc, W., Amann, B., Anselmetti, F. S., Arnaud, F., Blanchet, J., Brauer, A., Czymzik, M., Giguet-Covex, C., Gilli, A., Glur, L., Grosjean, M., Irmler, R., Nicolle, M., Sabatier, P., Swierczynski, T., and Wirth, S. B.: Impact of warmer climate periods on flood hazard in the European Alps, Nat. Geosci., 15, 118–123, doi:10.1038/s41561-021-00878-y, 2022.

Willkofer, F., Schmid, F.-J., Komischke, H., Korck, J., Braun, M., and Ludwig, R.: The impact of bias correcting regional climate model results on hydrological indicators for Bavarian catchments, Journal of Hydrology: Regional Studies, 19, 25–41, doi:10.1016/j.ejrh.2018.06.010, 2018.

Willkofer, F., Wood, R. R., Trentini, F. von, Weismüller, J., Poschlod, B., and Ludwig, R.: A Holistic Modelling Approach for the Estimation of Return Levels of Peak Flows in Bavaria, Water, 12, 2349, doi:10.3390/w12092349, 2020.

Wood, R. R., Lehner, F., Pendergrass, A. G., and Schlunegger, S.: Changes in precipitation variability across time scales in multiple global climate model large ensembles, Environ. Res. Lett., doi:10.1088/1748-9326/ac10dd, 2021.

Wood, R. R. and Ludwig, R.: Analyzing Internal Variability and Forced Response of Subdaily and Daily Extreme Precipitation Over Europe, Geophys. Res. Lett., 47, doi:10.1029/2020GL089300, 2020.

3 Conclusions

The scope of this dissertation is the assessment of the impacts of climate change on hydrological extreme events for the catchments of the Hydrological Bavaria within the 21^{st} century. It is analyzed whether extreme flood events will become more frequent and more intense in response to the strong increase in extreme precipitation events, as is expected for Central Europe and the Alps shown by many studies (Fowler et al., 2021; Li et al., 2021; Martel et al., 2020; Wood et al., 2021). The findings directly related to the scope of the dissertation not only confirm the results of the only other large-scale study by Hattermann et al. (2018) showing an increase in the frequency and magnitude of extreme floods in the Bavarian Danube and its tributaries but also extends the conclusions to the Bavarian Main and its tributaries. Furthermore, a novel approach of employing a RCM SMILE for hydrological modelling to foster extreme value analysis is presented. In the following, the major findings of the thesis are highlighted in association with their respective research question stated in section 1.3.

Q1: Can a holistic model setup parameterized towards high flow representation provide sufficient performance across heterogeneous catchments?

In Willkofer et al. (2020), the new approach of a semi-global and semi-automatized parameterization of a single fully-distributed and process-based hydrological model - in high spatio-temporal resolution (500m, 3h) - for the entire Hydrological Bavaria lead to a single regionalized set of parameters for all catchments. The results show that the new holistic modelling approach is able to provide a good model performance for most of the 98 gauges of the Hydrological Bavaria. Values of the NSE and KGE exceed 0.6 for the majority of the catchments for the reference period between 1981 and 2010. Therefore, this approach is able to provide sufficient performance across heterogeneous catchments. Although the presented new holistic single model approach might limit model performance of individual gauges as catchment specific characteristics can only be considered to a certain extent, it also circumvents problems arising from equifinality and over-parameterization similar to a global approach as presented in Gaborit et al. (2015). Furthermore, as presented in Willkofer et al. (2020) there is a potential to further enhance the performance of the model by the application of improved inputs. In particular, the model could benefit from an increase in precipitation over the Alps, since observational records exhibit an undercatch for this region (Poschlod et al., 2020; Prein & Gobiet, 2017) which propagated into the gridded meteorological reference dataset.

Q2: Does this holistic model setup qualify for the simulation of flood events with higher return periods?

To determine whether the model setup developed in Willkofer et al. (2020) was applicable to simulate peak discharges of higher return periods, mean high flows served as a first indicator. Furthermore, a concept for a Level of Trust (LOT) for return periods of 5-, 10-, and 20-year events was introduced which shows the confidence in the simulations based on relative deviations between model results and observations. Relative deviations in MHF depict satisfactory results with values around $\pm$20% for the majority gauges. However, for some downstream gauges along the Danube and Main river the deviations in MHF exceed 20%. The LOT depicts a similar behavior with moderate (<30% deviation) to high (<10% deviation) confidence in the simulations of the majority of gauges. This behavior confirms results by Ricard et al. (2013) for a regionalized parameter set gained through the new holistic modelling approach for the Hydrological Bavaria. Although flood events of higher return periods are of particular interest, the short reference period of only 30 years only allows for a robust estimation of peak flows of the moderately high return periods presented in this thesis. The estimation of discharges of higher return periods are prone to different sources of uncertainties such as from the natural variability of the system, the selected period for their calculation as well as from the selected approach to perform the extreme value analysis (Schulz & Bernhardt, 2016). Furthermore, the model serves to assess impacts of climate change on the hydrology of the Hydrological Bavaria. Thus, as stated in Willkofer et al. (2020) and Gaborit et al. (2015), the deviations in absolute values between the model and observations are less important for CCI studies as these focus on relative changes between periods and future trends. In conclusion, the newly developed holistic model presented in Willkofer et al. (2020) is considered to be applicable to assess future dynamics of extreme flood events, for the first time for the entire Hydrological Bavaria in high spatio-temporal resolution.

Q3: Which method for bias correction is recommended in terms of best representation of observed discharge regimes and hydrological indicators?

Willkofer et al. (2018) compared three approaches (with monthly and yearly correction factors) for bias correction for their capability to adjust RCM outputs for reproducing observed annual discharge regimes as well as a variety of hydrological indicators (LF, MLF, 7LF2, MF, MHF, HF, HF2 HF100). The analysis is based on three RCMs (two 3-member RCMs, one single member RCM) driving a single hydrological model for two catchments within the Hydrological Bavaria. Furthermore, this is the first study for this region not only employing BC on precipitation and air temperature, but also on relative air humidity, global radiation, and wind speed, as they are required for the hydrological model. The presented results show that all methods are able to reproduce the observed annual regime and mean flow indicators reasonable well, while high flow indicators (HF, HF100) show larger deviations from the observations. However, among the tested BC methods, quantile-mapping (qm) either shows the best adjustment performance or is close to other best performing methods in terms of deviations in discharge regime and flow indicators from observations. While qm has been recommended for performing bias correction of RCMs over other regions before (e.g., Themeßl et al., 2011 for the entire Austria), it is also recommended for the Hydrological Bavaria (using monthly correction factors). Further, the holistic correction of all climate variables as presented in Willkofer et al. (2018) is recommended over correcting only temperature and precipitation.

Q4: How does bias correction affect the climate change signal of hydrological indicators?

Willkofer et al. (2018) further assesses the impact of employing corrected RCM outputs for hydrological climate change impact modelling by means of climate change signals of multiple hydrological indicators. Regardless of the individual performance of a BC method, there are similar effects on the investigated flow indicators. The results show that differences in climate change signal (CCS) of the hydrological indicators between raw and corrected outputs are small for mean flow indicators across all methods with differences in relative CCS up to 15 percent points. However, regardless of the approach, bias correction has a stronger impact on the CCS of extreme value indicators (HF, HF100) with differences in actual change values exceeding 100% in extreme cases.

As stated in Willkofer et al. (2018), the applied hydrological model setups were not explicitly tailored to the representation of high flows but rather for a good representation of mean conditions, which impacts the significance of these findings. The results confirm the findings by Stagl and Hattermann (2015) and Muerth et al. (2013) for the impact of BC on the CCS of hydrological indicators for catchments within the upper Danube area. The results of this thesis extend these findings by examining multiple BC methods on multiple hydrological indicators for two catchments within the Hydrological Bavaria, one within the upper Danube and one within the Main area, providing more detailed insights.

Q5: Can SMILEs of RCMs contribute to facilitate robust estimates of peak discharges of high return periods?

The analysis presented in Willkofer et al. (2023) shows that the application of a RCM SMILE as a driver for hydrological models is beneficial for the robust estimation of peak flows of high return periods. The results for the estimation of the 100-year flood employing empirical probabilities on 1,500 events (annual maxima) provided by the hydrological SMILE leads to a robust value. In comparison, the results gained by 1,000 random samples of 30, 100, and 200 yearly peak flows exhibit a considerable variability around this robust value representing uncertainties arising from the different sample data. Although this variability decreases with increasing sample sizes as expected, even for the 200-years samples the spread of the statistical estimates is rather large compared to the robust value (between 0.56 and 1.6 times the robust value). Hence, the application of a RCM SMILE is beneficial for the estimation of these events and clearly contributes to facilitate robust values. These results confirm the synthetically generated findings by Schulz and Bernhardt (2016) regarding the uncertainties due to data and sample size through the application of a hydrological SMILE. Furthermore, the findings in Willkofer et al. (2023) confirm the results by van der Wiel et al. (2019) with regard to the benefit of the application of large ensembles to directly deduce changes in extreme values. This thesis further emphasizes this benefit for a regional impact analysis on extreme high flows through the application of a high resolution climate and hydrological model as well as transient simulations.

Q6: Will the frequency and intensity of peak discharges of high return periods decrease or increase due to climate change?

In Willkofer et al. (2023) the created hydrological SMILE is exploited to empirically deduce high flows of high return periods for a robust assessment of climate change impacts on their dynamics (i.e., changes in frequency and intensity). This impact of climate change on the dynamics of 100-year flood events varies for the different regime types. Willkofer et al. (2023) highlight that the most severe changes in dynamics are expected for catchments with nival (snowmelt) influenced flow regime characteristics within or close to the Alps. The region depicts a steady and in parts severe increase in frequency and intensity for the different periods with the largest increase of 36% to 104% in intensity and a 7 to 12 times more frequent occurrence at the end of the century. For catchments north of the Alps up to the northern parts of the Hydrological Bavaria where regimes become rainfall dominated (pluvial), the changes in dynamics of the 100-year flood are less pronounced with more than 50% of gauges depicting an increase in intensity of at least 10% to 20% and occurring 1.5 to 2 times more often at the end of the century. A few gauges even exhibit a decrease in frequency and intensity in the future. The presented results widen the spatial focus provided by Hattermann et al. (2018) (Bavarian Danube) to the entire Hydrological Bavaria including analysis for the Main river and its tributaries. Furthermore, in this study transient simulations of a RCM SMILE were employed for the first time for a large scale, high resolution, fully-distributed, and process-based hydrological model for this region and for this type of analysis on future dynamics of flood events.

In conclusion, the findings of this thesis contribute to the current scientific state of assessing the impacts of climate change on the dynamics of extreme peak flows within the Hydrological Bavaria through the integration of a newly developed RCM SMILE within the hydrological impact modelling chain.

4 Scientific Outreach

Poschlod, B., Willkofer, F., Ludwig, R. (2020). Impact of Climate Change on the Hydrological Regimes in Bavaria. Water 12 (6), p.1599, DOI: 10.3390/w12061599: In this publication the effects of climate change on the hydrological regimes of Bavarian catchments is presented. It employs results of the WaSiM hydrological large ensemble (WaSiM-LE) created for the ClimEx project (https://www.climex-project.org) which was driven by the CRCM5-LE using the RCP8.5 radiative forcing. First, the WaSiM-LE was utilized to distinguish six different discharge regimes through a hierarchical cluster analysis. Under the rather intense change conditions provided by the CRCM5-LE, the results of the hydrological model depict severe changes in discharge characteristics of all regimes types for Bavarian catchments and a shift of regime classes for almost half of the catchments towards the end of the 21^{st} century. The results used in this analysis originate from the work in Willkofer et al. (2020).

Brunner, M.I., Swain, D.L., Wood, R.R., Willkofer, F., Done, J.M., Gilleland, E., Ludwig, R. (2021). An extremeness threshold determines the regional response of floods to changes in rainfall extremes. Commun Earth Environ 2 (1), DOI: 10.1038/s43247-021-00248-x: This publication is based on the WaSiM-LE using an excerpt of its simulated gauges to investigate the responsiveness of floods to a postulated increase in precipitation extremes towards the end of the 21^{st} century. The study introduces an extremeness threshold for annuality of flood events above which an increase in flood magnitude is a direct response to an increase in precipitation. Floods events below this extremeness threshold are further affected by other catchment processes. The simulations used in this analysis originate from the work in Willkofer et al. (2020).

References

Aalbers, E. E., Lenderink, G., van Meijgaard, E., & van den Hurk, B. J. J. M. (2018). Local-scale changes in mean and heavy precipitation in Western Europe, climate change or internal variability? *Climate Dynamics*, *50*(11-12), 4745–4766. https://doi.org/10.1007/s00382-017-3901-9

Anctil, F., Perrin, C., & Andréassian, V. (2004). Impact of the length of observed records on the performance of ANN and of conceptual parsimonious rainfall-runoff forecasting models. *Environmental Modelling & Software*, *19*(4), 357–368. https://doi.org/10.1016/S1364-8152(03)00135-X

Arora, V. K., Scinocca, J. F., Boer, G. J., Christian, J. R., Denman, K. L., Flato, G. M., Kharin, V. V., Lee, W. G., & Merryfield, W. J. (2011). Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophysical Research Letters*, *38*(5), L05805. https://doi.org/10.1029/2010GL046270

Arsenault, R., Brissette, F., & Martel, J.-L. (2018). The hazards of split-sample validation in hydrological model calibration. *Journal of Hydrology*, *566*, 346–362. https://doi.org/10.1016/j.jhydrol.2018.09.027

Bárdossy, A. (2007). Calibration of hydrological model parameters for ungauged catchments. *Hydrology and Earth System Sciences*, *11*(2), 703–710. https://doi.org/10.5194/hess-11-703-2007

Bayerisches Landesamt für Umwelt. (2007a). Gewässerkundlicher Bericht: Hochwasser August 2005 (Bayerisches Landesamt für Umwelt, Ed.).

Bayerisches Landesamt für Umwelt. (2007b). Wasser in Bayern: Gewässerkundlicher Monatsbericht für Bayern - April 2007 (Bayerisches Landesamt für Umwelt, Ed.).

Bayerisches Landesamt für Umwelt. (2016). Niedrigwasser in Bayern: Grundlagen, Veränderungen und Auswirkungen (Bayerisches Landesamt für Umwelt, Ed.).

Bayerisches Landesamt für Umwelt. (2017a). Gewässerkundlicher Jahresbericht 2016 (Bayerisches Landesamt für Umwelt, Ed.).

Bayerisches Landesamt für Umwelt. (2017b). Sturzfluten- und Hochwasserereignisse Mai/Juni 2016: Wasserwirtschaftlicher Bericht (Bayerisches Landesamt für Umwelt, Ed.).

Bayerisches Landesamt für Umwelt. (2020). Niedrigwasser 2018 und 2019: Analysen und Auswirkungen für Bayern (Bayerisches Landesamt für Umwelt, Ed.).

Bayerisches Landesamt für Wasserwirtschaft. (2003). Hochwasser Mai 1999 (Bayerisches Landesamt für Wasserwirtschaft, Ed.).

Bednar-Friedl, B., Biesbroek, R., Schmidt, D. N., Alexander, P., Børsheim, K. Y., Carnicer, J., Georgopoulou, E., Haasnoot, M., Le Cozannet, G., Lionello, P., Lipka, O., Möllmann, C., Muccione, V., Mustonen, T., Piepenburg, D., & Whitmarsh, L. (2022). Europe. In H.-O. Pörtner, D. C. Roberts, M. Tignor, E. S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, & B. Rama (Eds.), *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1817–1927). Cambridge University Press. https://doi.org/10.1017/9781009325844.015

Berghuijs, W. R., Harrigan, S., Molnar, P., Slater, L. J., & Kirchner, J. W. (2019). The Relative Importance of Different Flood–Generating Mechanisms Across Europe. *Water Resources Research*. https://doi.org/10.1029/2019WR024841

Beven, K. (2001). How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences*, *5*(1), 1–12. https://doi.org/10.5194/hess-5-1-2001

Beven, K. (2019). How to make advances in hydrological modelling. *Hydrology Research*, *50*(6), 1481–1494. https://doi.org/10.2166/nh.2019.134

Blöschl, G., Nester, T., Komma, J., Parajka, J., & Perdigão, R. A. P. (2013). The June 2013 flood in the Upper Danube Basin, and comparisons with the 2002, 1954 and 1899 floods. *Hydrology and Earth System Sciences*, *17*(12), 5197–5212. https://doi.org/10.5194/hess-17-5197-2013

Blöschl, G., Gaál, L., Hall, J., Kiss, A., Komma, J., Nester, T., Parajka, J., Perdigão, R. A. P., Plavcová, L., Rogger, M., Salinas, J. L., & Viglione, A. (2015). Increasing river floods: fiction or reality? *WIREs Water*, *2*(4), 329–344. https://doi.org/10.1002/wat2.1079

Brath, A., Montanari, A., & Toth, E. (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. https://doi.org/10.1016/j.jhydrol.2003.12.044

Brunner, M. I., Slater, L., Tallaksen, L. M., & Clark, M. (2021a). Challenges in modeling and predicting floods and droughts: A review. *WIREs Water*, *8*(3). https://doi.org/10.1002/wat2.1520

Brunner, M. I., Swain, D. L., Wood, R. R., Willkofer, F., Done, J. M., Gilleland, E., & Ludwig, R. (2021b). An extremeness threshold determines the regional response of floods to changes in rainfall extremes. *Communications Earth & Environment*, *2*(1). https://doi.org/10.1038/s43247-021-00248-x

Cassalho, F., Beskow, S., de Mello, C. R., de Moura, M. M., Kerstner, L., & Ávila, L. F. (2018). At-Site Flood Frequency Analysis Coupled with Multiparameter Probability Distributions. *Water Resources Management*, *32*(1), 285–300. https://doi.org/10.1007/s11269-017-1810-7

Chen, D., Rojas, M., Samset, B. H., Cobb, K., Diongue Niang, A., Edwards, P., Emori, S., Faria, S. H., Hawkins, E., Hope, P., Huybrechts, P., Meinshausen, M., Mustafa, S. K., Plattner, G.-K., & Tréguie, A.-M. (2021). Framing, Context, and Methods. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 147–286).

Chen, J., Brissette, F. P., & Leconte, R. (2011a). Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *Journal of Hydrology*, *401*(3-4), 190–202. https://doi.org/10.1016/j.jhydrol.2011.02.020

Chen, J., Brissette, F. P., Poulin, A., & Leconte, R. (2011b). Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed. *Water Resources Research*, *47*(12). https://doi.org/10.1029/2011WR010602

Chen, Y. (2017). Distributed Hydrological Models. In Q. Duan, F. Pappenberger, J. Thielen, A. Wood, H. L. Cloke, & J. C. Schaake (Eds.), *Handbook of Hydrometeorological Ensemble Forecasting* (pp. 1–24). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40457-3_23-1

Cloke, H. L., Wetterhall, F., He, Y., Freer, J. E., & Pappenberger, F. (2013). Modelling climate impact on floods with ensemble climate projections. *Quarterly Journal of the Royal Meteorological Society*, *139*(671), 282–297. https://doi.org/10.1002/qj.1998

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W. J., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A. J., & Wehner M. (2013). Long-term Climate Change: Projections, Commitments and Irreversibility. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, & P. M. Midgle (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Cubasch, U., Wuebbles, D., Chen, D., Facchini, M. C., Frame, D., Mahowald, N., & Winthe, J.-G. (2013). Introduction. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, & P. M. Midgle (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Curceac, S., Atkinson, P. M., Milne, A., Wu, L., & Harris, P. (2020). An evaluation of automated GPD threshold selection methods for hydrological extremes across different scales. *Journal of Hydrology*, *585*, 124845. https://doi.org/10.1016/j.jhydrol.2020.124845

Dai, A. (2006). Precipitation Characteristics in Eighteen Coupled Climate Models. *Journal of Climate*, *19*(18), 4605–4630. https://doi.org/10.1175/JCLI3884.1

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

Di Virgilio, G., Ji, F., Tam, E., Nishant, N., Evans, J. P., Thomas, C., Riley, M. L., Beyer, K., Grose, M. R., Narsey, S., & Delage, F. (2022). Selecting CMIP6 GCMs for CORDEX Dynamical Downscaling: Model Performance, Independence, and Climate Change Signals. *Earth's Future*, *10*(4). https://doi.org/10.1029/2021EF002625

Duan, Q. Y., Gupta, V. K., & Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, *76*(3), 501–521. https://doi.org/10.1007/BF00939380

Durocher, M., Burn, D. H., & Ashkar, F. (2019). Comparison of Estimation Methods for a Nonstationary Index–Flood Model in Flood Frequency Analysis Using Peaks Over Threshold. *Water Resources Research*, *55*(11), 9398–9416. https://doi.org/10.1029/2019WR025305

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., & Liebert, J. (2012). HESS Opinions "Should we apply bias correction to global and regional climate model data?" *Hydrology and Earth System Sciences*, *16*(9), 3391–3404. https://doi.org/10.5194/hess-16-3391-2012

Enayati, M., Bozorg-Haddad, O., Bazrafshan, J., Hejabi, S., & Chu, X. (2021). Bias correction capabilities of quantile mapping methods for rainfall and temperature variables. *Journal of Water and Climate Change*, *12*(2), 401–419. https://doi.org/10.2166/wcc.2020.261

Fiddes, J., & Gruber, S. (2014). TopoSCALE v.1.0: downscaling gridded climate data in complex terrain. *Geoscientific Model Development*, *7*(1), 387–405. https://doi.org/10.5194/gmd-7-387-2014

Fischer, E. M., & Knutti, R. (2016). Observed heavy precipitation increase confirms theory and early models. *Nature Climate Change*, *6*(11), 986–991. https://doi.org/10.1038/nclimate3110

Fischer, S., & Schumann, A. H. (2022). Handling the stochastic uncertainty of flood statistics in regionalization approaches. *Hydrological Sciences Journal*, *67*(10), 1449–1465. https://doi.org/10.1080/02626667.2022.2091410

Fowler, H. J., Ali, H., Allan, R. P., Ban, N., Barbero, R., Berg, P., Blenkinsop, S., Cabi, N. S., Chan, S., Dale, M., Dunn, R. J. H., Ekström, M., Evans, J. P., Fosser, G., Golding, B., Guerreiro, S. B., Hegerl, G. C., Kahraman, A., Kendon, E. J., … Whitford, A. (2021). Towards advancing scientific knowledge of climate change impacts on short-duration rainfall extremes. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, *379*(2195), 20190542. https://doi.org/10.1098/rsta.2019.0542

Fyfe, J. C., Derksen, C., Mudryk, L., Flato, G. M., Santer, B. D., Swart, N. C., Molotch, N. P., Zhang, X., Wan, H., Arora, V. K., Scinocca, J., & Jiao, Y. (2017). Large near-term projected snowpack loss over the western United States. *Nature communications*, *8*, 14996. https://doi.org/10.1038/ncomms14996

Gaborit, É., Ricard, S., Lachance-Cloutier, S., Anctil, F., & Turcotte, R. (2015). Comparing global and local calibration schemes from a differential split-sample test perspective. *Canadian Journal of Earth Sciences*, *52*(11), 990–999. https://doi.org/10.1139/cjes-2015-0015

Gädeke, A., Hölzel, H., Koch, H., Pohle, I., & Grünewald, U. (2014). Analysis of uncertainties in the hydrological response of a model-based climate change impact assessment in a subcatchment of the Spree River, Germany. *Hydrological Processes*, *28*(12), 3978–3998. https://doi.org/10.1002/hyp.9933

Gagnon, P., Rousseau, A. N., Mailhot, A., & Caya, D. (2012). Spatial Disaggregation of Mean Areal Rainfall Using Gibbs Sampling. *Journal of Hydrometeorology*, *13*(1), 324–337. https://doi.org/10.1175/JHM-D-11-034.1

Gampe, D., Schmid, J., & Ludwig, R. (2019). Impact of Reference Dataset Selection on RCM Evaluation, Bias Correction, and Resulting Climate Change Signals of Precipitation. *Journal of Hydrometeorology*, *20*(9), 1813–1828. https://doi.org/10.1175/JHM-D-18-0108.1

Garambois, P. A., Roux, H., Larnier, K., Castaings, W., & Dartus, D. (2013). Characterization of process-oriented hydrologic model behavior with temporal sensitivity analysis for flash floods in Mediterranean catchments. *Hydrology and Earth System Sciences*, *17*(6), 2305–2322. https://doi.org/10.5194/hess-17-2305-2013

Gewässerkundlicher Dienst Bayern. (2002). Hochwasser im August 2002 (Bayerisches Landesamt für Wasserwirtschaft, Ed.).

Gharib, A., Davies, E., Goss, G., & Faramarzi, M. (2017). Assessment of the Combined Effects of Threshold Selection and Parameter Estimation of Generalized Pareto Distribution with Applications to Flood Frequency Analysis. *Water*, *9*(9), 692. https://doi.org/10.3390/w9090692

Giorgi, F. (2019). Thirty Years of Regional Climate Modeling: Where Are We and Where Are We Going next? *Journal of Geophysical Research: Atmospheres*. https://doi.org/10.1029/2018JD030094

Giorgi, F., & Gutowski, W. J. (2015). Regional Dynamical Downscaling and the CORDEX Initiative. *Annual Review of Environment and Resources*, *40*(1), 467–490. https://doi.org/10.1146/annurev-environ-102014-021217

Giorgi, F., Jones, C., & Asrar, G. R. (2009). Addressing climate information needs at the regional level: the CORDEX framework. *World Meteorological Organization (WMO) Bulletin*, (58(3)), 175.

GRDC. (2021). *Summary Statistics by Country*. Retrieved June 9, 2023, from https://www.bafg.de/SharedDocs/ExterneLinks/GRDC/summary_stat_cc_pdf.pdf?__blob=publicationFile

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

Hattermann, F. F., Wortmann, M., Liersch, S., Toumi, R., Sparks, N., Genillard, C., Schröter, K., Steinhausen, M., Gyalai-Korpos, M., Máté, K., Hayes, B., del Rocío Rivas López, M., Rácz, T., Nielsen, M. R., Kaspersen, P. S., & Drews, M. (2018). Simulation of flood hazard and risk in the Danube basin with the Future Danube Model. *Climate Services*, *12*, 14–26. https://doi.org/10.1016/j.cliser.2018.07.001

Haugen, M. A., Stein, M. L., Moyer, E. J., & Sriver, R. L. (2018). Estimating Changes in Temperature Distributions in a Large Ensemble of Climate Simulations Using Quantile Regression. *Journal of Climate*, *31*(20), 8573–8588. https://doi.org/10.1175/JCLI-D-17-0782.1

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., & Thépaut, J.-N. (2023). *ERA5 hourly data on single levels from 1940 to present*. Copernicus Climate Change Service (C3S) Climate Data Store. https://doi.org/10.24381/cds.adbb2d47

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., & Iguchi, T. (2014). The Global Precipitation Measurement Mission. *Bulletin of the American Meteorological Society*, *95*(5), 701–722. https://doi.org/10.1175/BAMS-D-13-00164.1

Huang, S., Krysanova, V., & Hattermann, F. F. (2014). Does bias correction increase reliability of flood projections under climate change? A case study of large rivers in Germany. *International Journal of Climatology*, *34*(14), 3780–3800. https://doi.org/10.1002/joc.3945

IPCC. (2021). Summary for Policymakers. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 3–32). https://doi.org/10.1017/9781009157896.001

IPCC. (2022). *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* (H.-O. Pörtner, D. Roberts, A. Fischlin, M. Howden, C. Méndez, J. J. Pereira, R. A. Sánchez-Rodríguez, S. Semenov, P. Yanda, & T. M. Zatari, Eds.). Cambridge University Press.

Isotta, F. A., Frei, C., Weilguni, V., Perčec Tadić, M., Lassègues, P., Rudolf, B., Pavan, V., Cacciamani, C., Antolini, G., Ratto, S. M., Munari, M., Micheletti, S., Bonati, V., Lussana, C., Ronchi, C., Panettieri, E., Marigo, G., & Vertačnik, G. (2014). The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. *International Journal of Climatology*, *34*(5), 1657–1675. https://doi.org/10.1002/joc.3794

Jackson, E. K., Roberts, W., Nelsen, B., Williams, G. P., Nelson, E. J., & Ames, D. P. (2019). Introductory overview: Error metrics for hydrologic modelling – A review of common practices and an open source library to facilitate use and adoption. *Environmental Modelling & Software*, *119*, 32–48. https://doi.org/10.1016/j.envsoft.2019.05.001

Kay, A. L., Rudd, A. C., & Coulson, J. (2023). Spatial downscaling of precipitation for hydrological modelling: Assessing a simple method and its application under climate change in Britain. *Hydrological Processes*, *37*(2). https://doi.org/10.1002/hyp.14823

Kirchmeier-Young, M. C., Zwiers, F. W., & Gillett, N. P. (2017). Attribution of Extreme Events in Arctic Sea Ice Extent. *Journal of Climate*, *30*(2), 553–571. https://doi.org/10.1175/JCLI-D-16-0412.1

Kjellström, E., Boberg, F., Castro, M., Christensen, J. H., Nikulin, G., & Sánchez, E. (2010). Daily and monthly temperature and precipitation statistics as performance indicators for regional climate models. *Climate Research*, *44*(2-3), 135–150. https://doi.org/10.3354/cr00932

Kleinn, J. (2005). Hydrologic simulations in the Rhine basin driven by a regional climate model. *Journal of Geophysical Research: Atmospheres*, *110*(D4). https://doi.org/10.1029/2004JD005143

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

Kohn, I., Rosin, K., Freudiger, D., Belz, J. U., Stahl, K., & Weiler, M. (2014). Niedrigwasser in Deutschland 2011. *Hydrologie und Wasserbewirtschaftung*, (58(1)), 4–17. https://doi.org/10.5675/HyWa_2014,1_1

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

Kreienkamp, F., Philip, S. Y., Tradowsky, J. S., Kew, S. F., Lorenz, P., Arrighi, J., Belleflamme, A., Bettmann, T., Caluwaerts, S., Chan, S. C., Ciavarella, A., de Cruz, L., de Vries Hylke, Demuth, N., Ferrone, A., Fischer, E. M., Fowler, H. J., Goergen, K., Heinrich, D., … Wanders, N. (2021). *Rapid attribution of heavy rainfall events leading to the severe flooding in Western Europe during July 2021*. World Weather Attribution. Retrieved July 30, 2023, from https://www.worldweatherattribution.org/wp-content/uploads/Scientific-report-Western-Europe-floods-2021-attribution.pdf

Kunstmann, H., Heckl, A., & Rimmer, A. (2006). Physically based distributed hydrological modelling of the Upper Jordan catchment and investigation of effective model equations. *Advances in Geosciences*, *9*, 123–130. https://doi.org/10.5194/adgeo-9-123-2006

Lawrence, D. (2020). Uncertainty introduced by flood frequency analysis in projections for changes in flood magnitudes under a future climate in Norway. *Journal of Hydrology: Regional Studies*, *28*, 100675. https://doi.org/10.1016/j.ejrh.2020.100675

Leander, R., & Buishand, T. A. (2007). Resampling of regional climate model output for the simulation of extreme river flows. *Journal of Hydrology*, *332*(3-4), 487–496. https://doi.org/10.1016/j.jhydrol.2006.08.006

Leander, R., Buishand, T. A., van den Hurk, B. J., & de Wit, M. J. (2008). Estimated changes in flood quantiles of the river Meuse from resampling of regional climate model output. *Journal of Hydrology*, *351*(3-4), 331–343. https://doi.org/10.1016/j.jhydrol.2007.12.020

Leduc, M., Mailhot, A., Frigon, A., Martel, J.-L., Ludwig, R., Brietzke, G. B., Giguère, M., Brissette, F., Turcotte, R., Braun, M., & Scinocca, J. (2019). The ClimEx Project: A 50-Member Ensemble of Climate Change Projections at 12-km Resolution over Europe and Northeastern North America with the Canadian Regional Climate Model (CRCM5). *Journal of Applied Meteorology and Climatology*, *58*(4), 663–693. https://doi.org/10.1175/JAMC-D-18-0021.1

Lehner, F., Hawkins, E., Sutton, R., Pendergrass, A. G., & Moore, F. C. (2023). New Potential to Reduce Uncertainty in Regional Climate Projections by Combining Physical and Socio–Economic Constraints. *AGU Advances*, *4*(4). https://doi.org/10.1029/2023AV000887

Lenderink, G., Buishand, A., & van Deursen, W. (2007). Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach. *Hydrology and Earth System Sciences*, *11*(3), 1145–1159. https://doi.org/10.5194/hess-11-1145-2007

Li, C., Zwiers, F., Zhang, X., Li, G., Sun, Y., & Wehner, M. (2021). Changes in Annual Extremes of Daily Temperature and Precipitation in CMIP6 Models. *Journal of Climate*, *34*(9), 3441–3460. https://doi.org/10.1175/JCLI-D-19-1013.1

Liu, Z., Wang, Y., Xu, Z., & Duan, Q. (2017). Conceptual Hydrological Models. In Q. Duan, F. Pappenberger, J. Thielen, A. Wood, H. L. Cloke, & J. C. Schaake (Eds.), *Handbook of Hydrometeorological Ensemble Forecasting* (pp. 1–23). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40457-3_22-1

Madsen, H. (2003). Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Advances in Water Resources*, *26*(2), 205–216. https://doi.org/10.1016/S0309-1708(02)00092-1

Maher, N., Milinski, S., & Ludwig, R. (2021). Large ensemble climate model simulations: introduction, overview, and future prospects for utilising multiple types of large ensemble. *Earth System Dynamics*, *12*(2), 401–418. https://doi.org/10.5194/esd-12-401-2021

Maher, N., Milinski, S., Suarez–Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., Kröger, J., Takano, Y., Ghosh, R., Hedemann, C., Li, C., Li, H., Manzini, E., Notz, D., Putrasahan, D., Boysen, L., Claussen, M., Ilyina, T., Olonscheck, D., … Marotzke, J. (2019). The Max Planck Institute Grand Ensemble: Enabling the Exploration of Climate System Variability. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2050–2069. https://doi.org/10.1029/2019MS001639

Maniak, U. (2010). *Hydrologie und Wasserwirtschaft*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-05396-2

Maraun, D. (2016). Bias Correcting Climate Change Simulations - a Critical Review. *Current Climate Change Reports*, *2*(4), 211–220. https://doi.org/10.1007/s40641-016-0050-x

Marke, T. (2008). *Development and Application of a Model Interface to couple Land Surface Models with Regional Climate Models for Climate Change Risk Assessment in the Upper Danube Watershed*. Ludwig-Maximilians-Universität München.

Martel, J.-L., Mailhot, A., & Brissette, F. (2020). Global and Regional Projected Changes in 100-yr Subdaily, Daily, and Multiday Precipitation Extremes Estimated from Three Large Ensembles of Climate Simulations. *Journal of Climate*, *33*(3), 1089–1103. https://doi.org/10.1175/JCLI-D-18-0764.1

McCollum, J., & Beighley, E. (2019). Flood Frequency Hydrology with Limited Data for the Weser River Basin, Germany. *Journal of Hydrologic Engineering*, *24*(3). https://doi.org/10.1061/(ASCE)HE.1943-5584.0001713

Meehl, G. A., Boer, G.J., Covey, C., Latif, M., & Stouffer, R. J. (2000). The Coupled Model Intercomparison Project (CMIP). *Bulletin of the American Meteorological Society*, (81(2)), 313–318.

Meinshausen, M., Nicholls, Z. R. J., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J. G., Daniel, J. S., John, A., Krummel, P. B., Luderer, G., Meinshausen, N., Montzka, S. A., Rayner, P. J., Reimann, S., … Wang, R. H. J. (2020). The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. *Geoscientific Model Development*, *13*(8), 3571–3605. https://doi.org/10.5194/gmd-13-3571-2020

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J.-F., Matsumoto, K., Montzka, S. A., Raper, S. C. B., Riahi, K., Thomson, A., Velders, G. J. M., & van Vuuren, D. P. (2011). The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, *109*(1-2), 213–241. https://doi.org/10.1007/s10584-011-0156-z

Meresa, H., Tischbein, B., & Mekonnen, T. (2022). Climate change impact on extreme precipitation and peak flood magnitude and frequency: observations from CMIP6 and hydrological models. *Natural Hazards*, *111*(3), 2649–2679. https://doi.org/10.1007/s11069-021-05152-3

Merz, R., Parajka, J., & Blöschl, G. (2011). Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resources Research*, *47*(2). https://doi.org/10.1029/2010WR009505

Mittermeier, M., Braun, M., Hofstätter, M., Wang, Y., & Ludwig, R. (2019). Detecting Climate Change Effects on Vb Cyclones in a 50–Member Single–Model Ensemble Using Machine Learning. *Geophysical Research Letters*, *46*(24), 14653–14661. https://doi.org/10.1029/2019GL084969

Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. V., & Kumar, R. (2019). On the choice of calibration metrics for "high-flow" estimation using hydrologic models. *Hydrology and Earth System Sciences*, *23*(6), 2601–2614. https://doi.org/10.5194/hess-23-2601-2019

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., Kram, T., Meehl, G. A., Mitchell, J. F. B., Nakicenovic, N., Riahi, K., Smith, S. J., Stouffer, R. J., Thomson, A. M., Weyant, J. P., & Wilbanks, T. J. (2010). The next generation of scenarios for climate change research and assessment. *Nature*, *463*(7282), 747–756. https://doi.org/10.1038/nature08823

Mpelasoka, F. S., & Chiew, F. H. S. (2009). Influence of Rainfall Scenario Construction Methods on Runoff Projections. *Journal of Hydrometeorology*, *10*(5), 1168–1183. https://doi.org/10.1175/2009JHM1045.1

Muerth, M. J., Gauvin St-Denis, B., Ludwig, R., & Caya, D. (2012). Evaluation of different sources of uncertainty in climate change impact research using a hydro-climatic model ensemble. *International Congress on Environmental Modelling and Software*, (377). https://doi.org/10.5282/ubm/epub.14094

Muerth, M. J., Gauvin St-Denis, B., Ricard, S., Velázquez, J. A., Schmid, J., Minville, M., Caya, D., Chaumont, D., Ludwig, R., & Turcotte, R. (2013). On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff. *Hydrology and Earth System Sciences*, *17*(3), 1189–1204. https://doi.org/10.5194/hess-17-1189-2013

Nakicenovic, N., & Swart, R. (Eds.). (2000). *Special Report on Emission Scenarios*. Cambridge University Press.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

O'Gorman, P. A., & Schneider, T. (2009). Scaling of Precipitation Extremes over a Wide Range of Climates Simulated with an Idealized GCM. *Journal of Climate*, *22*(21), 5676–5685. https://doi.org/10.1175/2009JCLI2701.1

Pan, X., Rahman, A., Haddad, K., & Ouarda, T. B. M. J. (2022). Peaks-over-threshold model in flood frequency analysis: a scoping review. *Stochastic Environmental Research and Risk Assessment*, *36*(9), 2419–2435. https://doi.org/10.1007/s00477-022-02174-6

Paul, P. K., Zhang, Y., Mishra, A., Panigrahy, N., & Singh, R. (2019). Comparative Study of Two State-of-the-Art Semi-Distributed Hydrological Models. *Water*, *11*(5), 871. https://doi.org/10.3390/w11050871

Pechlivanidis, I. G., Jackson, B. M., McIntyre, N. R., & Wheater, H. S. (2011). Catchment Scale Hydrological Modelling: A Review of Model Types, Calibration Approaches and Uncertainty Analysis Methods in the Context of recent Developments in Technology and Applications. *Global NEST Journal*, *2011*(13(3)), 193–214. https://doi.org/10.30955/gnj.000778

Piani, C., Weedon, G. P., Best, M., Gomes, S. M., Viterbo, P., Hagemann, S., & Haerter, J. O. (2010). Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *Journal of Hydrology*, *395*(3-4), 199–215. https://doi.org/10.1016/j.jhydrol.2010.10.024

Pokhrel, P., Yilmaz, K. K., & Gupta, H. V. (2012). Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures. *Journal of Hydrology*, *418-419*, 49–60. https://doi.org/10.1016/j.jhydrol.2008.12.004

Poschlod, B., Ludwig, R., & Sillmann, J. (2021). Ten-year return levels of sub-daily extreme precipitation over Europe. *Earth System Science Data*, *13*(3), 983–1003. https://doi.org/10.5194/essd-13-983-2021

Poschlod, B., Willkofer, F., & Ludwig, R. (2020). Impact of Climate Change on the Hydrological Regimes in Bavaria. *Water*, *12*(6), 1599. https://doi.org/10.3390/w12061599

Prein, A. F., & Gobiet, A. (2017). Impacts of uncertainties in European gridded precipitation observations on regional climate analysis. *International Journal of Climatology*, *37*(1), 305–327. https://doi.org/10.1002/joc.4706

Rajczak, J., Kotlarski, S., Salzmann, N., & Schär, C. (2016). Robust climate scenarios for sites with sparse observations: a two-step bias correction approach. *International Journal of Climatology*, *36*(3), 1226–1243. https://doi.org/10.1002/joc.4417

Rajczak, J., & Schär, C. (2017). Projections of Future Precipitation Extremes Over Europe: A Multimodel Assessment of Climate Simulations. *Journal of Geophysical Research: Atmospheres*, *122*(20), 10, 773–10, 800. https://doi.org/10.1002/2017JD027176

Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., & Gratzki, A. (2013). A Central European precipitation climatology ◎ Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS). *Meteorologische Zeitschrift*, *22*(3), 235–256. https://doi.org/10.1127/0941-2948/2013/0436

Ricard, S., Bourdillon, R., Roussel, D., & Turcotte, R. (2013). Global Calibration of Distributed Hydrological Models for Large-Scale Applications. *Journal of Hydrologic Engineering*, *18*(6), 719–721. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000665

Salas, J. D., Obeysekera, J., & Vogel, R. M. (2018). Techniques for assessing water infrastructure for nonstationary extreme events: a review. *Hydrological Sciences Journal*, *63*(3), 325–352. https://doi.org/10.1080/02626667.2018.1426858

Salathé, E. P., Leung, L. R., Qian, Y., & Zhang, Y. (2010). Regional climate model projections for the State of Washington. *Climatic Change*, *102*(1-2), 51–75. https://doi.org/10.1007/s10584-010-9849-y

Schmidli, J., Frei, C., & Vidale, P. L. (2006). Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods. *International Journal of Climatology*, *26*(5), 679–689. https://doi.org/10.1002/joc.1287

Schulla, J. (2021). Model Description WaSiM (Water balance Simulation Model) (Hydrology Software Consulting J. Schulla, Ed.).

Schulz, K., & Bernhardt, M. (2016). The end of trend estimation for extreme floods under climate change? *Hydrological Processes*, *30*(11), 1804–1808. https://doi.org/10.1002/hyp.10816

Sharifi, E., Saghafian, B., & Steinacker, R. (2019). Downscaling Satellite Precipitation Estimates With Multiple Linear Regression, Artificial Neural Networks, and Spline Interpolation Techniques. *Journal of Geophysical Research: Atmospheres*, *124*(2), 789–805. https://doi.org/10.1029/2018JD028795

Shen, H., Tolson, B. A., & Mai, J. (2022). Time to Update the Split–Sample Approach in Hydrological Model Calibration. *Water Resources Research*, *58*(3). https://doi.org/10.1029/2021WR031523

Slater, L. J., Anderson, B., Buechel, M., Dadson, S., Han, S., Harrigan, S., Kelder, T., Kowal, K., Lees, T., Matthews, T., Murphy, C., & Wilby, R. L. (2021). Nonstationary weather and water extremes: a review of methods for their detection, attribution, and management. *Hydrology and Earth System Sciences*, *25*(7), 3897–3935. https://doi.org/10.5194/hess-25-3897-2021

Smiatek, G., Kunstmann, H., & Senatore, A. (2016). EURO–CORDEX regional climate model analysis for the Greater Alpine Region: Performance and expected future change. *Journal of Geophysical Research: Atmospheres*, *121*(13), 7710–7728. https://doi.org/10.1002/2015JD024727

Solari, S., & Losada, M. A. (2012). A unified statistical model for hydrological variables including the selection of threshold for the peak over threshold method. *Water Resources Research*, *48*(10). https://doi.org/10.1029/2011WR011475

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., & Miller, H. L. (Eds.). (2007). *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Sperna Weiland, F., Stuparu, D., de Winter, R., & Haasnoot, M. (2022). Improving hydrological climate impact assessments using multirealizations from a global climate model. *Journal of Flood Risk Management*, *15*(2). https://doi.org/10.1111/jfr3.12787

Stagl, J., & Hattermann, F. (2015). Impacts of Climate Change on the Hydrological Regime of the Danube River and Its Tributaries Using an Ensemble of Climate Scenarios. *Water*, *7*(11), 6139–6172. https://doi.org/10.3390/w7116139

Stahl, N., & Hofstätter, M. (2018). Vb-Zugbahnen und deren Auftreten als Serie mit Bezug zu den resultierenden Hochwassern in Bayern und Auswirkungen auf Rückhalteräume im Isareinzugsgebiet. *Hydrologie und Wasserbewirtschaftung*, *62*(2), 77–97. https://doi.org/10.5675/HyWa_2018,2_2

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., J. Boschung, J., Nauels, A., Xia, Y., Bex V., & Midgley, P. M. (Eds.). (2013). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Suarez-Gutierrez, L., Li, C., Müller, W. A., & Marotzke, J. (2018). Internal variability in European summer temperatures at 1.5 °C and 2 °C of global warming. *Environmental Research Letters*, *13*(6), 064026. https://doi.org/10.1088/1748-9326/aaba58

Suman, M., Maity, R., & Kunstmann, H. (2022). Precipitation of Mainland India: Copula–based bias–corrected daily CORDEX climate data for both mean and extreme values. *Geoscience Data Journal*, *9*(1), 58–73. https://doi.org/10.1002/gdj3.118

Szymczak, S., Backendorf, F., Bott, F., Fricke, K., Junghänel, T., & Walawender, E. (2022). Impacts of Heavy and Persistent Precipitation on Railroad Infrastructure in July 2021: A Case Study from the Ahr Valley, Rhineland-Palatinate, Germany. *Atmosphere*, *13*(7), 1118. https://doi.org/10.3390/atmos13071118

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Teutschbein, C., & Seibert, J. (2010). Regional Climate Models for Hydrological Impact Studies at the Catchment Scale: A Review of Recent Modeling Strategies. *Geography Compass*, *4*(7), 834–860. https://doi.org/10.1111/j.1749-8198.2010.00357.x

Teutschbein, C., & Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology*, *456-457*, 12–29. https://doi.org/10.1016/j.jhydrol.2012.05.052

Themeßl, M. J., Gobiet, A., & Leuprecht, A. (2011). Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *International Journal of Climatology*, *31*(10), 1530–1544. https://doi.org/10.1002/joc.2168

Thieken, A. H., Kienzler, S., Kreibich, H., Kuhlicke, C., Kunz, M., Mühr, B., Müller, M., Otto, A., Petrow, T., Pisi, S., & Schröter, K. (2016). Review of the flood risk management system in Germany after the major flood in 2013. *Ecology and Society*, *21*(2). https://doi.org/10.5751/ES-08547-210251

Tolson, B. A., & Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, *43*(1). https://doi.org/10.1029/2005WR004723

Tong, R., Parajka, J., Salentinig, A., Pfeil, I., Komma, J., Széles, B., Kubáň, M., Valent, P., Vreugdenhil, M., Wagner, W., & Blöschl, G. (2021). The value of ASCAT soil moisture and MODIS snow cover data for calibrating a conceptual hydrologic model. *Hydrology and Earth System Sciences*, *25*(3), 1389–1410. https://doi.org/10.5194/hess-25-1389-2021

van der Wiel, K., Wanders, N., Selten, F. M., & Bierkens, M. F. P. (2019). Added Value of Large Ensemble Simulations for Assessing Extreme River Discharge in a 2 °C Warmer World. *Geophysical Research Letters*, *46*(4), 2093–2102. https://doi.org/10.1029/2019GL081967

van Kempen, G., van der Wiel, K., & Melsen, L. A. (2021). The impact of hydrological model structure on the simulation of extreme runoff events. *Natural Hazards and Earth System Sciences*, *21*(3), 961–976. https://doi.org/10.5194/nhess-21-961-2021

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J., & Rose, S. K. (2011). The representative concentration pathways: an overview. *Climatic Change*, *109*(1-2), 5–31. https://doi.org/10.1007/s10584-011-0148-z

Vergara–Temprado, J., Ban, N., & Schär, C. (2021). Extreme Sub–Hourly Precipitation Intensities Scale Close to the Clausius–Clapeyron Rate Over Europe. *Geophysical Research Letters*, *48*(3). https://doi.org/10.1029/2020GL089506

Wagner, W., Blöschl, G., Pampaloni, P., Calvet, J.-C., Bizzarri, B., Wigneron, J.-P., & Kerr, Y. (2007). Operational readiness of microwave remote sensing of soil moisture for hydrologic applications. *Hydrology Research*, *38*(1), 1–20. https://doi.org/10.2166/nh.2007.029

Westra, S., Alexander, L. V., & Zwiers, F. W. (2013). Global Increasing Trends in Annual Maximum Daily Precipitation. *Journal of Climate*, *26*(11), 3904–3918. https://doi.org/10.1175/JCLI-D-12-00502.1

Wheater, H. S., Jakeman, A. J., & Beven, K. J. (1995). Progress and directions in rainfall-runoff modelling. In A. J. Jakeman (Ed.), *Modelling change in environmental systems* (Reprinted.). Wiley.

Willkofer, F., Wood, R. R., & Ludwig, R. (2023). Assessing the impact of climate change on high return levels of peak flows in Bavaria applying the CRCM5 Large Ensemble. *Hydrology and Earth System Sciences*.

Willkofer, F., Schmid, F.-J., Komischke, H., Korck, J., Braun, M., & Ludwig, R. (2018). The impact of bias correcting regional climate model results on hydrological indicators for Bavarian catchments. *Journal of Hydrology: Regional Studies*, *19*, 25–41. https://doi.org/10.1016/j.ejrh.2018.06.010

Willkofer, F., Wood, R. R., von Trentini, F., Weismüller, J., Poschlod, B., & Ludwig, R. (2020). A Holistic Modelling Approach for the Estimation of Return Levels of Peak Flows in Bavaria. *Water*, *12*(9), 2349. https://doi.org/10.3390/w12092349

Wood, R. R., & Ludwig, R. (2020). Analyzing Internal Variability and Forced Response of Subdaily and Daily Extreme Precipitation Over Europe. *Geophysical Research Letters*, *47*(17). https://doi.org/10.1029/2020GL089300

Wood, R. R., Lehner, F., Pendergrass, A. G., & Schlunegger, S. (2021). Changes in precipitation variability across time scales in multiple global climate model large ensembles. *Environmental Research Letters*. https://doi.org/10.1088/1748-9326/ac10dd

Xu, C.-y., Xiong, L., & Singh, V. P. (2017). Black-Box Hydrological Models. In Q. Duan, F. Pappenberger, J. Thielen, A. Wood, H. L. Cloke, & J. C. Schaake (Eds.), *Handbook of Hydrometeorological Ensemble Forecasting* (pp. 1–48). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40457-3_21-1

Zelle, H., van Jan Oldenborgh, G., Burgers, G., & Dijkstra, H. (2005). El Niño and Greenhouse Warming: Results from Ensemble Simulations with the NCAR CCSM. *Journal of Climate*, *18*(22), 4669–4683. https://doi.org/10.1175/JCLI3574.1

Zhao, X., Zhang, Z., Cheng, W., & Zhang, P. (2019). A New Parameter Estimator for the Generalized Pareto Distribution under the Peaks over Threshold Framework. *Mathematics*, *7*(5), 406. https://doi.org/10.3390/math7050406