Modiri, Ehsan

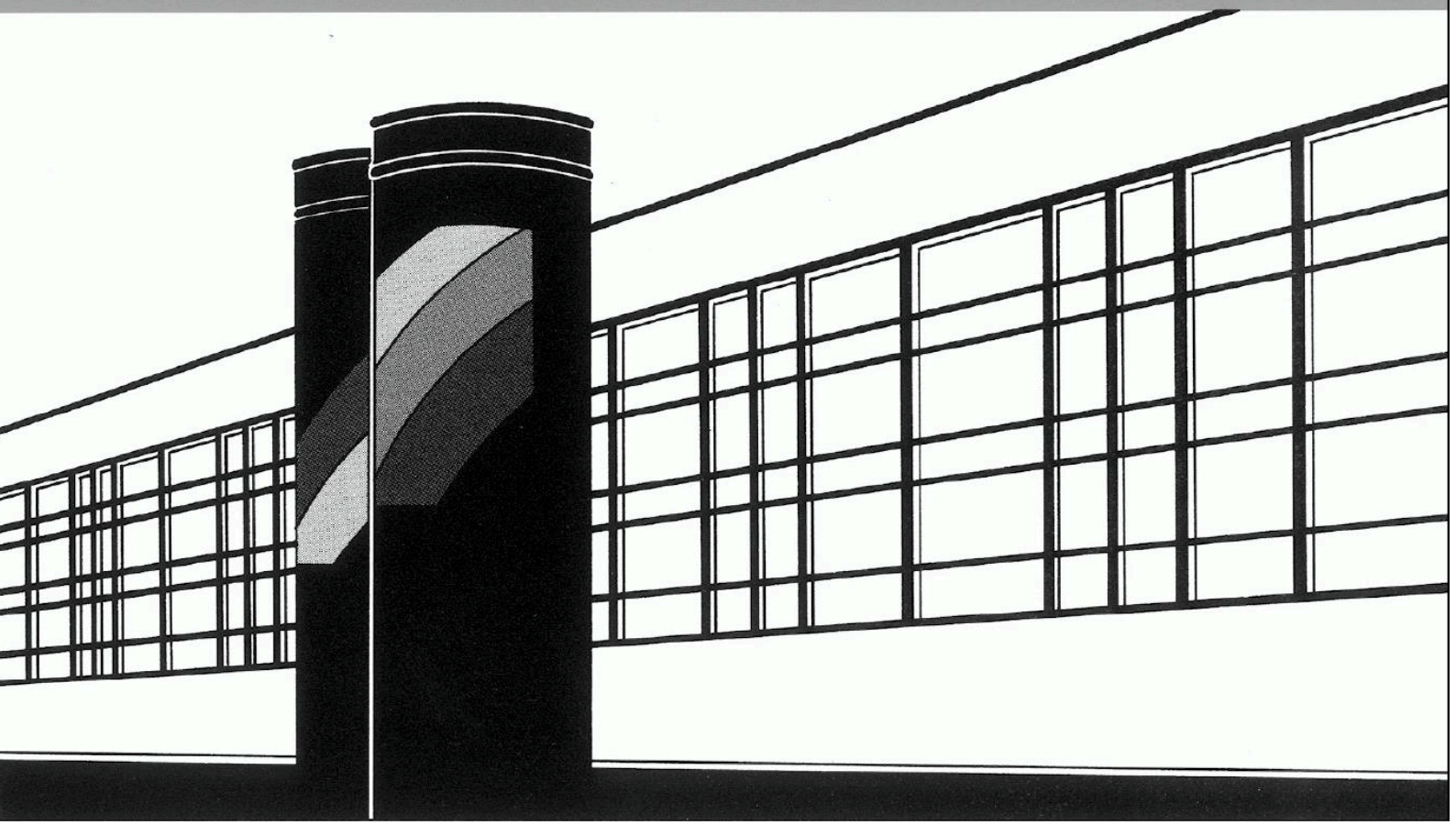
Clustering simultaneous occurrences of extreme floods in the Neckar catchment

Mitteilungen. Institut für Wasser- und Umweltsystemmodellierung, Universität Stuttgart

Universität Stuttgart

Institut für Wasser- und Umweltsystemmodellierung

Mitteilungen



Heft 288 Ehsan Modiri

Clustering simultaneous occurrences of extreme floods in the Neckar catchment

Clustering simultaneous occurrences of extreme floods in the Neckar catchment

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

vorgelegt von

Ehsan Modiri

aus Teheran, Iran

Hauptberichter:	Prof. Dr. rer. nat. Dr.-Ing. András Bárdossy
Mitberichter:	Prof. Dr. Bruno Merz

Tag der mündlichen Prüfung: 03.03.2022

Institut für Wasser- und Umweltsystemmodellierung
der Universität Stuttgart
2022

Gedruckt und/oder veröffentlicht mit Unterstützung des Deutschen Akademischen Austauschdienstes

Heft 288

Clustering simultaneous occurrences of extreme floods in the Neckar catchment

von
Dr.-Ing.
Ehsan Modiri

Eigenverlag des Instituts für Wasser- und Umweltsystemmodellierung
der Universität Stuttgart

D93 Clustering simultaneous occurrences of extreme floods in the Neckar catchment

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

Modiri, Ehsan:
Clustering simultaneous occurrences of extreme floods in the Neckar catchment, Universität Stuttgart. - Stuttgart: Institut für Wasser- und Umweltsystemmodellierung, 2022

(Mitteilungen Institut für Wasser- und Umweltsystemmodellierung, Universität Stuttgart: H. 288)
Zugl.: Stuttgart, Univ., Diss., 2022
ISBN 978-3-942036-92-4

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

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

Druck: DCC Kästl e.K., Ostfildern

Acknowledgment

In the beginning, I would like to express my sincere gratitude to my supervisor, Prof. Dr. rer. nat. Dr.-Ing. András Bárdossy for his kind supports, warm encouragements, competent guidances, and valuable comments. He established a friendly research environment in the department of hydrology and geohydrology (LHG) of the institute for modeling hydraulic and environmental systems (IWS) at the University of Stuttgart. He gave me the freedom to create my own path and scientific network. What I have learned from him is not only scientific knowledge but also the way of thinking. For all of this, I owe him more than I can describe.
I would like to express my deep gratitude to Prof. Dr. Bruno Merz for giving me the opportunity to work with him. His precious comments have opened new doors for research possibilities, from which my thesis benefited tremendously.
My gratitude extends to Prof. Dr.-Ing. habil. Jörn Birkmann for agreeing to evaluate this thesis. During my qualification and final exams, his comments and questions let me think more about some other aspects of my research.

I am deeply grateful to my dear brother Dr. -Ing Sadegh Modiri for his generous help and non-stop support since 1989. Some implemented ideas of this Ph.D. research have been laid through discussions with him and his scientific network. He always provides solutions to my questions and scripting. Indeed, my younger brother played an advisory role in my thesis. I am really proud of you, Sadegh.
I would like to acknowledge the International Doctoral Program Environment Water (ENWAT) at the University of Stuttgart for providing an academic framework for my study. Also, sincere thanks to Dr. -Ing. Gabriele M. Hartmann, the course director of the ENWAT program, who supports me to begin my journey as a Ph.D. student.

I would like to acknowledge the Sustainable Water Management program (NaWaM) from the Federal Ministry of Education and Research (BMBF) and the German Academic Exchange Service (DAAD). Furthermore, thanks to the International Office of the University of Stuttgart presenting the DAAD-STIBET scholarship and Bürkert Stiftung to support my Ph.D.

I also much appreciate all of my colleagues at LHG for their help, feedback, and assistance. I would like to give my special thanks to my friend and colleague, Vahid Shoarinezhad for his positive energy, Faizan Anwar, Dhiraj Raj Gyawali, Dr. -Ing Dirk Schlabing, and Dr. rer.nat. Jochen Seidel for their friendly support. Many thanks go to my former officemates Micha Eisele and Naibin Song. I would like to express my dear colleagues, Antônio Alves, Bushra Amin, Abbas El Hachem, Iman Fatehi, Masoud Mehrvand, Claudia Teutsch, Ning

Wang, and Dr. -Ing Jieru Yan, my former colleague, as well as my dear friend, Mostafa Hoseini from the Norwegian University of Science and Technology, Norway.
Special thanks to Astrid Lemp, whose door was always opened for help, support, and discussions. She was always available to back us up to handle our official and bureaucratic issues.

Last but not least, I am very grateful for my parents, who are always and forever my supervisors, and my beloved siblings Elham and Sadegh, for all the encouragement and support they gave to me. This dissertation is dedicated to them. You are the one who motivates in getting me to this point.
I especially thank Zeinab, my love and my constant friend. For many years, she has been not only my wife but also my best friend. I thank you from the bottom of my heart for your efforts. My dear, without you, it would not have been possible to walk through this path. The birth of our son, Niksan, a year before my Ph.D. defense, promises a sweeter life for us. I present this thesis to you, dear ones.

Contents

List of Figures

List of Tables

Abbreviations

AHCT	Agglomerative Hierarchical Cluster Tree
ARI	Adjusted Rand index
BW	Baden-Württemberg
CH	Calinski-Harabasz
CDF	Cumulative Distribution Function
DB	Davies-Bouldin
DEM	Digital Elevation Model
DWD	Deutscher Wetterdienst
FDC	Flow duration curve
FFT	fast Fourier transform
FM	Fowlkes–Mallows index
KS	Kolmogorov-Smirnov test
LS line	Least-Squares linear fit
LUBW	Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg
MDS	Multidimensional Scaling
PCA	Principal Component Analysis
PRESS	Predicted sum of squares
RFFA	Regional flood frequency analysis
RI	Rand index
RSA	Robust Simulated Annealing
SA	Simulated Annealing
SSE	Sum of Squared Errors
SVD	Singular Value Decomposition
UPGMA	Unweighted Pair Group Method with Arithmetic mean
VRC	Variance Ratio Criterion

Abstract

Floods are rare phenomena that can impose severe human and financial consequences. Usually, in flood analysis, the differences in simultaneity and interaction of floods in different regions are not considered. Insufficient knowledge on managing these extreme events may lead to a catastrophic disaster. Due to different precipitation intensities and forms, distinct atmospheric circulation patterns, and seasons, floods have different types. Also, extreme events' triggering mechanisms could be different when some individual events happen at the same time in multi-site locations. Large-scale precipitation combined with melting snow in different locations of a region brings discharge flow peaks from different areas together and simultaneously leading extreme floods. Coincidence flood occurrences can cause unpredictable losses and destruction downstream. Another approach of flood frequency analysis is investigating the general flood occurrences. These floods can happen individually in different parts of a catchment. However, they can be shared partly with simultaneous floods. Traditionally, researchers focus more on the univariate analysis and have less attention on the multivariate analysis. The multivariate extreme analysis is a method that can investigate spatiotemporal information on flood behavior. The main concern of this research is understanding flood behavior regarding the simultaneous and general occurrence that can contribute to efficient flood risk management. Also, this research's objective is to comprehend the nexus of a catchment, river system processes, geological aspects, and their interactions leading to peak river floods. Understanding the clustering of extreme floods can give researchers more insight into flood defense and protection. Clustering can divide a region into distinct areas that demand different action plans for flood management and risk analysis. Also, the insurance and reinsurance industries can calculate new coefficients to estimate risk and damages for each cluster area. The multidimensional flood clustering of two flood occurrence types is another considering point applied in this research. According to the problems faced, clustering methods are assessed for simultaneous and general flood occurrences to suggest appropriate methods for further research. With this backdrop, this dissertation discusses the spatiotemporal multivariate analysis and quantitative characteristics of simultaneous and general floods of the Neckar River and its tributaries to provide new insights to flood protection and risk analysis. The Neckar has an annual discharge average of 158 m^3/sec and the highest calculated daily areal precipitation of 97 mm over the catchment.

This thesis investigates 55 years of daily discharge flow from 1961 to 2015 for 46 measurement gauges in the Neckar catchment. First, simultaneous occurrences of extreme floods are investigated. Then, hierarchical clustering and multidimensional scaling are performed to cluster the sub-catchments that react similarly. To perform clustering methods, two dissimilarity metrics and five linkage methods are calculated and applied. Obtained clusters are then evaluated to determine the best possible clustering map. As a result, the catchment is divided into three cluster areas. Due to the geophysical complexity of the upper Neckar,

this area is separated into west and east. The rest of the catchment is also grouped as a separate cluster. In the next step, the Principal Component Analysis (PCA) is applied to the data set to capture the primary behavior of the discharge signals over the Neckar. Here, the PCA cross-validation is performed to find the best number of components for the time series reconstruction. The three first principal components are chosen to reconstruct a new time series. Then, the difference between the original and reconstructed time series matrices is obtained as a residual time series. Afterward, independent simultaneous peaks per year are selected from the absolute residual matrix. Subsequently, as the distance matrix, the correlation between the pair sets of selected peaks is determined to assess the relationship of the simultaneous events in different sub-catchments. Next, hierarchical clustering is applied to cluster the multi-site catchments into different groups with a similar reaction. As a result, clusters are defined to show the region with different behavior concerning a flood reacting mechanism. Finally, the region is divided into the three main clusters, two of which belong to the specific geological karst and crystalline rock basin and the highest elevation. The results illustrate a particular pattern for flood occurrence magnitude corresponding to each group of sub-catchments. The applied PCA-AHCT (Agglomerative Hierarchical Cluster Tree) are compared with those of the AHCT methods in the previous steps. Consequently, the similarities among acquired clusters are statistically assessed. Thus, this clustering can be a fundamental and innovative method in multivariate analysis of risk analysis and hydrological modeling by considering the simultaneous occurrence of floods.

The general floods are clustered using a distribution-based dissimilarity matrix. In the first instance, the general floods are identified. Then the Kolmogorov-Smirnov statistics are calculated from the empirical cumulative distribution functions of the general flood series. Finally, hierarchical clustering is performed to find possible groups of floods in different sub-catchments. The obtained clusters are evaluated and validated with three clustering evaluation criteria. Also, the optimum number of clusters for the Neckar catchments is determined. Five different linkage methods lead to different clustering results; however, the catchment center area is clustered in the same group in all of them. Usually, Jagst and Kocher rivers react distinctly from the other part. Also, the first upstream sub-catchments plus a part of the Enz are in the same cluster. An optimization scheme clustering is the applied idea to innovatively cluster the general floods with a less possible presumption for clustering, such as the number of clusters. The simulated annealing is employed in an intertwined and robust way to cluster extreme flood series. Here, the Silhouette coefficient is used for optimization's objective function role. As a result, the Robust Simulated Annealing is introduced as a new tool for multidimensional clustering. However, this method needs to be investigated and developed further.

The applied flood frequency analysis showed that the flow-connected regions on the mainstream reacted similarly when general floods happened. The topography and geography had more impacts on simultaneous occurrences of severe floods. Due to this, the upper Neckar with high elevations and different geological properties are separated from the rest of the basin and divided into two distinct areas, where the Neckar and Fils rivers flow. It demonstrates that simultaneous incidences distinctly respond from the general occurrences of floods.

"My knowledge has reached the point where I know that I do not know."
Avicenna - Ibn-e-Sina

Kurzfassung

Überschwemmungen sind seltene Phänomene, die viele menschliche und finanzielle Kosten verursachen können. In der Regel werden bei der Hochwasseranalyse die Unterschiede in der Gleichzeitigkeit und der Wechselwirkung von Hochwassern in verschiedenen Regionen nicht berücksichtigt. Das unzureichende Wissen für den Umgang mit diesen Extremereignissen kann zu Katastrophen führen. Aufgrund unterschiedlicher Niederschlagsintensität und -typen, unterschiedlicher atmosphärischer Zirkulationsmuster und Jahreszeiten gibt es verschiedene Arten von Überschwemmungen. Auch die Auslösemechanismen von Extremereignissen können unterschiedlich sein, wenn einige Einzelereignisse zur gleichen Zeit an mehreren Orten auftreten. Großflächige Niederschläge in Kombination mit Schneeschmelze in verschiedenen Orten einer Region bringen Abflussspitzen aus verschiedenen Gebieten zusammen und führen so zu extremen Hochwassern. Gleichzeitig auftretende Hochwasser können unvorhersehbare Verluste und Zerstörungen flussabwärts verursachen. Ein anderer Ansatz der Hochwasserhäufigkeitsanalyse ist die Untersuchung des allgemeinen Hochwasservorkommens. Diese Hochwasser können individuell in verschiedenen Teilen eines Einzugsgebietes auftreten. Sie können aber auch teilweise mit gleichzeitigen Hochwassern zusammen fallen. Traditionell konzentrieren sich Forscher mehr auf die univariate Analyse und schenken der multivariaten Analyse weniger Aufmerksamkeit. Die multivariate Extremwertanalyse ist ein Algorithmus der raum-zeitliche Informationen liefern kann. Das Hauptanliegen dieser Forschung ist das Verständnis des Überschwemmungsverhaltens hinsichtlich des gleichzeitigen und allgemeinen Auftretens, das zum Hochwasserrisikomanagement beitragen kann. Außerdem ist es das Ziel dieser Forschung, den Nexus eines Einzugsgebiets, die Prozesse des Flusssystems, geologische Aspekte und ihre Wechselwirkungen zu verstehen, die zu Spitzenhochwassern führen. Das Verständnis des Clustering von extremen Hochwassern kann den Forschern mehr Einblicke in den Hochwasserschutz geben. Die Clusterbildung kann eine Region in verschiedene Gebiete unterteilen, die unterschiedliche Aktionspläne für das Hochwassermanagement und die Risikoanalyse erfordern. Außerdem kann die Versicherungs- und Rückversicherungsbranche neue Koeffizienten zur Abschätzung von Risiko und Schäden für jedes Clustergebiet berechnen. Das mehrdimensionale Hochwasserclustering von zwei Hochwasserereignistypen ist ein weiterer Gesichtspunkt, der in dieser Forschung angewendet wird. Entsprechend der Problemstellung werden Clustermethoden für gleichzeitige und allgemeine Hochwasserereignisse bewertet, um geeignete Methoden für weitere Forschungen vorzuschlagen.

Daher werden in dieser Dissertation die raum-zeitliche multivariate Analyse und die quantitativen Eigenschaften von simultanen und allgemeinen Hochwassern des Neckars und seiner Nebenflüsse betrachtet, um neue Erkenntnisse für den Hochwasserschutz und die Risikoanalyse zu gewinnen. Der Neckar hat einen mittleren Abfluss von 158 m^3/sec und den höchsten aufgezeichneten Tagesniederschlag von 97 mm im Einzugsgebiet.

In dieser Doktorarbeit werden 55 Jahre Tagesabflüsse von 1961 bis 2015 für 46 Messpegel im Neckareinzugsgebiet verwendet. Zunächst wird das gleichzeitige Auftreten von Extremhochwassern untersucht. Dann werden das hierarchische Clustering und die multidimensionale Skalierung durchgeführt, um die Teileinzugsgebiete zu clustern, die ähnlich reagiert haben. Zur Durchführung der Clustering-Methoden werden zwei Dissimilaritätsmetriken und fünf Verknüpfungsmethoden berechnet und angewendet. Dann werden die erhaltenen Cluster ausgewertet, um die bestmögliche Clusterkarte zu bestimmen. Als Ergebnis wird das Einzugsgebiet in drei Clusterbereiche unterteilt. Aufgrund der geophysikalischen Komplexität des oberen Neckars wird dieser Bereich in West und Ost unterteilt. Der Rest des Einzugsgebietes wird ebenfalls als eigener Cluster gruppiert. Im nächsten Kapitel wird die "Principal component analysis" (PCA) auf den Datensatz angewendet, um das primäre Verhalten der Abflusssignale über dem Neckar zu erfassen. Hier wird die PCA-Kreuzvalidierung durchgeführt, um die beste Anzahl von Komponenten für die Zeitreihenrekonstruktion zu finden. Die drei ersten Hauptkomponenten werden ausgewählt, um eine neue Zeitreihe zu rekonstruieren. Dann wird die Differenz zwischen der ursprünglichen und der rekonstruierten Zeitreihenmatrix als Restzeitreihe ermittelt. Danach werden aus der absoluten Residualmatrix unabhängige gleichzeitige Peaks pro Jahr ausgewählt. Anschließend wird als Distanzmatrix die Korrelation zwischen den Paarsätzen ausgewählter Peaks bestimmt, um die Beziehung der gleichzeitigen Ereignisse in verschiedenen Teileinzugsgebieten zu beurteilen. Als nächstes wird hierarchisches Clustering angewendet, um die Multi-Site-Einzugsgebiete in verschiedene Gruppen mit ähnlicher Reaktion zu clustern. Als Ergebnis werden Cluster definiert, um die Region mit unterschiedlichem Verhalten bezüglich eines Hochwasserreaktionsmechanismus darzustellen. Schließlich wird die Region in die drei Hauptcluster unterteilt, von denen zwei zu dem spezifischen geologischen Karst- und Kristallinfelsbecken und der höchsten Erhebung gehören. Die Ergebnisse zeigen ein bestimmtes Muster für das Ausmaß des Hochwasserereignisses, das jeder Gruppe von Teileinzugsgebieten entspricht. Die Ergebnisse der angewandten PCA-AHCT (Agglomerative Hierarchical Cluster Tree) werden mit denen der AHCT-Methoden aus dem vorherigen Kapitel verglichen. Folglich werden die Ähnlichkeiten zwischen den gewonnenen Clustern statistisch ausgewertet. Somit kann dieses Clustering eine grundlegende und innovative Methode in der multivariaten Analyse der Risikoanalyse und der hydrologischen Modellierung unter Berücksichtigung des gleichzeitigen Auftretens von Hochwassern sein.

Die allgemeinen Überschwemmungen werden mithilfe einer verteilungsbasierten Dissimilaritätsmatrix geclustert. Zunächst werden die allgemeinen Hochwasser identifiziert. Dann werden die Kolmogorov-Smirnov-Statistiken aus den empirischen kumulativen Verteilungsfunktionen der allgemeinen Hochwasserreihen berechnet. Schließlich wird ein hierarchisches Clustering durchgeführt, um mögliche Gruppen von Hochwassern in verschiedenen Teileinzugsgebieten zu finden. Die erhaltenen Cluster werden mit drei Clustering-Evaluationskriterien bewertet und validiert. Außerdem wird die optimale Anzahl von Clustern für die Neckareinzugsgebiete ermittelt. Die fünf verschiedenen Verknüpfungsmethoden erhalten unterschiedliche Cluster; bei allen wird jedoch das Einzugsgebietszentrum in einen Cluster eingeordnet. In der Regel werden die Flüsse Jagst und Kocher deutlich vom anderen Teil abgesetzt. Auch die ersten flussaufwärts gelegenen Teileinzugsgebiete sowie ein Teil der Enz befinden sich im gleichen Cluster. Ein Optimierungsschema Clustering ist die ange-

wandte Idee, um die allgemeinen Hochwasser innovativ zu clustern, mit weniger notwendigen Vorgaben für das Clustering, wie z.B. die Anzahl der Cluster. Das Simulated Annealing wird in einer verflochtenen und robusten Weise eingesetzt, um extreme Hochwasserserien zu clustern. Hier spielt der Silhouette-Koeffizient die Rolle der Zielfunktion für die Optimierung. Als Ergebnis wird das Robust Simulated Annealing als neues Werkzeug für mehrdimensionales Clustering eingeführt. Diese Methode muss jedoch noch weiter erforscht und entwickelt werden.

Die angewandte Hochwasserhäufigkeitsanalyse zeigte, dass die mit dem Fluss verbundenen Regionen am Hauptstrom ähnlich reagieren, wenn allgemeine extreme Hochwasser auftreten. Die Topographie und die Geographie hatten jedoch mehr Einfluss auf das gleichzeitige Auftreten von schweren Hochwassern. So ist der obere Neckar mit hohen Erhebungen und unterschiedlichen geologischen Eigenschaften vom Rest des Einzugsgebietes getrennt und in zwei unterschiedliche Bereiche aufgeteilt, in denen Neckar und Fils fließen. Es zeigt sich, dass sich die gleichzeitigen Ereignisse deutlich von den allgemeinen Hochwasserereignissen abheben.

„Mein Wissen hat den Punkt erreicht, an dem ich weiß, dass ich es nicht weiß."

Avicenna - Ibn-e-Sina

1 Introduction

1.1 Motivation

Extreme floods are triggered by large-scale precipitation or a combination of snowmelt and rainfall. These events are considered as one of the principal natural disasters in the world. Approximately one billion people in the world live in floodplains (Baldassarre et al., 2013). In Europe, floods have led to more than 426 billion euros of loss between 1980 – 2017 (European Environment Agency, 2019). Notably, in Germany, two catastrophic floods in 2002 and 2013 have resulted in losses of 26 billion dollars (Schröter et al., 2015). Additionally, in 2016, the losses were estimated to be 4.1 billion dollars for a couple of extreme events (van Oldenborgh et al., 2016; MunichRe, 2019). These massive losses compel us to explore upcoming threats in ways that were not considered in traditional frequency analysis.

The usual flood assessments in the occurrence frequencies have been exerted by a vast range of techniques to perform univariate analyses of extreme events. Nevertheless, univariate statistics cannot discover the flood spatial interactions within catchments, because the flood leading mechanisms in different catchment locations can be different. Therefore, the multivariate extreme value statistics might cover both time and space characteristics of the floods at the same time, although, multivariate analysis of such variables is rarely performed for extreme values (Favre et al., 2004). Traditional multivariate extreme value statistics have constraints on handling the dependence structure (Wyncoll and Gouldby, 2015).

Moreover, the time series of discharge flows have various influential elements, such as linear trends and seasonal and cyclical terms (Ming et al., 2017; Milliner et al., 2018). These series have vivid periodic terms that are probably caused by some geomorphological features and meteorological signals, which should be considered (Schmidt et al., 2008; Moeeni et al., 2017).

This recently developed understanding of floods requires new approaches to investigate. By investigating the spatial characteristics of large-scale floods, the flood protection network can be adjusted by recognizing new spatiotemporal patterns of floods to apply a robust method to notify danger before a disaster. Multivariate analysis of extreme floods is the focus of this study, which can overcome some disadvantages of the univariate analysis. The proposed thesis aims to investigate the flood interactions and the coincidence of flood events to comprehend the relevant processes and causes leading to extreme floods. In addition, this research tries to understand better the spatial aspect of the flood generating mechanism. Moreover, we assess the possibility of improving the current hydrological models to perform better simulation, based on newly proposed methods for identifying spatial clustering of extreme flood occurrences.

1.2 Challenges in flood analysis

Flood risk assessment requires precise statistical analyses of discharge flows for insurance and reinsurance purposes or for the design of appropriate flood protection systems (Ehmele and Kunz, 2019). Besides, flood defenses are critical components to protect against them. Therefore, it is essential to deal accurately with their effects and consider the concurrency of flood occurrences at multiple locations. The risk arises when massive floods coincide in the mainstream and its tributaries. This enhanced risk is characterized by flood magnitude and occurrence date (Chen et al., 2012). Further, the multivariate analysis can be inappropriate when the extreme values of all variables are unlikely to occur together. However, recent developments in multivariate extreme value methods have removed and solved some of these constraints and opened opportunities for improving flood risk analysis methods (Heffernan and Tawn, 2004). Falter et al. (2015) proposed a novel approach innovatively evaluating river flood risk by considering magnitudes of peak discharge, but still, the concurrency of floods was not assessed in their work.

This dissertation mainly focuses on three approaches to the clustering of flood occurrences. The first challenge is how simultaneous occurrences occur; the second one is how do general floods react and interact with each other? i.e., to what extent they happened in the case study's basin? The third one is the spatial extent to which floods relate to each other. Different techniques of clustering are developed to present the most significant algorithms for the clustering of extreme events. In the end, this thesis will answer the difference between these two approaches, and if the general and concurrent floods behave differently in their spatiotemporal dependencies? The three challenges are discussed in the following sections.

1.2.1 Simultaneously occurrences of floods

Floods that simultaneously affect many sites might be considerable challenges to systematize flood disaster management for the indemnification and reinsurance industry (de Moel et al., 2009). So far, much has been investigated in terms of crisis mitigation and risk evaluation (Brázdil et al., 2006; Baldassarre et al., 2013; Alfieri et al., 2014; Haigh et al., 2016; Dewan et al., 2019; Diederen et al., 2019). Nevertheless, they mostly did not take the simultaneous occurrence of floods into account, which plays a significant role in planning and flood risk management. Flood hazards could remarkably increase if floods arrive simultaneously in rivers (Mirza, 2003; McPhillips et al., 2018; Kundzewicz, 2019). The synchronized forms of flooding approach to an accumulation of losses with the risk assessment required to be expanded to a notion representing the spatial risk of flooding (Uhlemann et al., 2010). A similar simultaneous occurrence of peak flows caused devastating damage (Mirza, 2002). Therefore, the depth and extent of floods and associated damage are extensive when the major rivers reach their peaks simultaneously with distinct causative mechanisms (United Nations, 1964; Bertle, 1973). The contemporaneous incidences of extreme floods may have a detrimental impact on society and urban areas more than an individual occurrence (Leonard et al., 2014; Hao et al., 2018; McPhillips et al., 2018). After the destructive flood on the Meuse river in the Netherlands with the characteristic of coinciding in several basins, the government decided to evacuate around two hundred thousand people living in the floodplain areas (Geertsema

et al., 2018). Furthermore, simultaneous occurrences of floods are crucial in designing power plants and hydraulic structures (Newton, 1983). These floods depend on weather circulation patterns (Pedrozo-Acuña et al., 2014). Therefore, assessment of coincident occurrences of extreme floods plays a pivotal role in diminishing possible damages and levels of risk. Consequently, in this regard, two questions come up: To what extent simultaneous behavior of extremes is similar to each other in different catchments? Furthermore, what are the reasons for having probable distinct clusters in the catchment? Accordingly, the thesis will answer the above questions. Also, one of the primary concerns of this research is to understand better the synchronous occurrence of floods in upstream sub-basins that can contribute to flooding risk management by defining areas that may have distinct exposure factors.

1.2.2 General flood behavior

In addition to the synchronization extent of flood events in the previous section, the spatial relationship of general floods between various drainage basins is studied in this part. Here, the challenge is, how do the designated algorithms for clustering in the previous section work for the peaks without considering the concurrent events?. This challenge will focus on general floods. Traditionally, researchers have tried to find and identify the flood types and patterns (Thieken, 2009; Glaser et al., 2010; Hattermann et al., 2012; Hall and Blöschl, 2018). Nevertheless, it is not clear that how do general floods react in different sub-catchments in a basin. However, recently, it became visible that flood patterns have changed in terms of magnitude and timing due to climate change (Blöschl et al., 2017, 2019). Therefore, another challenge in incidences of extreme events examination is to discover probable changing points in time series in different catchment areas. Therefore, there may be different clusters of floods prior-to these possible alteration points.

This part will endeavor to answer the question: Is a pairwise assessment of flood in both general (common) and simultaneous terms sufficient to describe the multidimensional behavior? Or some other innovative methods should be taken into account? In conclusion, we will answer whether the resulting clusters for simultaneous events are the same as the general flood clusters or not and what factors caused these possible differences. Therefore, the main achievement of solving this challenge is understanding to what extent floods react together in time and space.

1.2.3 Clustering algorithms

In addition to the last two challenges, the method of performing clustering is also one of the challenges of this research. Such that, what kind of clustering method is suitable for extreme floods, or what are the differences between the various methods, was formulated during the analysis in the previous two sections. Several different methods have been studied and implemented in this dissertation that have the ability to analyze multidimensional space properly. Finally, a new method is developed for clustering occurrences of extreme events by applying the optimization scheme.

The thesis attempts to answer the following questions. Which regions of the catchments

reacted differently in severe events and what was the reason? How different are the results of employed clustering methods in this study? Furthermore, how near the different resulting clusters are to each other, and how far are they? Is there any alternative way for a multivariate extreme value analysis for flood risk management?

In summary, the thesis's objective is to understand the flood triggering and reaction areas coherently due to the interactions of meteorological variables, catchment properties, and river network processes. Furthermore, the question of how they extend in space and time will be answered. Therefore, clustering simultaneously occurrence and general largest floods in a catchment is proposed using distinct clustering methods in the multidimensional space.

1.3 Scopes and chapters of the thesis

This study aims to address the three challenges discussed in the last section. The thesis is comprised of seven chapters as follows:

Chapter one : Introduction, explains the necessity of investigating the clustering of extreme events in hydrology. First, the simultaneous occurrence of flood events is described and then the general behavior of flood occurrences is demonstrated. Third, the clustering challenges are expressed.

Chapter two : Study area and data, gives a brief assessment of the study domain and the discharge data. The major tributaries of the Neckar catchment are introduced. This chapter illustrates the catchment delineation based on discharge gauges coordinates. Next, it explains some of the physical characteristics of the regions, like streams, distance to the outlet, longitudinal profile of the river, and the slope's steepness. In the following parts, probable trend and the slope of changes in the time series is investigated. Finally, it presents probable breakpoint/s in time series.

Chapter three : Hierarchical cluster analysis of the simultaneous occurrence of floods, focuses on the floods, which happened at the same time and tries to find out the interconnection of the concurrent most enormous floods in space. The simultaneous flood identification distinguishes this part from the general flood behavior. Next, hierarchical clustering is employed to explain to what extent the synchronous floods react in the basin. Also, Multidimensional scaling as a method for dimension reduction is applied here. The last part demonstrates the difference between the two mentioned methods.

Chapter four : PCA-based clustering of the simultaneous occurrence of floods, starts with an overview of principal component techniques, such as time series reconstruction and residual series. Then, the most prominent and common behavior, and terms are captured to work with the residual time series. Next, flood identification is assessed with the contemporary highest values in the residual series. Furthermore, the rank associations of these series are scrutinized and used as an input for hierarchical cluster analysis. Eventually, this chapter shows all the possible cluster maps and compares the distinct methods used in this chapter and the previous one.

Chapter five : Distribution-based clustering of general floods (KS clustering), shifts to explore the general flood behavior. Here, instead of concentrating on the correlation of flood series, a distribution-based analysis is examined. This chapter evaluates the pairwise distance of flood series by utilizing the Kolmogorov-Smirnov test. Further, the similarity of flood distribution based on the previous step is calculated. Finally, the chapter maps the conclusive clusters.

Chapter six : Simulated Annealing clustering of general floods, introduces an innovative clustering algorithm, employing optimization schemes into machine learning and cluster analysis. This chapter presents two intertwined Simulated Annealing optimization schemes as a robust technique to find the most similar flood reaction areas.

Chapter seven : Conclusion, presents a summary of the thesis to answer research questions.

The flow diagram of this thesis is shown in Figure 1.1.

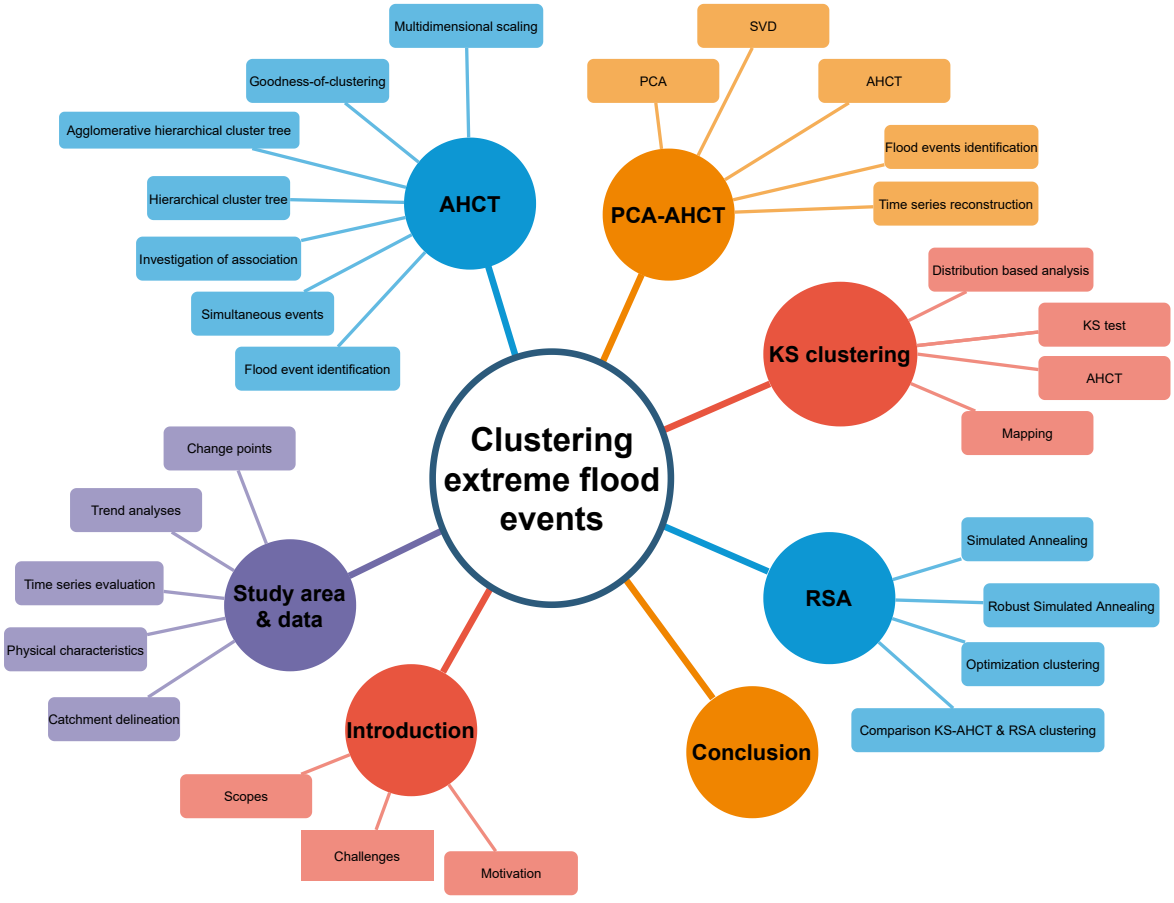


Figure 1.1: The flow diagram of the thesis

2 Study area and data

In hydrology, understanding the catchment area and its properties as well as evaluating and pre-processing data play a key role. In this thesis, the study area is the Neckar catchment and the data is daily discharge flow measurements across the basin. Besides, the precipitation time series is employed briefly to check the identification of simultaneous occurrences of floods. In the continuation of the chapter, initially, the catchment is delineated. Then some of the physical features of the catchment are investigated. Finally, a trend test, the slope of changes and breakpoints of discharge time series are analyzed and detected. This information may show a significant relationship between the possible clusters and regions with similar trends or slope of changes.

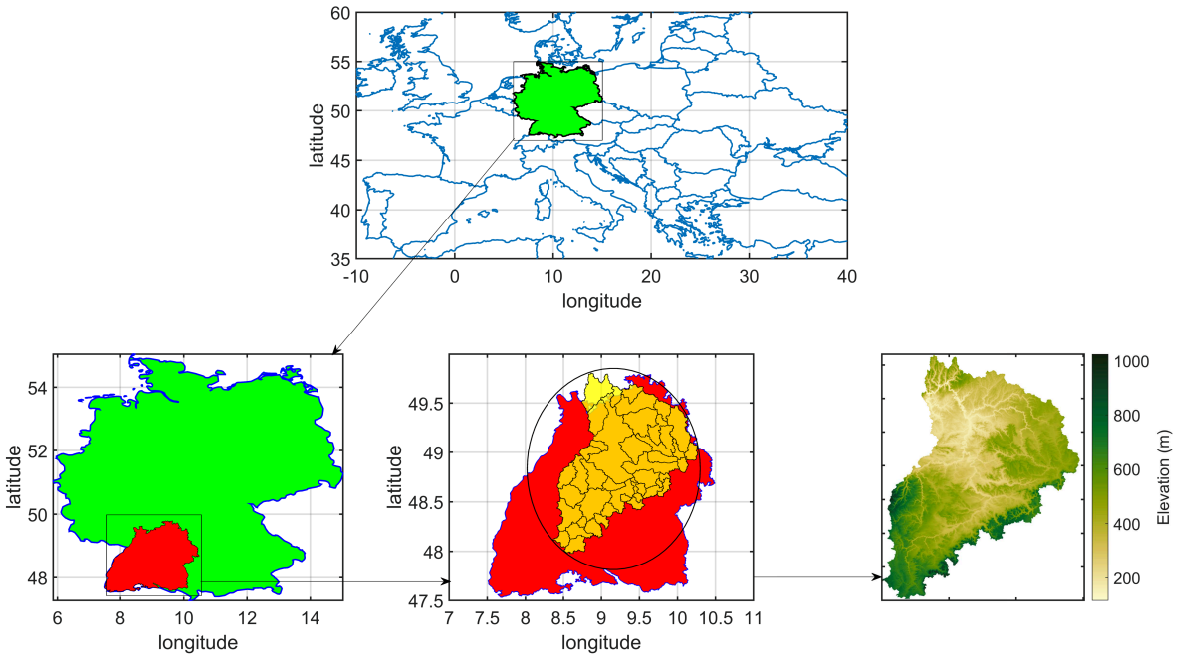


Figure 2.1: Coordinate of the Neckar catchment in the southwest of Germany

The Neckar is a tributary of the Rhine river with a length of 367 km and is located in the southwest of Germany. The catchment area is approximately 14000 square kilometers, covering around 40% of Baden-Württemberg (BW) state's area. The study region is split into 46 sub-catchments with various drainage sizes between 60 to 14000 km^2. The average flow at the outlet of the Neckar river near Heidelberg is 158 m^3/sec. According to the Digital Elevation Model (DEM) map of the catchment in Figure 2.1, the highest elevation point

is located in the south and southwest of the area (Greene et al., 2019). The elevation in the study area varies from 97 to 851 m. The most important tributaries of the Neckar, respectively, are Enz, Kocher and Jagst rivers. Afterward, Nagold, Rems, and Fils are other vital branches of the river (Table 2.1).

Table 2.1: Major tributaries of the Neckar River

Rivers	**Catchment area (km^2)**	**Length of record**
Enz	2223	1961-2015
Kocher	1989	1961-2015
Jagst	1837	1961-2015
Nagold	1144	1961-2015
Fils	699	1961-2015
Rems	583	1961-2015

The local topography mainly adjusts the precipitation in the region. The highest mean annual precipitation of 2000 mm was recorded in the northern Black Forest, which is in the southwest of the Neckar (Emblemsvåg, 2012; Seidel et al., 2012). By moving to the east of the region, the mean annual precipitation is reduced to around 800-1000 mm (Bürger et al., 2006). Besides, the most extensive daily precipitation recorded among selected stations from 1961 to 2015 is 97 mm.

The geological unite of the Neckar basin is characterized by a gentle dipping of the formations towards the southeast. Nine major geological units were recognized, of which six were categorized as freshwater aquifers. Furthermore, the impermeable crystalline rock was specified as the bottom of the aquifer system (Götzinger et al., 2008). Diverse geological formations form the uppermost aquifer in distinct catchment regions (Jagelke and Barthel, 2005). Consequently, the existence of the dipping layers may lead to a complex hydrogeological condition.

In addition, some anthropogenic changes have affected the river and subsequent discharge flow in the independent tributaries of the study area. For example, Kalweit et al. (1993); Bormann (2010) mentioned that around 60 constructed reservoirs enhanced the catchment's complexity.

2.1 Catchment delineation

Regarding coordinates of measurement gauges in the Neckar basin, sub-catchments are delineated (Schwanghart and Kuhn, 2010; Schwanghart and Scherler, 2014). All the stations and their river tributaries in the catchment are shown in Table 2.2.

Table 2.2: Discharge measurement gauges in the Neckar catchment

Nr.	ID	Area(km^2)	Name	River	Nr.	ID	Area(km^2)	Name	River
1	406	$4.12*10^2$	Rottweil	Neckar	24	2452*	$3.75*10^2$	Nagold	Nagold
2	409	$6.53*10^2$	Oberndorf	Neckar	25	2471**	$7.50*10$	Gruol	Stunzach
3	411	$1.11*10^3$	Horb	Neckar	26	2477**	$1.77*10^2$	Oberensingen	Aich
4	420	$2.32*10^3$	Kirchentellinsfurt	Neckar	27	2489*	$1.49*10^2$	Tübingen	Steinlach
5	422	$1.77*10^2$	Riederich	Erms	28	3421	$1.65*10^3$	Vaihingen	Enz
6	427	$3.96*10^3$	Plochingen	Neckar	29	3465	$1.94*10^3$	Stein	Kocher
7	434**	$5.04*10^2$	Murr	Murr	30	3470	$1.82*10^3$	Untergriesheim	Jagst
8	454	$1.26*10^4$	Rockenau	Neckar	31	3498	$1.29*10^3$	Kocherstetten	Kocher
9	460*	$4.56*10^2$	Meckesheim	Elsenz	32	4408*	$2.10*10^2$	Owingen	Eyach
10	463*	$6.64*10$	Süßen	Lauter	33	4416⊗	$1.37*10^4$	Heidelberg	Neckar
11	473**	$2.46*10^2$	Abtsgmünd	Lein	34	4421*	$1.56*10^2$	Mosbach	Elz
12	477	$1.02*10^3$	Dörzbach	Jagst	35	4422	$1.06*10^3$	Pforzheim	Enz
13	478*	$2.13*10^2$	Sennfeld	Seckach	36	4427	$7.02*10^2$	Plochingen	Fils
14	1411*	$1.65*10^2$	Schwabsberg	Jagst	37	4428	$7.41*10^2$	Gaildorf	Kocher
15	1412	$8.08*10^2$	Elpershofen	Jagst	38	36056**	$4.16*10^2$	Pforzheim	Würm
16	1439*	$1.45*10^2$	Geislingen	Fils	39	40670*	$1.26*10^2$	Rangendingen	Starzel
17	1452*	$5.83*10$	Unterlenningen	Lauter	40	44603*	$1.16*10^2$	Bad Urach	Erms
18	1458*	$8.82*10$	Kirchheim	Lindach	41	46349	$4.79*10^2$	Wöllstein	Kocher
19	1462	$2.87*10^3$	Wendlingene	Neckar	42	46358*	$1.12*10^2$	Hüttlingen	Kocher
20	1470*	$5.69*10^2$	Neustadt	Rems	43	76121*	$1.15*10^2$	Geislingen	Eyb
21	2406*	$1.23*10^2$	Sachsenheim	Metter	44	76123*	$2.82*10^2$	Neuenbürg	Enz
22	2431	$1.91*10^2$	Wendlingen	Lauter	45	76178**	$2.77*10^2$	Geislingen	Bühler
23	2446*	$1.67*10^2$	Horgen	Eschach	46	76179	$3.86*10$	Salach	Fils

⊗: The outlet of the Neckar catchments in the Ziegelhausen in Heidelberg.

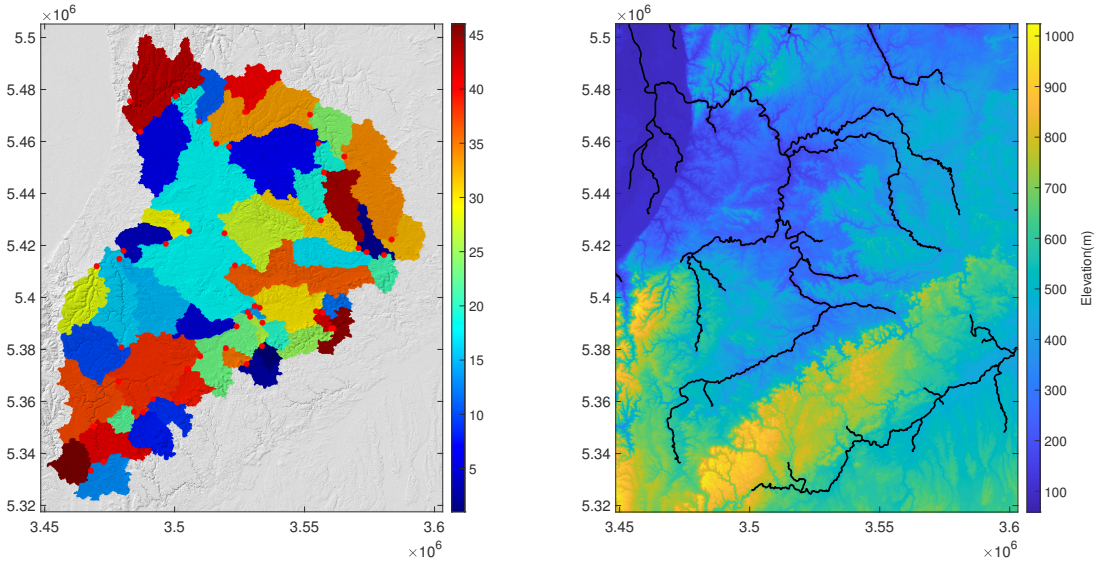
*: The headwater subcatchment.

**: The headwater subcatchment not on the outer border of the catchment.

The subcatchment area in this thesis is calculated based on all areas behind the selected gauge. The subcatchments areas range between almost 39 km^2 in Salach on Fils river (station Nr.46 (76179)) and 13722 km^2 in Heidelberg on the Neckar river. Station Nr.33 (4416) in Heidelberg is the outlet of the study area. The headwater subcatchments like station Nr.14 in Schwabsberg on the Jagst River are the stations where there is no measurement gauge before that. Usually, these gauges are located on the border of a catchment. However, due to the topography of the Neckar catchment, some headwater stations are in the middle of the case study (such as station Nr.38 in Pforzheim on the Würm River. Therefore, it is possible to have distinct behavior in the headwater subcatchment areas.

The result of the catchment delineation based on 46 stations is illustrated in Figure 2.2, where each station is the outlet of its sub-catchment.

In the left panel of Figure 2.2a, each sub-catchments area is mapped with distinct colors and the red dots are shown the location of outlets for each of them. Figure 2.2b is illustrated the



(a) Sub-catchments in the Neckar basin

(b) Biggest streams in the Baden Württemberg

Figure 2.2: The Neckar catchment delineation, based on measurement stations

main streams of the Neckar that are mainly in the BW state. In the middle of the map, the major rivers of the Neckar are clearly shown that lead to the Rhein river in the Northwest (top left) corner of the map. Yellow color parts are the highest elevation in the region of the Swabian Jura and the Black forest. Also, it is possible to point to the Danube River source in the south and southeast on this map.

2.2 Physical characteristics of the catchment

2.2.1 Main river

The major streams in the study area are demonstrated in the previous map (see fig. 2.2). Here, it is essential to recognize the main river and its tributaries in the catchment area. Figure 2.3 shows the mainstream and other adjunct rivers.

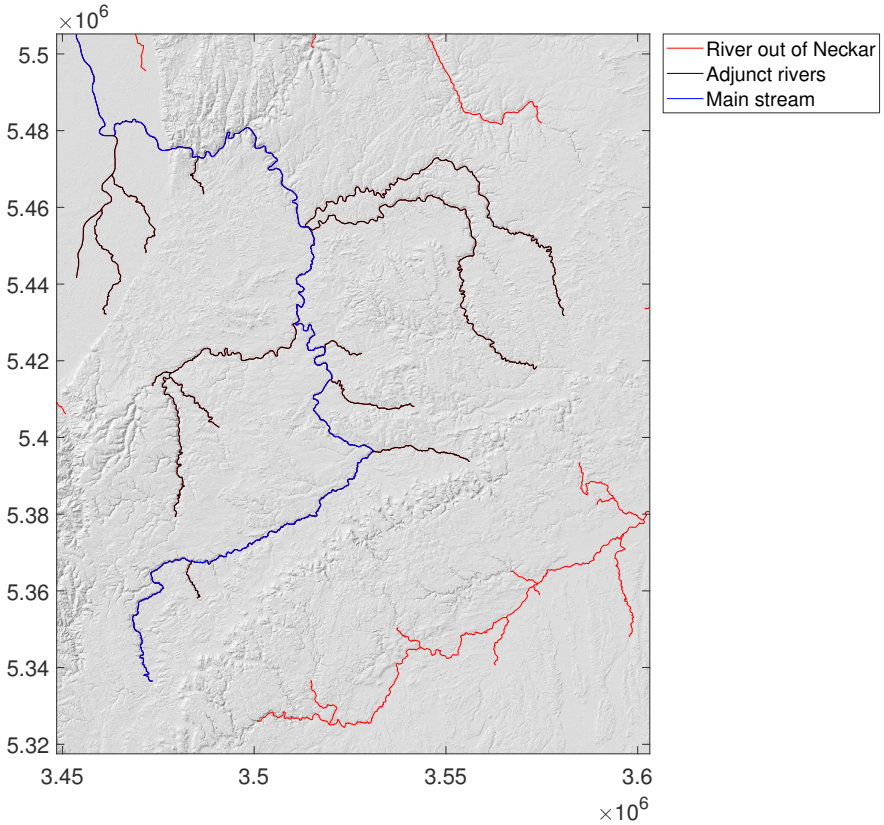
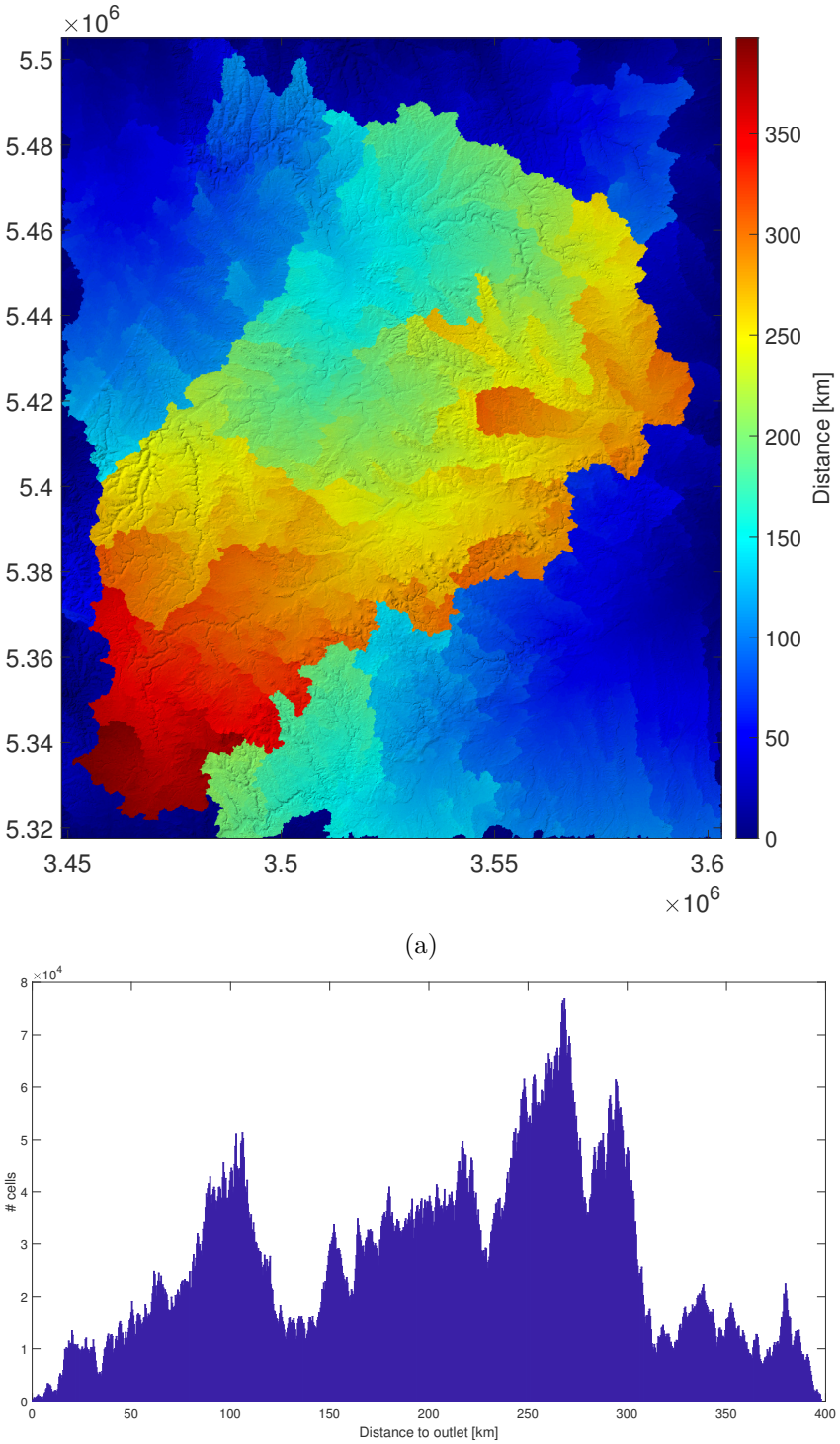


Figure 2.3: Mainstream in the Neckar

The blue line shows the mainstream of the catchment, which all other tributaries lead the flow in the Neckar catchment. The dark color streams are taken other paths and then are joined the main river. This indicates the different topographic sections and divisions in this area.

The distances from each drainage sub-basin along the flow network are calculated to gain more insights into the physical characteristics of the Neckar. From two different perspectives, the distance of catchments from the outlet of the Neckar has been studied. The first is the top view and the second one is the cell by cell longitudinal profile. At the end of the thesis, it will determine any relationship between resulted clusters and distances to the outlet.



(a)

(b)

Figure 2.4: Cellwise distance to the outlet in the Neckar catchment, a) Distance from each drainage basin along with the flow network, b) Density function of cell distances to the outlet

In contrast to the Euclidean distance, the hydrological distance presented in Figure 2.4 shows distinct behavior. In some parts of the catchment, such as Northeast and partly west, the hydrological distance is longer than the Euclidean distance (Farther distances areas are shown in red). Also, in part *b* of Figure 2.4, the most frequent distance to the outlet is around 250 to 300 *km*. Then, at approximately 100 *km* from the catchment outlet, the most comprehensive section is visible.

2.2.2 Longitudinal profile

The mainstream of the Neckar starts 360-370 *km* from the outlet in Heidelberg. The highest elevation point of the main river is 550 *m* and the lowest elevation is 97 *m* above sea level. In the lower elevation, many fluctuations are visible on the map between 40 to 120 *km* from the outlet (see fig. 2.5).

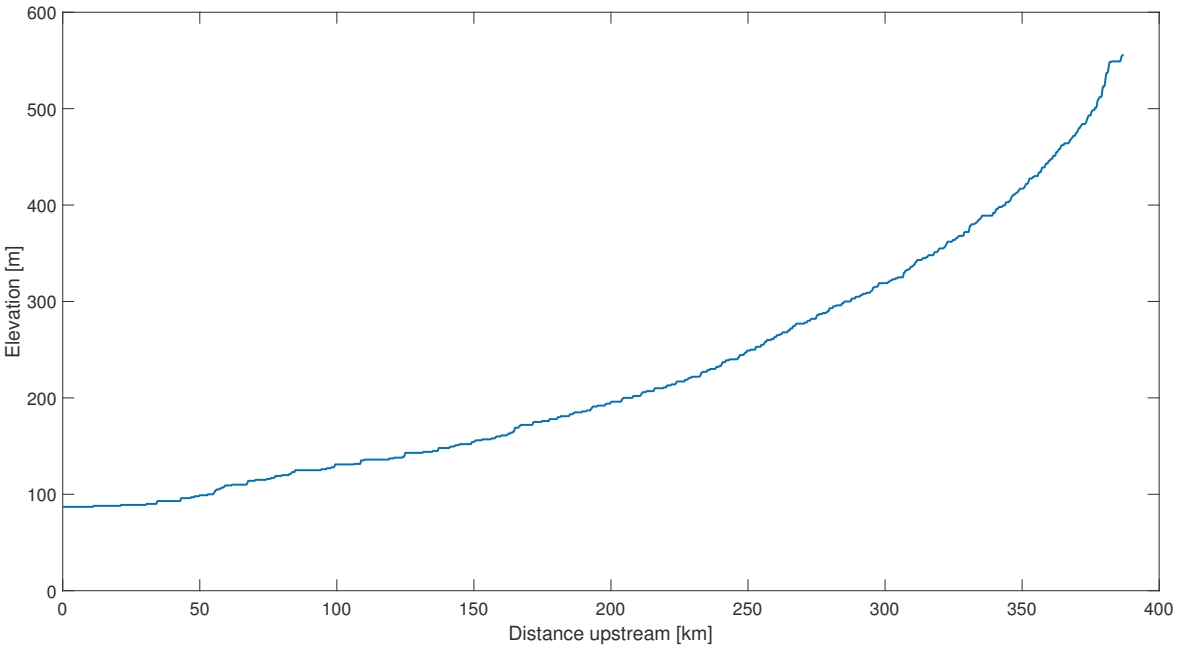
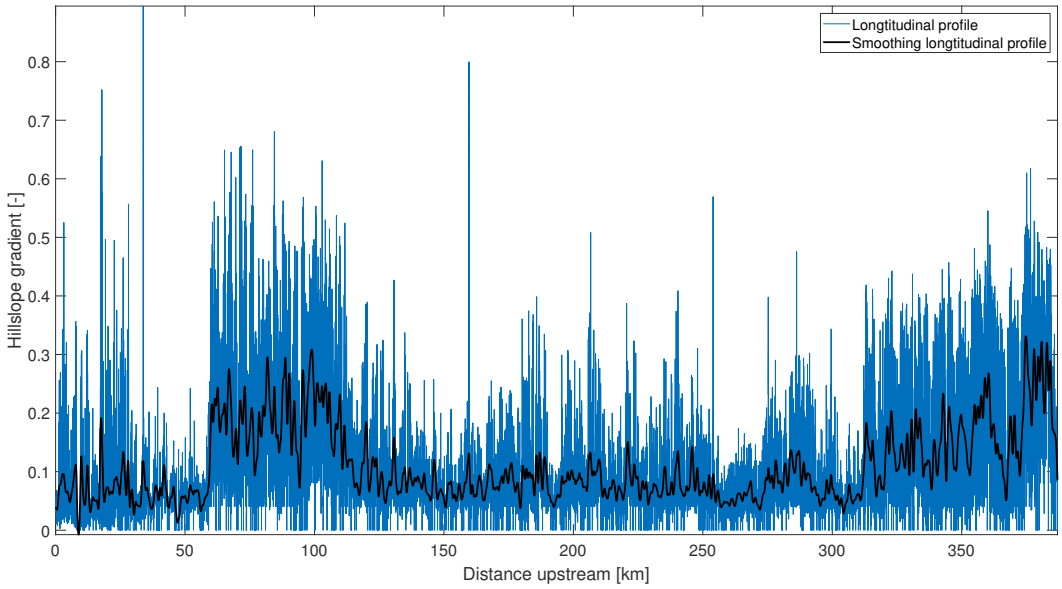


Figure 2.5: The longitudinal profile of the mainstream

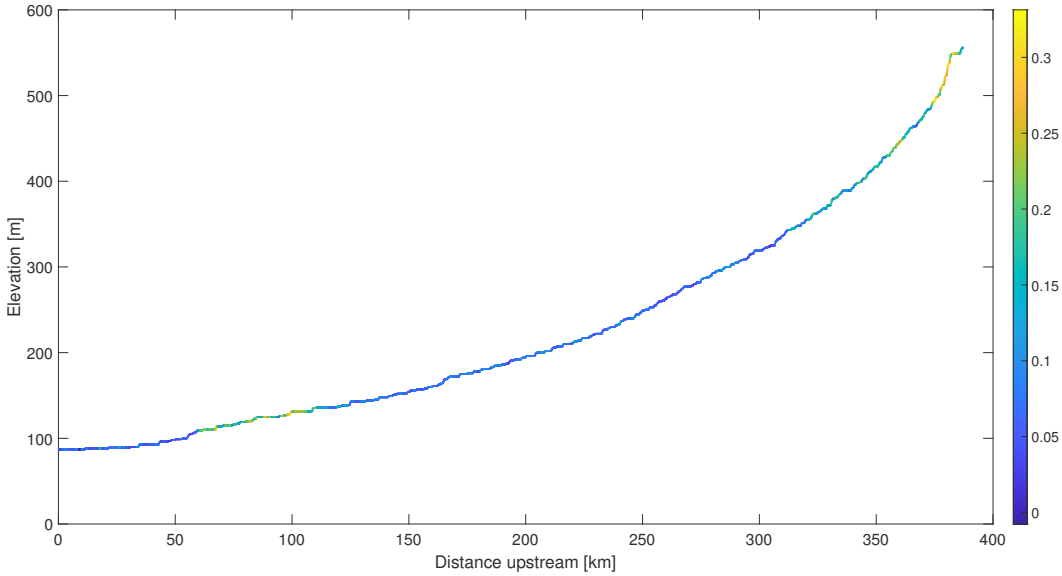
2.2.3 Steepness of slopes

In order to understand better the longitudinal profile of the catchment, the steepness of slopes is calculated (Schwanghart and Scherler, 2017) and the results in three parts are presented in Figure 2.6.

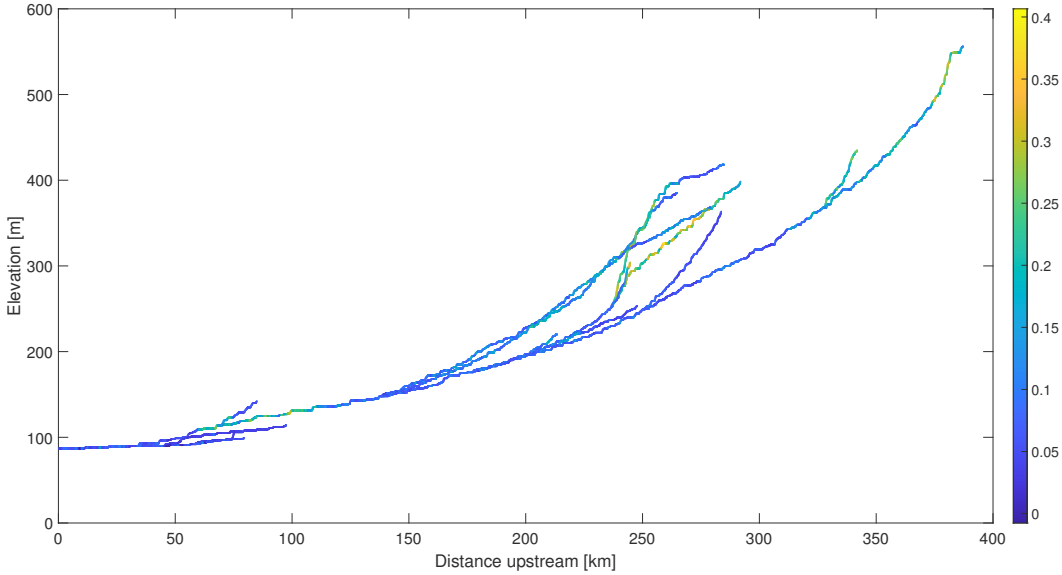
In Figures 2.6a and 2.6b, the hillslope gradient on the mainstream is plotted. These figures show more visible information in comparison to Figure 2.5. Besides, the most considerable gradient is obtained for two sections of 60 to 110 *km* and after 350 *km* to the outlet. Furthermore, the longitudinal profiles of the other branches are also shown in Figure 2.6c. Compared to the main river route, the highest gradient exists at an altitude of 250 to 400 meters. Also, this figure shows that after joining most of the river tributaries to the main river at a distance of 140 *km* from the outlet, the gradient of the slope changed more sharply.



(a) The steepness of slopes adjacent to the river network



(b) The hillslope gradient of the mainstream

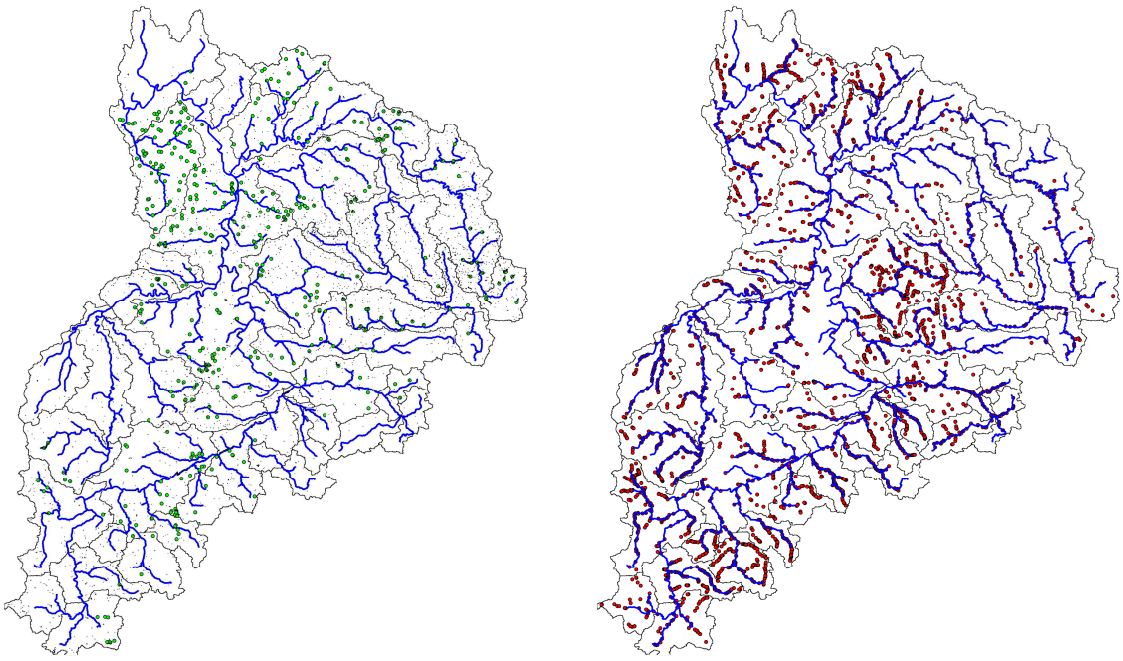


(c) The hillslope gradient of the network streams

Figure 2.6: The steepness of slopes over the Neckar catchment

2.3 River regulations

River regulation controls the variability of stream flows or river water levels to fulfill human demands for different purposes. Water supplies, irrigation agriculture, hydroelectric power generation, navigation, and flood control, and land drainage are examples of river regulation (Petts, 1999). It is an essential tool for socio-economic development, which is controlled by the coordinated regulation of flows throughout entire river basins (Nikitina et al., 2010). River regulation can be made in four main ways. First, flow regulation is achieved by building large dams, often in the headwaters of rivers or in canyons downstream. This form of regulation aims to reduce floods, increase flows in the river during the dry season, or save water from one year to the next. Second, it is possible to regulate rivers by building a chain of major dams or hydraulic structures. Third, constructing a series of run-of-river impoundments to maintain water levels, managed by navigation weirs and locks. Usually, they include low-head hydroelectric power plants. Fourth, building a channel to regulate rivers or lead streamflow to small lakes to protect the environment and to do sport (Petts, 1999).



(a) Dams, reservoirs in the Neckar basin (b) Hydraulic structures in the Neckar basin

Figure 2.7: The Neckar catchment delineation, based on measurement stations

In Figure 2.7a, the green dots are dams (Stauanlage), including the built flood protection structures (Hochwasserschutz), fish ponds (Fischteich), bathing lakes (Badesee), energy generation structure (energiegewinnung), environmental protection structure (Naturschutz), screes (Schlamm) and sport water structures. The tiny black shapes are the small water bodies such as reservoirs and lakes in the Neckar catchment. In total, there are 4111 water bodies and 423 dams in the Neckar basin before the outlet near Heidelberg (LUBW, 2021). In Figure 2.7b, the red dots are all hydraulic structures in three categories of the regulatory, river bed, and hydropower structures which equal to 4406 constructions. Among these constructed

buildings, 94% of structures are currently active (LUBW, 2021). Figure 2.7 shows that the Neckar is a highly regulated river.

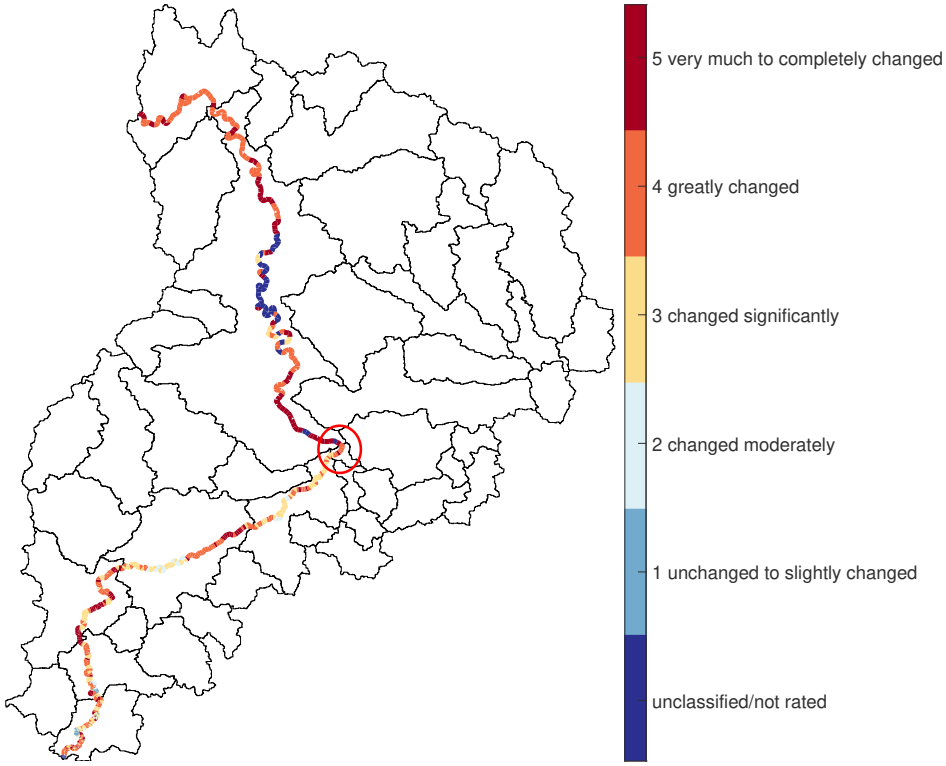


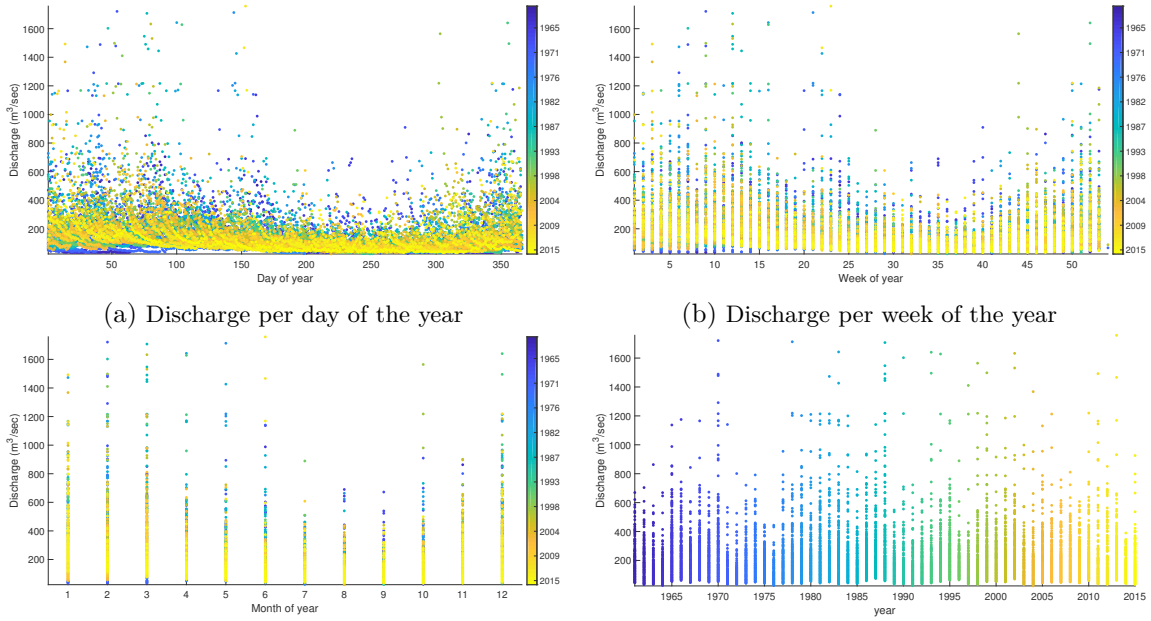
Figure 2.8: The main river channel structural changes

The Neckar River originates from the southwest and flows to the south of the study area. Then the streamflow completely changes direction and continues to the north. The Plochingen station (Nr. 6) is the place of flow diversion and it is the outlet of the upper Neckar (shown as red circle). Almost, every part of the Neckar river after this point was greatly to completely changed, which is shown in orange and red color in Figure 2.8. Before this point, usually, the revise was slightly to significantly changed (LUBW, 2021). Also, a part of the map in dark blue color is not rated yet.

2.4 Data and time series evaluation

The discharge flow measurement data of the Neckar basin is considered for this thesis, which is kindly provided by Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg (LUBW). Also, precipitation data issued by Deutscher Wetterdienst (DWD) is used to verify individual events. The daily discharge time series from 1961 to 2015 is studied for 46 gauges in different parts of the basin. Also, the Copula-based method is employed for infilling missing values in time series (Bárdossy and Pegram, 2014).

In order to accurately evaluate the input data, the data is first divided into different time intervals: day of the year, week of the year, month of the year and yearly distribution. Then, the discharge flows per different temporal resolutions are displayed in Figure 2.9. Figure



(a) Discharge per day of the year

(b) Discharge per week of the year

(c) Discharge per month of the year

(d) Discharge per year

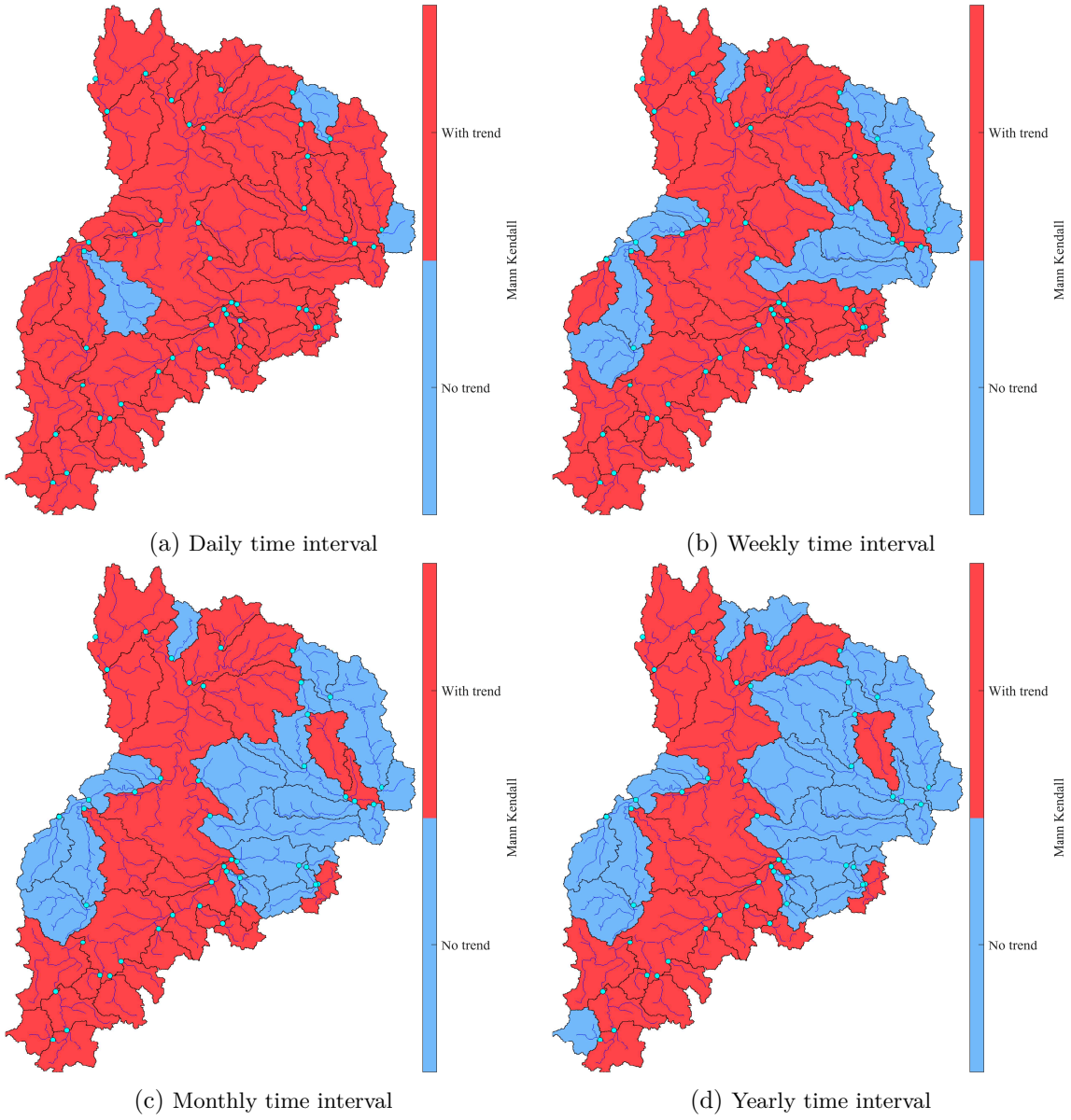
Figure 2.9: Discharge flow in different period of time intervals

2.9b shows a more clear difference among decades. Measured discharges after 2004 have shown less volatility. Most of the time, their maximum measured volumes are less than the maximum discharge volumes in previous years, except for the first five weeks of the year. On the monthly scale, the highest river discharge volumes have a downward deviation, except for January. However, in the first three months of the year, the minimum volumes of river water in the early years are much lower than the average of the lowest discharge measured in other years (see fig. 2.9c). The scatter points in different years are the occurred floods in a year at the top of the plot. Each year is plotted with different colors to see the extent of scattering data in different years as the symbol of extreme events (see fig. 2.9d). In total, the results express that the range of discharge series has been changed and it is necessary to calculate the amount of likely changes per year.

2.4.1 Trend analyses

A non-parametric trend test using the *Mann-Kendall Tau-b* technique is performed for all designated temporal resolutions to investigate the stationarity of the discharge time series (Mann, 1945; Kendall, 1948). Putting the trend significance of a time series on a map shows the trend pattern in space. Then a probable trend line is plotted by employing ordinary least-squares and *Sen*'s slope to see the possible orientation of changes. In general, trends usually meaningful for the independent data. Here, the independent data could be the extremes. Therefore, it is recommended to apply the phase randomisation method on the dependent data to catch the possible trend/s (Radziejewski et al., 2000).

The results of Figure 2.10 show that changes in time intervals significantly alter the result of the Mann Kendall test so that by moving from daily to yearly resolution, fewer stations



(a) Daily time interval

(b) Weekly time interval

(c) Monthly time interval

(d) Yearly time interval

Figure 2.10: Mann Kendall trend analysis in different time intervals for the discharge time series in the Neckar at the significance level of 0.05

are reported to have a significant trend. In Figure 2.10a, only three small sub-catchments in different parts of the region have no trend over time. The highlighted point in the four parts of Figure 2.10 is that all the sub-basins along the main river, always have a significant trend at $\alpha = 0.05$ level in different time intervals.

2.4.2 Theil-Sen's slope

The *Theil–Sen*'s slope estimator, or the *Kendall* robust line-fit method, also known as *Sen*'s slope, is a robust linear regression method. It chooses the median slope among all lines through pairs of sample points (Sen, 1968; Theil, 1992). *Theil–Sen*'s slope can be calculated

efficiently and is insensitive to outliers. It can be significantly more accurate than simple linear regression for skewed and heteroskedastic data and can be estimated besides simple least-squares even for normally distributed data (Peng et al., 2008; Danziger, 2020).

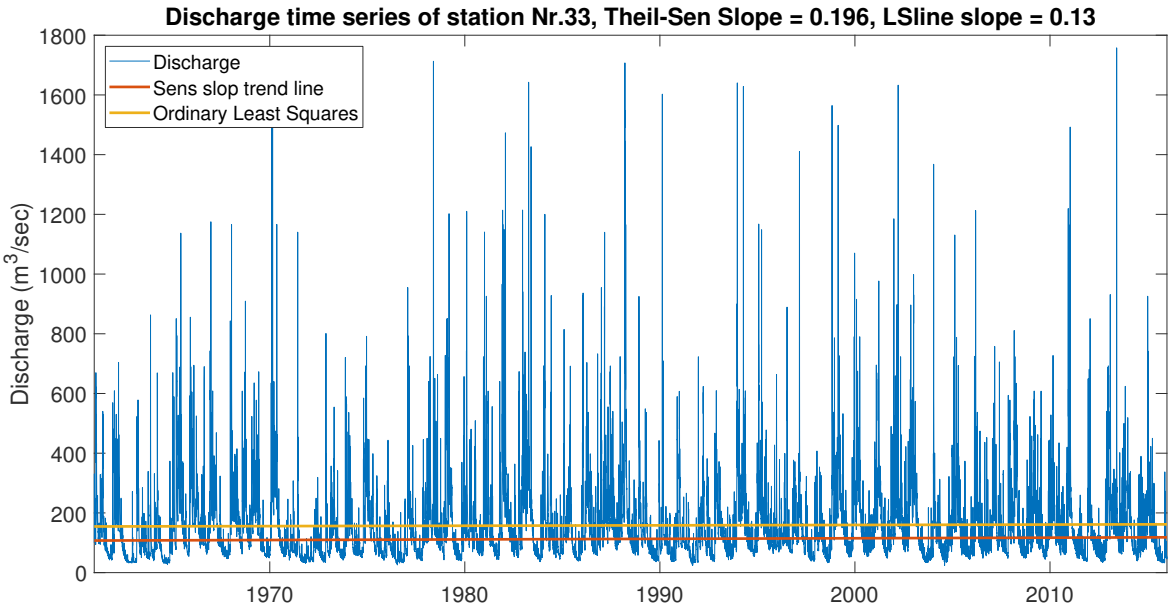


Figure 2.11: The slope of changes in the discharge time series

Figure 2.11 shows a positive slope for both methods. The least-squares linear regression presents 0.13 m^3/sec per year increase of discharge and the *Theil-Sen*'s slope shows a higher ascent with around 0.20 m^3/sec per year. This slope of the line indicates that over time, the river flow has consistently increased despite seasonal fluctuations.

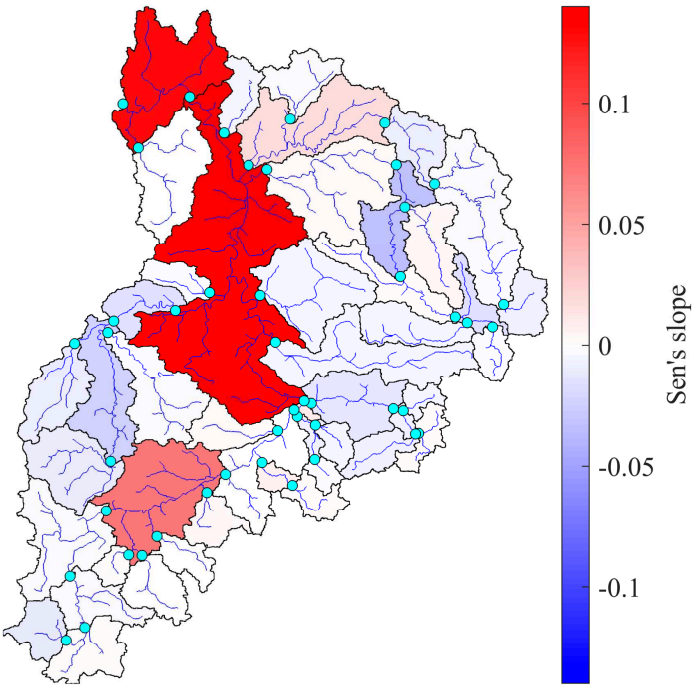


Figure 2.12: The Sen's slope of changes in the discharge time series

Different sub-catchments in the Neckar show distinct slopes of changes in discharge time series. The sub-catchments near the outlet have a solid positive slope of changes; however, most areas have no slope or small negative slopes (see fig. 2.12). The reason for this difference can be flood protection structures built after the Plochingen as the outlet of the upper Neckar.

2.4.3 Change points detection

According to previous steps and observing the various slopes of change in the discharge time series, it is necessary to determine any possible breakpoint/s at different time intervals. Also, Blöschl et al. (2017) detected noticeable patterns of flood timing shift over European countries for the last fifty years. The following steps are taken to detect probable change points in a signal, i.e., time series (Lavielle, 2005; Killick et al., 2012):

1. A spot is chosen and the signal is split into two sections.
2. An empirical estimate of the least squared line for each section is computed.
3. At each point within a section, the property deviates from the empirical estimate is determined and the deviations for all points are added.
4. The total residual error using section by section deviation addition is found.
5. The location of the division point is changed until the total residual error is achieved a minimum.

The target is to minimize the Sum of Squared Errors (SSE) of linear regression and observation for a given signal $x_m, x_{m+1}, \ldots, x_n$. The algorithm employs as total deviation the sum

of squared differences between the signal values and the Least-Squares linear fit (LS line) via the values, which is shown as follows (Killick et al., 2012):

$$\hat{x}(t) = \frac{S_{xt}|_m^n}{S_{tt}|_m^n}(t - \mu([t_m \quad \dots t_n])) + \mu([x_m \quad \dots \quad x_n]) \tag{2.1}$$

and the SSE is equal to:

$$\begin{aligned}\sum_{i=m}^{n} \Delta(x_i; \chi([x_m, \cdots x_n])) &= \sum_{i=m}^{n} (x_i - \widehat{x}(t_i))^2 \\ &= S_{xx}|_m^n - \frac{S_{xt}^2|_m^n}{S_{tt}|_m^n} \\ &= (n-m+1)\sigma^2([x_m \cdots x_n]) \\ &\quad - \frac{(\sum_{i=m}^{n}(x_i - \mu([x_m \cdots x_n]))(i - \mu([m \quad m+1 \ \cdots \ n])))^2}{(n-m+1)\sigma^2([m \quad m+1 \ \cdots \ n])},\end{aligned} \tag{2.2}$$

where $S_{xy}|_m^n$, μ, σ^2, χ and Δ are the sum of squares of two sections x and y, mean, variance and a given the section empirical estimate and the deviation measurement, respectively. If there are K changing points in the series, then equation 2.3 should minimize as follows:

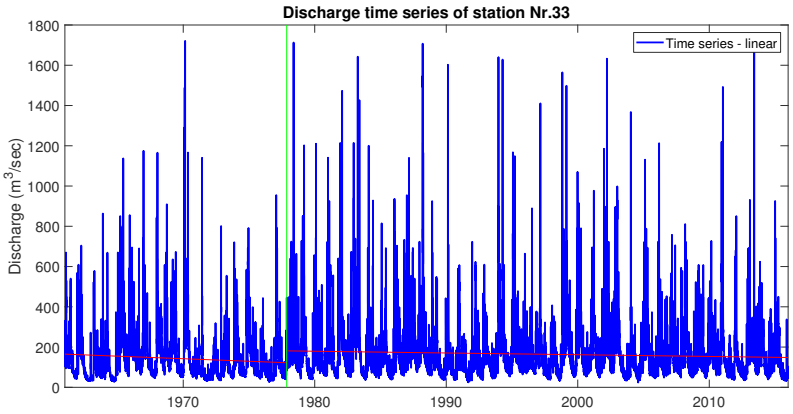
$$J(K) = \sum_{r=0}^{K-1} \sum_{i=k_r}^{k_{r+1}-1} \Delta(x_i; \chi([x_{k_r}, \cdots x_{k_{r+1}-1}])) + \beta K, \tag{2.3}$$

where k_0 and k_K are the first and the last sample of the signal, respectively, the β specifies the minimum improvement in total residual error for each changing point as a minimum threshold.

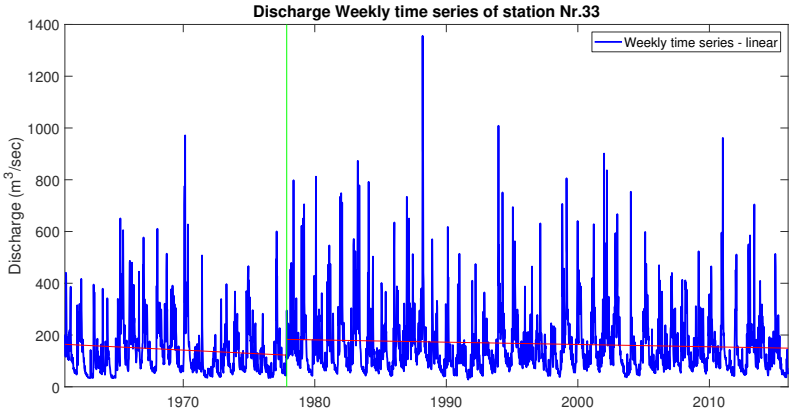
Figure 2.13 illustrates a linear regression line into different sections of the discharge time series of the station Nr. 33 in Heidelberg.

A breakpoint appears in all time intervals in the year 1989. Also, both separated sections have shown a downward tendency; however, in Figure 2.11, a positive slope is reported. As it is clear, especially in Figures 2.13c and 2.13d, that the slopes of the red lines are steeper in the first part between 1961 to 1989 than in years between 1989 to 2015. This difference and breaking point may also have occurred in a series of floods. As a result, it can be considered and clustered separately for these two time periods. Moreover, there is the possibility to detect the potential change points using phase randomization in the dependence data set (Radziejewski et al., 2000).

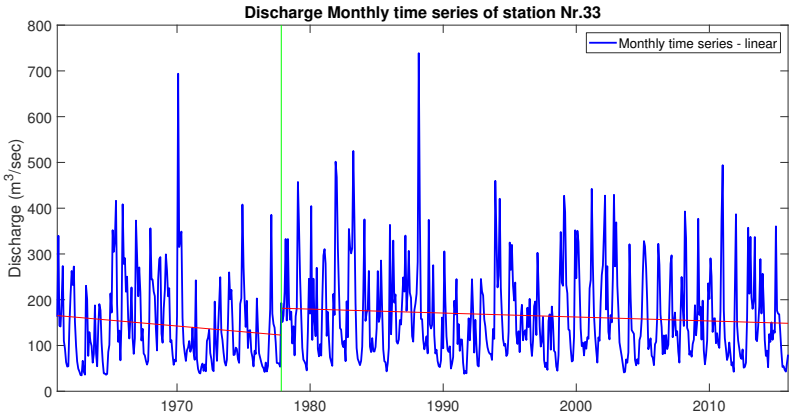
This section gives us better information and perspective on the status of the data and the characteristics of the catchment area, which will lead to the interpretation of the clustering maps in the following chapters.



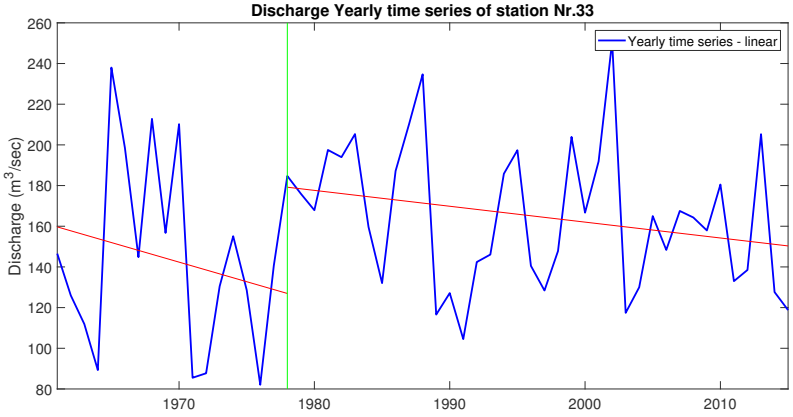
(a) Daily time series



(b) Weekly time series



(c) Monthly time series



(d) Yearly time series

Figure 2.13: The detection of change point/s in the outlet of the Neckar catchment

3 Hierarchical clustering of the simultaneous occurrence of floods

3.1 Introduction

Floods are one the most destructive natural disasters which have to be controlled and monitored to protect society. These extreme events brought hundreds of billion economic losses only in Europe in the last decades. However, the flood protections and consequential damages are various in different countries. Particularly, in Germany, a couple of catastrophic floods happened in 2002, 2013 and 2016, which caused more than 30 billion dollars damages (Schröter et al., 2015; van Oldenborgh et al., 2016; MunichRe, 2019).

Therefore, flood risk assessment and flood defenses require precise analyses to design appropriate flood management systems. After these massive floods in the Elbe and Danube catchments, the German water and planning regulations were subjected to several amendmentations (Petrow et al., 2006). The socio-economic losses of the floods proved that it is substantial to accurately deal with and consider the concurrency of flood occurrences. Floods that simultaneously affect many sites might be a problem for systematic flood disaster management and the indemnification and reinsurance industry. So far, they have not taken the simultaneous occurrences of extremes into account, which play a significant role for planers and water resources executives (de Moel et al., 2009). Thus, a tremendous challenge still remains for quantifying the potential and actual flood damages.

This chapter focuses on the spatiotemporal multivariate analysis of extreme floods, which can overcome some disadvantages of the univariate analysis. The univariate flood frequency analysis is broadly used in hydrological studies. Nevertheless, univariate statistics cannot discover the flood spatial interactions within catchments. The peaks, primary behavior, and dynamics of floods are required as the main properties for every veritable analysis. Furthermore, in the spatiotemporal space, these elements can be taken into account. Comprehensive multivariate analysis is rarely performed in this issue because the minimal available number of multivariate models are not well suited to represent extreme values (Favre et al., 2004). The limited approach is inappropriate when the extreme values of all variables are unlikely to occur together or when the interest supports the joint distribution, where only some subsets of components are extremes.

The proposed research aims to investigate the coincident occurrence of flood events. Moreover, we assess the possibility of improving the current hydrological models to perform better on the peak values based on newly proposed clusters of extremes. Also, one of the primary concerns of this research is to understand better the synchronous occurrence of floods

in upstream sub-basins that can contribute to flooding risk management. Due to different spatiotemporal precipitation distributions and the resulting outflows in these sub-basins.

3.2 Methodology

3.2.1 Flood events identification

In this research, the two biggest events per year for 55 years are selected for each location. These two events have to be independent of each other ($\vee_{i,1} \perp\!\!\!\perp \vee_{i,2}$). On the other hand, the second biggest peak should not be in ten-day intervals of the highest discharge in a year. Furthermore, each peak should come from a single and separate rainfall event (not the accumulation of complementary precipitation events in a period of $\pm$ 10 days).

$$t_{\vee_{i,1}} - 10 < t_{\vee_{i,2}} < t_{\vee_{i,1}} + 10 \tag{3.1}$$

where $\vee_{i,1}$ and $\vee_{i,2}$ are the two biggest independent maximums in the ith year and $t_{\vee_{i,1}}$ and $t_{\vee_{i,2}}$ are the day indices of their occurrences in that year.

3.2.2 Simultaneous events and their corresponding peaks

The flood event identification presented in the last step illustrates actual independent extreme floods in the time series. Some floods arose at different times and some at the same time window. Two days before and after each incident has been selected to find the simultaneous occurrences. If there is a peak at one station and at the other it is not, the corresponding peak in the time interval has been found. This procedure does not neglect any of these highest values and keeps all possible information about the extreme event.

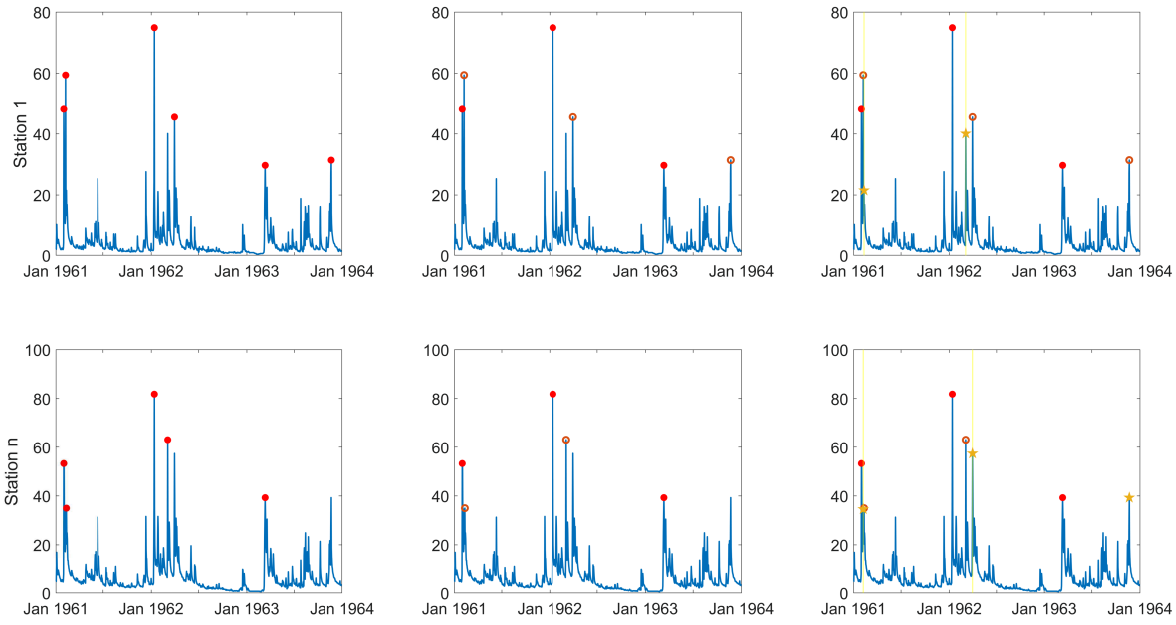


Figure 3.1: The procedure of flood events identification

In Figure 3.1, the top panel shows the time series of a station, and the bottom panel is for another station. Two left figures show the identified events independently from 1961 to 1965 in red dots. Some of the events happened at the same time (within two days intervals), and some individually took place. In the middle, individual events are displayed in red circles. These circles are the events that occurred independently. The figures on the right illustrate the corresponding discharge with the star shape symbol for each individual event (red circles). The highlighted yellow bound is a time window in which the corresponding discharge has the highest value.

3.2.3 Investigation of association and distance matrix - (pairwise investigation)

Pairwise associations and correlations are calculated for all discharge series. Peak discharges are analyzed to identify the frequencies of simultaneous floods. The time series are reduced to the time steps where there was a flood in at least one of the investigated catchments. Based on these correlations, Euclidean distance and Kendall distance are calculated. This procedure may lead to different clusters of simultaneously occurring floods depending on the considered series.

Every two stations have a pair set of coinciding extremes, which can be a minimum number of 110 pairs (i.e., two events per year). The minimum condition can arise when all peaks took place at the same time and be synchronized. Furthermore, 220 pairs would be a maximum range, thus pointing to a lack of coincidence in the peaks. To find the similarity of the orderings of the data when ranked by each of the quantities, Kendall's Tau correlation of each pair is calculated, and the square correlation matrix is computed. The distance matrix to clarify the discrepancies among stations is evaluated with two the following distance methods regarding this matrix.

Kendall's Tau coefficient

Kendall's Tau is grounded on counting the number of (i, j) couples, for $i < j$, that are compatible, for which $X_{a,i} - X_{a,j}$ and $Y_{b,i} - Y_{b,j}$ have the same sign (Kendall, 1948). Additionally, the equation for Kendall's Tau-b subtends an improvement in the normalizing constant for the ties.

The coefficient (τ) is determined for column X_a in matrix X and column Y_b in matrix Y, as:

$$\tau = \frac{2K}{n(n-1)} \tag{3.2}$$

where

$$K = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \xi^*(X_{a,i}, X_{a,j}, Y_{b,i}, Y_{b,j}) \tag{3.3}$$

and

$$\xi^*(X_{a,i}, X_{a,j}, Y_{b,i}, Y_{b,j}) = \begin{Bmatrix} 1 & if & (X_{a,i} - X_{a,j})(Y_{b,i} - Y_{b,j}) > 0 \\ 0 & if & (X_{a,i} - X_{a,j})(Y_{b,i} - Y_{b,j}) = 0 \\ -1 & if & (X_{a,i} - X_{a,j})(Y_{b,i} - Y_{b,j}) < 0 \end{Bmatrix} \tag{3.4}$$

The range of correlation is from -1 to $+1$. A value of -1 portends that one column ranking is the converse of the other, while a value of $+1$ demonstrates that the two rankings are identical. A value of 0 designates no relationship between the columns.

Distances are performed as follow (Equations 3.5 and 3.6):

Euclidean distance:

$$e_{jk} = \sqrt{\sum_{n=1}^{n}(x_{ij} - x_{ik})^2},\ 0 < e_{jk} < \infty \tag{3.5}$$

where j and k are two objects, n is the number of attributes available in the data objects, x_{ij} and x_{ik} are the values of the ith attribute of objects j and k, respectively.

Kendall's Tau distance:

$$d_{ij} = \sqrt{1 - \tau_{ij}},\ 0 < d_{ij} < +1 \tag{3.6}$$

The term of "Kendall distance", defined as Equation 3.6, where τ_{ij} is the rank correlation between selected peak pair set of objects i and j, d_{ij} is the distance of two objects i and j in terms of rank correlation, respectively.

3.2.4 Hierarchical cluster tree

Among the statistical methods used to investigate the spatiotemporal variations of climatological variables, the application of multivariate techniques by cluster analysis has been increasing in recent years (Modaresi Rad and Khalili, 2015). Clustering is the unsupervised classification of pattern recognition (Jain et al., 1999) and can delineate homogeneous regions and identify regional and global climate patterns (Unal et al., 2003; Lyra et al., 2014; Corporal-Lodangco and Leslie, 2017). The hierarchical cluster analysis technique is closely related to the quality and types of variables that can be grouped under different aspects and to identify similarity and dissimilarity patterns among study variables. Thus, due to the flexibility of combinations of similarity methods and metrics, hierarchical cluster analysis is suitable for different purposes and situations, making its application comprehensive and valuable for different types of variables and studies (Santos et al., 2019).

The clustering of catchments based on floods' simultaneous occurrence can be done by considering different dependence measures. Since a visual identification of clusters with similar flood behavior is impossible for multivariate cases, a geometric approach will be selected instead. Hierarchical algorithms do not construct a single partition with k clusters, but they deal with all values of k in the same run. It means if $k = 1$, all objects would be together in the same group, and if $k = n$, every object configures in an individual group includes only one value (Kaufman and Rousseeuw, 2009).

The hierarchical algorithm produces a dendrogram representing the nested gathering of patterns and similarity levels at which clusters change (Jain et al., 1999). The clustering process is performed by combining the most similar patterns in the cluster set to form a bigger one. The different hierarchical clustering algorithms are investigated, including Unweighted Pair Group Method with Arithmetic mean (UPGMA), Ward, Single, and Complete linkage

methods (Day and Edelsbrunner, 1984; Bouguettaya, 1996; Jain et al., 1999; Murtagh and Contreras, 2011, 2012; Bouguettaya et al., 2015).

3.2.4.1 Agglomerative Hierarchical Cluster Tree (AHCT)

Albeit pairwise investigations are unable to detect higher-order dependencies, but in some cases may occur that output derived from clustering is likely to have more severe consequences than foreseen from the pairwise investigation. Clustering may occur on slightly lower quantiles than pairwise tail dependence, which might lead to seemingly tail-independent pairs of catchments becoming dependent on large-scale flooding (Heffernan and Tawn, 2004). Hence, Davidson and Ravi (2005) presented an agglomerative hierarchical clustering by definition of some constraints, which improved the result of clustering. Agglomerative constructs hierarchy in the opposite direction, which makes different results. This method begins when all objects are away from each other, then two clusters are merged in each level. It will continue until one object remains only (Kaufman and Rousseeuw, 2009).

3.2.4.2 Construct agglomerative clusters for linkage

Linkage is defined as the distance between two clusters. Pursuant mathematics describes the linkages employed by diverse methods:

Cluster r is organized from clusters p and q. n_r is the number of objects in cluster r. x_{ri} is the ith object in cluster r.

Single linkage uses the smallest distance between objects in the two clusters, because of that it is also called nearest neighbor.

$$d(r,s) = min(dist(x_{ri}, x_{sj})), i \in (i, ..., n_r), j \in (1, ..., n_s) \tag{3.7}$$

Complete linkage as the farthest neighbor exerts the largest distance between objects in the two clusters.

$$d(r,s) = max(dist(x_{ri}, x_{sj})), i \in (1, ..., n_r), j \in (1, ..., n_s) \tag{3.8}$$

Average linkage employs the average distance between all pairs of objects in any two clusters.

$$d(r,s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} dist(x_{ri}, x_{sj}) \tag{3.9}$$

Weighted average linkage employs a recursive definition for the distance between two clusters. If cluster r was constructed by combining clusters p and q, the distance between clusters r and s is defined as the average of the distance between p and s and the distance between q and s.

$$d(r,s) = \frac{(d(p,s) + d(q,s))}{2} \tag{3.10}$$

Ward Jr (1963) presented a method that has been widely applied in semi-supervised learning. Ward is the Unique one among the agglomerative clustering methods, which is relayed

on an inner squared distance criterion, producing clusters that minimize within-groups dispersion between objects. Besides, Ward's method is fascinating because it seeks clusters in multivariate space (Murtagh and Legendre, 2014).

$$d(r,s) = \sqrt{\frac{2n_r n_s}{(n_r + n_s)}} ||\bar{x_r} - \bar{x_s}||_2 \tag{3.11}$$

where $||\ ||_2$ is the Euclidean distance. n_r and n_s are the numbers of elements in clusters r and s. Also, $\bar{x_r}$ and $\bar{x_s}$ are the centroids of clusters r and s.

3.2.4.3 Optimal leaf tree

An optimal leaf order catches a dendrogram tree, the corresponding distance $(tree, D)$, and reflects an optimal leaf ordering for the hierarchical dual cluster tree. This optimal ordering tree maximizes the sum of the similarities between adjacent leaves by flipping tree branches without dividing the clusters (Bar-Joseph et al., 2001). Also, the optimal leaf order returned as a length-M vector, where M is the number of leaves. Leaf order is a permutation of the first mth vector, giving an optimal leaf ordering based on the specified distances and similarity transformation.

3.2.4.4 Goodness-of-clustering

3.2.4.4.1 Inconsistency coefficient

The inconsistency coefficient is computed for each link k, as (Zahn, 1971; Jain and Dubes, 1988):

$$Y_k = \frac{Z_k - \bar{h}}{\sigma_h} \tag{3.12}$$

Where Y is the inconsistency coefficient for links in the hierarchical cluster tree Z. Also, $\bar{h}$ is the mean and σ_h is the standard deviation of all the links' heights included in the calculation. For links that have no further links below them, the inconsistency coefficient is set to 0.

3.2.4.4.2 Cophenetic coefficient

The cophenetic distance of two objects measures how similar those two objects have to be to be grouped into the same cluster. In terms of clustering in the dendrogram form, the cophenetic distance between two objects is the height of the dendrogram, where the two branches that include the two objects merge into a single branch. Thus, it is a measure of how accurately the tree demonstrates the dissimilarities among observations. The cophenetic distance between two observations is demonstrated in a dendrogram by the altitude of the connection at which those two observations are first attached. This altitude is the distance between the two sub-clusters that are merged by that link. The cophenetic correlation for a cluster tree (Z) is described as the linear correlation coefficient between the cophenetic distances acquired from the tree, and the actual distances/dissimilarities from the cluster configuration (Sokal and Rohlf, 1962; Farris, 1969; Holgersson, 1978). The largest cophenetic

correlation coefficient, the more desirable hierarchical clusters. The magnitude of this factor should be near to 1 for a high-quality solution. This measure can be employed to compare alternative cluster solutions obtained using different algorithms. The cophenetic correlation between Z (hierarchical cluster tree) and Y (distance matrix) is defined as:

$$C = \frac{\sum_{i<j} (Y_{ij} - y)(Z_{ij} - z)}{\sqrt{\sum_{i<j} (Y_{ij} - y)^2 \sum_{i<j} (Z_{ij} - z)^2}} \tag{3.13}$$

Where Y_{ij} is the distance between objects i and j in Y. Z_{ij} is the cophenetic distance between objects i and j, from Z. y and z are the mean of Y and Z, respectively. Also, the output value (C), is the cophenetic correlation coefficient.

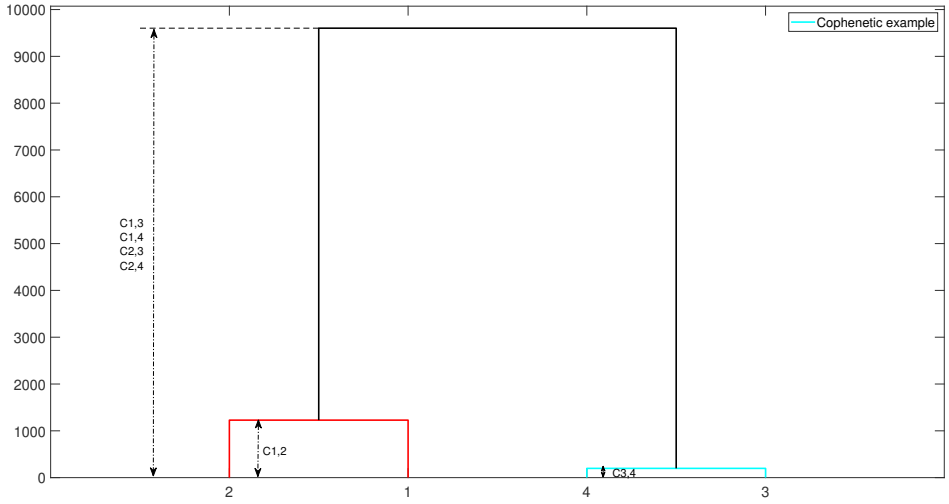


Figure 3.2: An example of dendrogram and its cophenetic distances $c_{i,j}$ between four stations.

An example of the dendrogram and the associated cophenetic distance is shown in Figure 3.2. Two downstream subcatchments and two upstream on the main river are selected, and their floods are identified in Table 3.1. Naturally, the magnitude of floods is higher in downstream than upstream. Therefore, The cophenetic distance between the first and second stations and third and fourth stations are low. It means, subcatchments with similar floods are clustered in the same cluster. Bigger substatements are in the red cluster and small ones are in cyan. The difference between the two low flood objects is small, then their cophenetic distance ($C_{3,4}$) is much lower than the two biggest in downstream ($C_{1,2}$). Cophenetic distances between stations 2 and 4 are huge and therefore, they cannot be similar to each other and be defined as separate groups. The Pearson's correlation coefficient corresponds to a quantitative comparison of the linear correlation between arrows, equal to 0.99 in this example. The Spearman's correlation coefficient (ρ) corresponds to a qualitative comparison of the clustering, which considers the ranks of the cophenetic distances between shapes. Here, ρ equals to 0.85.

In this example, the average flood in the first station (Rockenau) is 865.25 (m^3/sec), and the standard deviation equals 425.20 (m^3/sec). The average of floods in different locations

Table 3.1: Description of floods and selected station for the cophenetic example

Nr.	Mean*	SD	Area(km^2)	Location
1	865.25	425.20	$1.26 * 10^4$	Downstream
2	916.86	373.05	$1.37 * 10^4$	Downstream
3	50.50	24.38	$4.12 * 10^2$	Upstream
4	66.78	30.66	$6.53 * 10^2$	Upstream

*: in (m^3/sec)

shows apparent differences in up and mainstreams. Therefore, the high distances bring a high cophenetic coefficient, and low distances (similar series) have a low coefficient.

3.2.5 Silhouette value

The Silhouette value of each point is a scale of how similar that point is to points in its cluster compared to points in other clusters. In fact, one of the best cluster configuration validation techniques is the Silhouette coefficient (Chaimontree et al., 2010). The Silhouette value $s(i)$ for the ith point is defined as (Rousseeuw, 1987; Kaufman and Rousseeuw, 2009):

$$s(i) = \frac{b(i) - a(i)}{max(a(i), b(i))}, \quad \text{if } |C_i| > 1 \text{ and } i \in C_i \tag{3.14}$$

where $|C_i|$ is the number of members in the cluster that i is a member of it and $d(i,j)$ is the distance between data points i and j in the cluster C_i. $a(i)$ is the mean distance from the ith point to other points in the same cluster as i,

$$a(i) = \frac{1}{|C_i| - 1} \sum_{j \in C_i, i \neq j} d(i,j) \tag{3.15}$$

and $b(i)$ is the minimum average distance from the ith point to points in a different cluster (where $C_k \neq C_i$), and minimized over clusters.

$$b(i) = \min_{k \neq i} \frac{1}{|C_k|} \sum_{j \in C_k} d(i,j) \tag{3.16}$$

Global Silhouette score is define as:

$$S = \frac{1}{N} \sum_{i} s(i) \tag{3.17}$$

The range of the Silhouette value is from –1 to 1. A high Silhouette value indicates that i is well matched to its cluster, and poorly matched to other clusters. The clustering solution is appropriate when most points have a high Silhouette value, and if multiple points have a negative or low value, the clustering solution might have too many/few clusters. Silhouette values as a clustering assessment measure can be implemented with any distance metric.

3.2.6 Multidimensional scaling (MDS)

The Multidimensional Scaling (MDS) is a set of mathematical approaches to discover the "Hidden structure" of the dataset (Kruskal and Wish, 1978). This technique employs the proximities among various sorts of input objects. The output is a spatial representation on a map, consisting of a geometric configuration of points. Each point in the configuration corresponds to one of the objects. This configuration points out the hidden structure in the inputs and often makes the data much easier to comprehend. By reflecting the data structure, it means that the more considerable the dissimilarity (or the smaller the similarity) between the two objects, as shown by their proximity value, the further apart they should be in the spatial map. MDS is occasionally used indirectly to analyze data that are not proximities, by forming them as an intermediate step. The first theory of Multivariate analysis begun in the 1930s and was restricted to the multivariate normal distribution. Classical multidimensional scaling procedures stem back to Torgerson (1952); however, the basis of the recent multidimensional scaling was grounded by Seber (1984). Cox and Cox (2000) explained a narrow definition of multidimensional scaling, looking for low dimensional space. In other words, data in the form of points can be a plot with a scatter plot, but this plot might not be convenient in the pairwise distance form. While scatter plots of the input data make it straightforward to compare Euclidean distance, it does not work with other inter-point distances or more comprehensive dissimilarities. Also, it is complicated to illustrate distances in the multidimensional space. Hence, some dimension reduction is needed to represent data in a small number of dimensions. Therefore, an appropriate way to measure how near or far two set points is multidimensional scaling, which is a novel method that allows visualizing how near points are to each other for various sorts of dissimilarity or distance. MDS does not demand raw data, but only a matrix of pairwise distances or dissimilarities.

A distance matrix D estimates the inter-point distances of a point X configuration in a low-dimensional space p. That is, the elements of D, denoted d_{ij}, may be calculated from X by using the following equation Hintze (2019):

$$d_{ij} = \sqrt{\sum_{k=1}^{p} (x_{ij} - x_{ik})^2} \tag{3.18}$$

The classical MDS algorithm is as follows in these steps:

1. From D calculate $A = \left\{-\frac{1}{2} {d_{ij}}^2\right\}$.
2. From A calculate $B = \{a_{ij} - a_{i.} - a_{.j} + a_{..}\}$, where $a_{i.}$ is the average of all a_{ij} across j.
3. Find the p largest eigenvalues $\lambda_1 > \lambda_2 > ... > \lambda_p$ of B and corresponding eigenvectors $L = (L_1, L_2, ..., L_p)$ which are normalized so that ${L'}_i L_i = \lambda_i$. (Assumption: p is selected so that the eigenvalues are all relatively large and positive).
4. The coordinates of the objects are the rows of L.

The classical way is to minimize the sum of squared differences. That is when a straightforward solution is possible (i.e., when D is a Euclidean distance matric), the solution (L),

minimizes the sum of squared differences between the actual d_{ij}'s (elements of D) and the $\hat{d_{ij}}$'s based on L.

Here, multidimensional scaling is used as a tool to visualize clusters in a 2d space.

3.3 Results

3.3.1 Investigation of association and distance matrix

The Kendall's τ correlation is calculated among flood series. After that, every two sites had an absolute correlation between the greatest contemporary cases. Figure 3.3 shows the relationship between each station to the others.

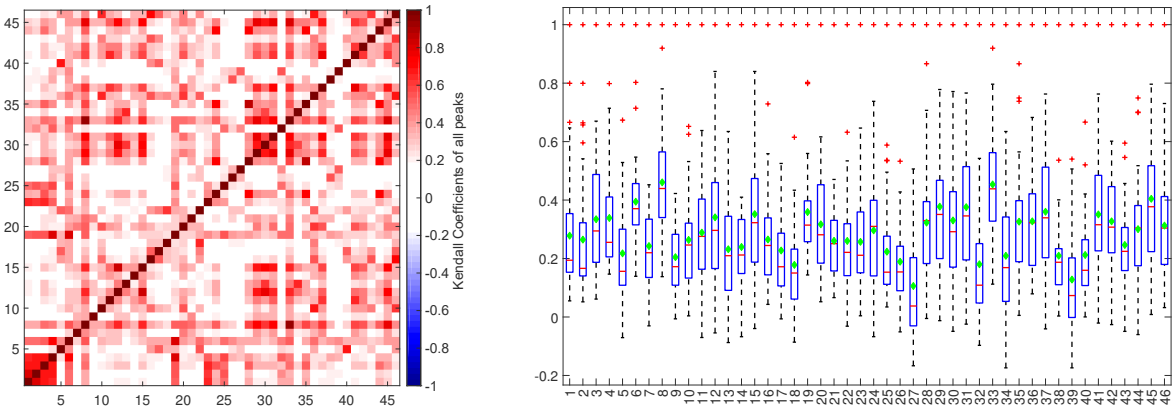


Figure 3.3: Rank correlation among all pair sets of extremes

On the left, the correlation matrix shows a positive correlation among almost all stations, caused by the proximity and the larger scale of weather patterns. The dark red shows a stronger relationship between two points, and pixels with near-zero values express no correlation. On the right figure, the range of changes in correlation is drawn. The large-size catchments have a higher correlation with other catchments, and smaller catchments generally have a lower correlation. This figure shows that some of the sub-catchments reacted opposite each other.

To convert this non-tangible plot to an explicit configuration, Figure 3.4 is illustrated. Then, for each measurement gauge as a reference, the rank correlation is traced in the space.

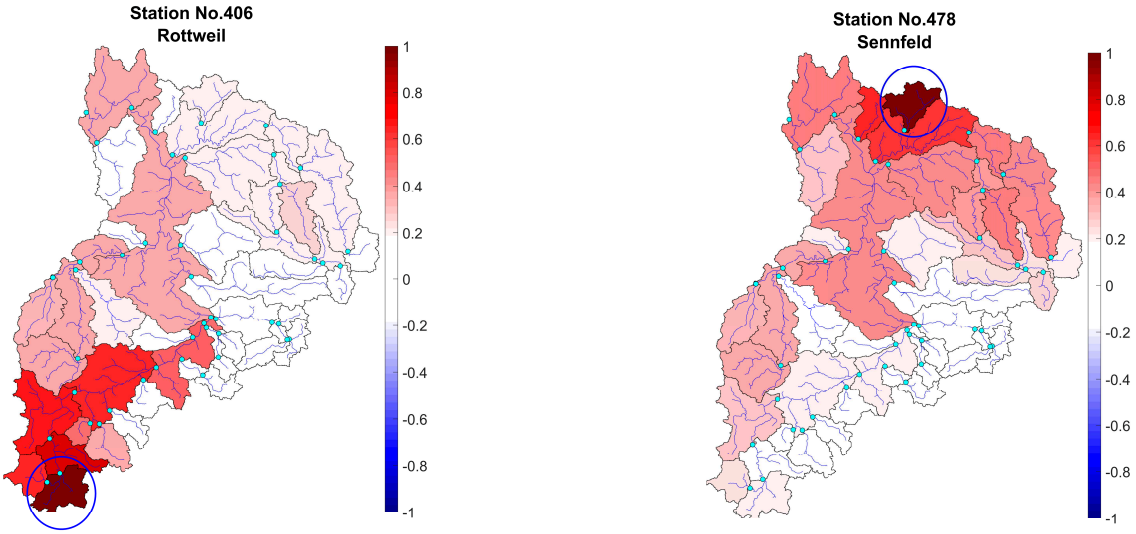


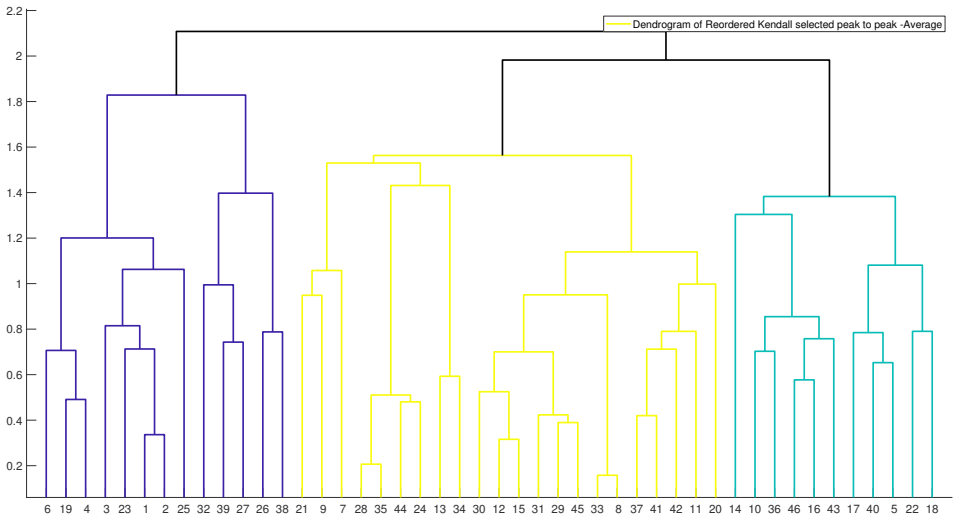
Figure 3.4: The simultaneous occurrence of the largest floods regarding the reference stations; Left: Rottweil - Right: Sennfeld

Two sub-catchments are selected from two distinct parts of the region to show various spatial reactions. Rottweil on the left figure as an upstream part is settled on; the darker red shows a strong association with the reference. This station is the origin of the main river (the Neckar river) in this case study area. It shows, neighboring sub-catchments react at the same time and with the same harmony in this part. Also, station Sennfeld in the north had different relationships to other catchments. All red color areas had a meaningful correlation with the reference catchments (highlighted in blue circle) in terms of vis-a-vis incidences. Although this station is located close to the outlet of the catchment. But it is located on the outer edge of the catchment and is one of the headwater subcatchments. Figure 3.4 shows an individual relationship of the selected subcatchments with other parts of the catchment. Therefore, still, the behavior of the region is almost obscure. The clustering analysis, by combining 46 dimensions, provides reasonable classes in the Neckar basin.

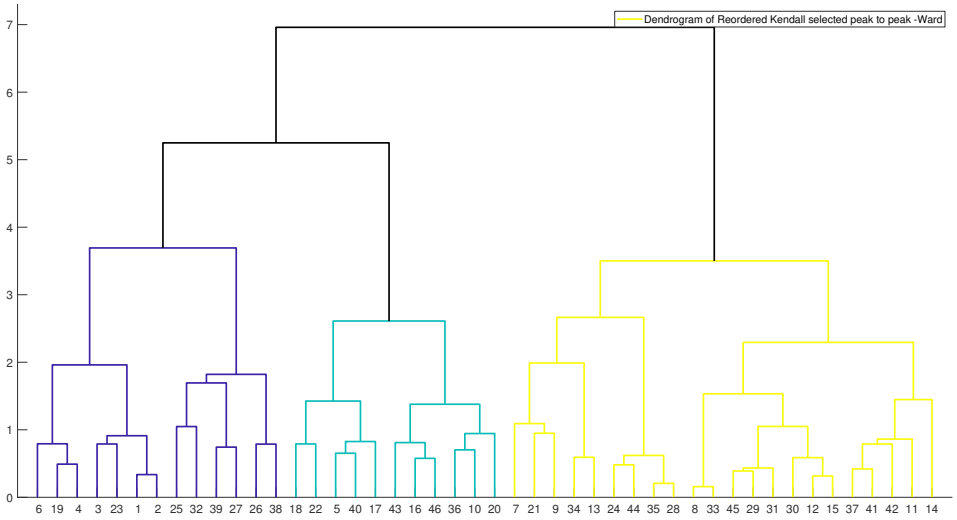
3.3.2 Cluster analyzing

3.3.2.1 Hierarchical tree and verification

Based on computed distance matrices, hierarchical cluster trees are calculated. This clustering reflects the simultaneous behavior of extremes in different catchments. In Figure 3.5, two samples of the linkage method are plotted with the highest Silhouette and cophenetic indices. The shape, structure and height of trees show distinguished groups.



(a) Average cluster tree



(b) Ward cluster tree

Figure 3.5: Hierarchical cluster tree of the simultaneous occurrence of the largest floods

Dendrogram presented similarities between the simultaneous peak discharges of the sampling sites. Furthermore, the three main clusters in this figure are entirely vivid. However, inside each batch, some sub-clusters illustrate the interconnection of sub-basins in a specific area. Therefore, to decide which linkage method or corresponding tree is more relevant to this research, and as a result, do the clustering based on that, three verification methods are exerted.

The mean of the inconsistency coefficient is reported in Table 3.2.

Table 3.2: Mean of inconsistency coefficient of the hierarchical cluster tree

	Dist.	Complete	Average	Weighted	Ward	Single
Inconsistency coefficient	Euc. dist.	0.542	0.540	0.542	0.544	0.515
	Kend. dist.	0.524	0.543	0.561	0.537	0.513
Number of zeros	Euc. dist.	17	16	16	18	15
	Kend. dist.	18	15	15	18	14

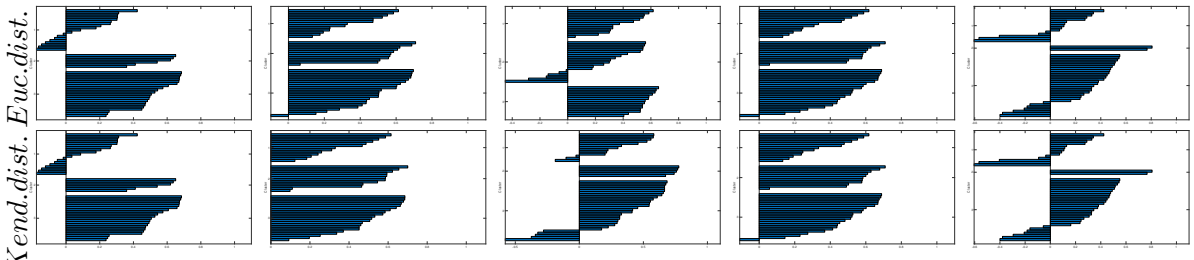
The lowest ratio is the Single one, and all the other factors are just about the same. It shows, different linkage methods reacted similarly. The higher the value of inconsistency coefficient, the less similar the objects connected by the link. Due to the low number of zeros in the Single method, it might be critical to quickly decide which linkage method is more appropriate in this research. The zero coefficients express the links which had no further links below them in the hierarchical tree. This factor is independent of the distance matrix; ergo, the cophenetic technique is applied in the pursuing step.

Also, the cophenetic coefficient is calculated and written in Table 3.3 for five linkage algorithms.

Table 3.3: The cophenetic coefficient of various linkage methods, based on distance matrices

	Complete	Average	Weighted	Ward	Single
Euclidean distance	0.725	0.787	0.676	0.730	0.630
Kendall distance	0.672	0.784	0.786	0.649	0.710

For the Euclidean distance, the best score is achieved by the Average algorithm and then Ward, Complete, Weighted, and Single, respectively. However, Average is still in the top two in the Kendall distance, but Weighted had the highest coefficient. The Single is in the third rank, and Complete and Ward are in order. So far, the Average method and then Ward or Weighted are reasonable for continuing phases of this research. However, the interpretation of the Single algorithm is for the shortest distance among the objects, and the shortest distance among dissimilarities means finding the most irrelevant catchments. Therefore, selecting the best two linkage methods took all three validation parameters into account.



(a) Complete (b) Average (c) Weighted (d) Ward (e) Single

Figure 3.6: Silhouette coefficient of the different hierarchical trees

The Silhouette factor is the most widely used clustering validation, owing to employing all possible related variables, including similarity metrics, estimated classes, distance matrix, and

tree. The x-axis of a plot is the Silhouette coefficient, and y-axis is all the stations into three clusters. The Silhouette coefficient shows the validity of each cluster, as it cleared the best two linkage methods to cluster in this area are Average (UPGMA) and Ward algorithms. According to Figure 3.6, the Complete, Weighted, and Single methods represented many negative values that show an unreliable cluster tree. Therefore, concerning the above plot, the two best possible ways of clustering the sub-basins are chosen.

3.3.2.2 Mapping clusters

Regardless of the results of the verification coefficients, the clusters are mapped into Figure 3.7. These maps illustrate massive differences by employing different nexus as well as distinct distance methods.

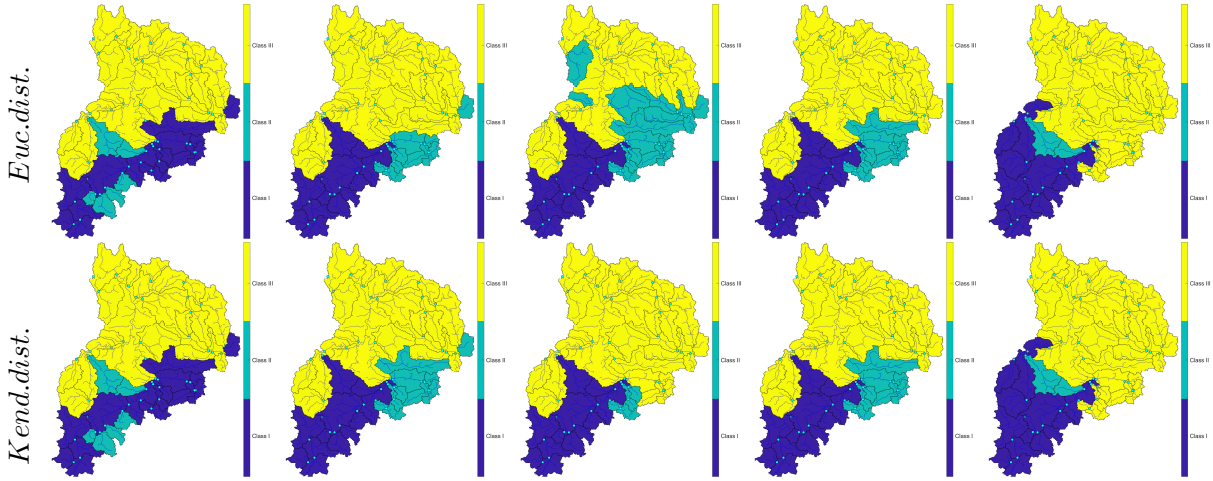
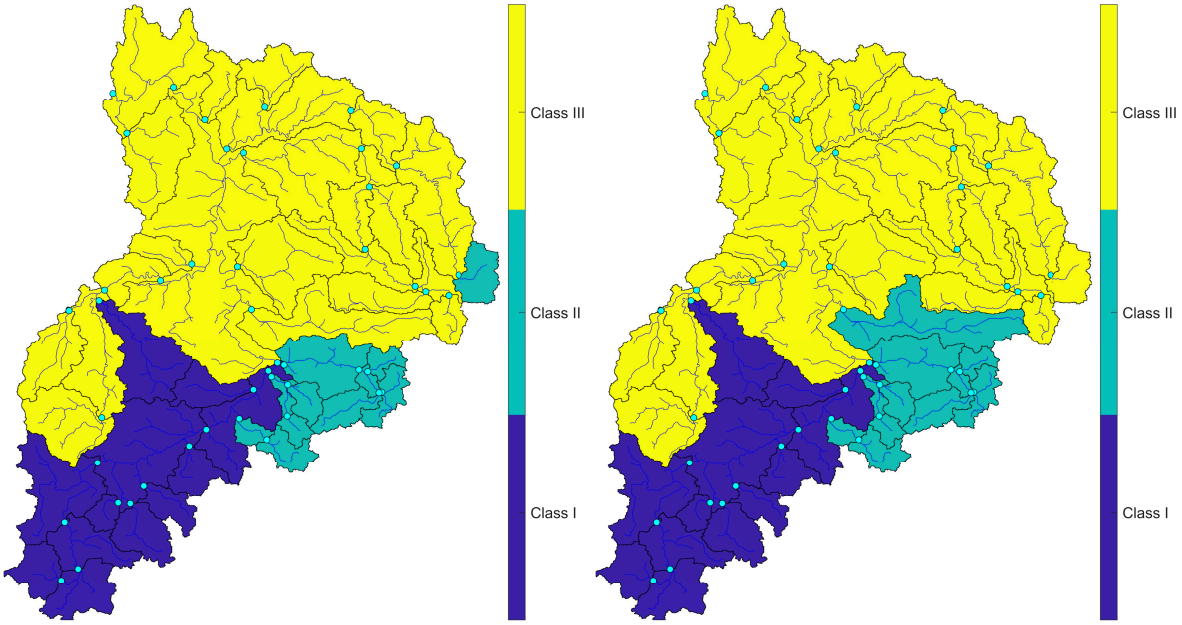


Figure 3.7: Spatial mapping of different hierarchical clustering

In Figure 3.7, it is evident that the application of each linkage manner can make a diverse pattern. However, concerning the research objectives and the result verification on dissimilarity and inconsistency and Silhouette value, the most considerable ways to interpret clustering are selected.

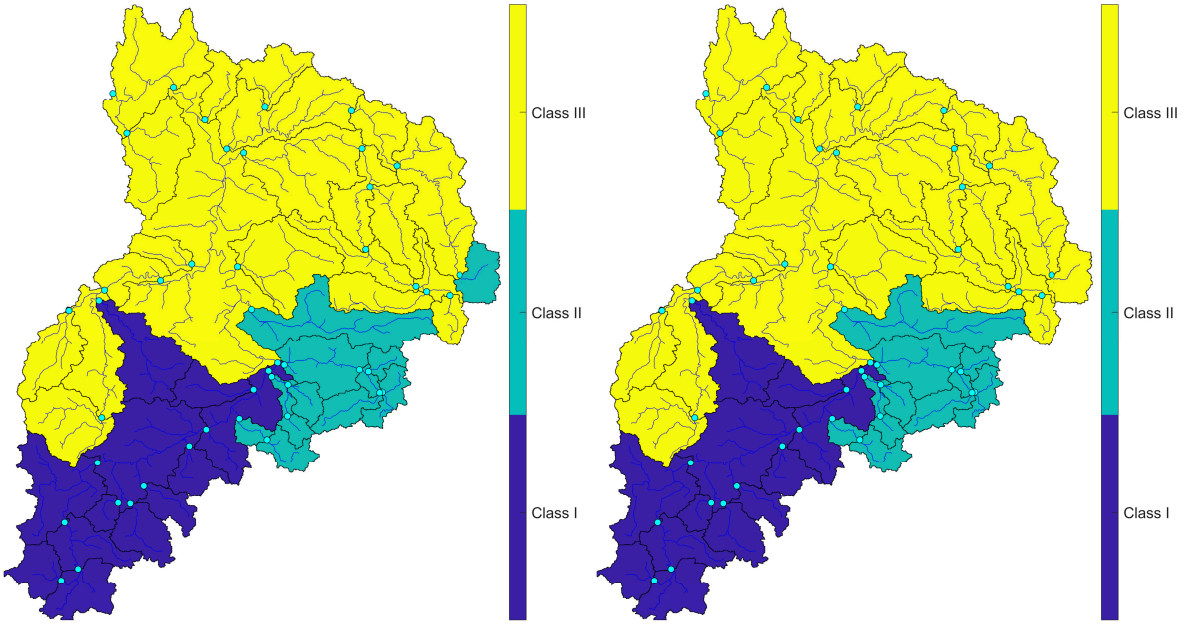
After realizing the best possible solidarity algorithm, the next step is to transform achieved outcomes and clusters to the map in a spatial domain using computed trees. In Figure 3.8, clustering is accomplished using Euclidean distance and with the two best linkage methods. Average and Ward linkage had the best verification factors among the others.



(a) Average linkage algorithm (b) Ward linkage algorithm

Figure 3.8: Clustering the simultaneous occurrences of the largest flood events using Euclidean distance

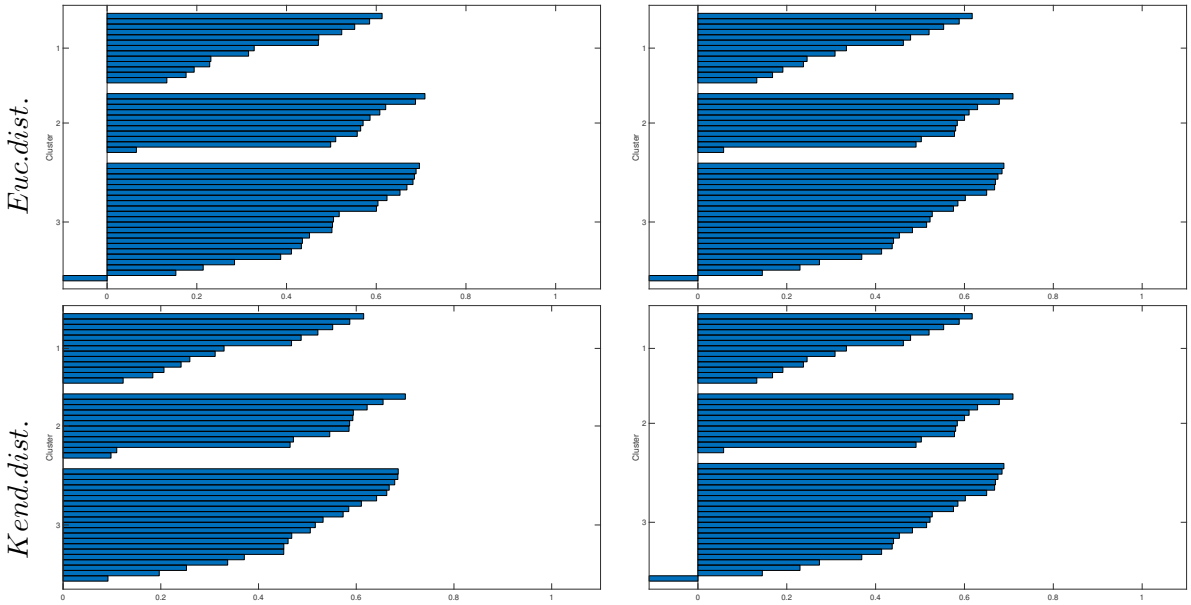
Results came from clustering represented three regions that have a well-defined flood mechanism. In Figure 3.8, synchronic occurrences of flood events are clustered using Euclidean distance. In this figure, the west part of the upper Neckar catchment is wholly separated from the other sub-catchments. The Average algorithm is used in Figure 3.8a, and the Ward linkage method is implemented in Figure 3.8b. The difference between these two plots is within clusters two and three. However, they had the most common reaction areas. Afterward, the distance method is changed to discover the difference between a varied distance manner to investigate the probable difference between these two methods.



(a) Average linkage algorithm (b) Ward linkage algorithm

Figure 3.9: Clustering the simultaneous occurrences of the largest flood events using Kendall distance

These two maps have the main pattern in clusters and only minor differences between second and third clusters, which is shown in Figure (3.9). Despite changing the distance algorithm, the clustered regions are mostly identical by using these two linkage methods. Only discharge gauges Nr. 1470 and 1411 reacted critically with this kind of clustering. The Silhouette diagram only for Average and Ward algorithms in two distance directions is projected to show the difference more visible. Then, by doing multidimensional scaling, the final clusters have become visible. These results show that the upper Neckar (described under Figure 2.8) is fully separated from the rest of the catchment. Also, the two disputed subcatchments are located on the Rems and Jags rivers with dozens of dams and hydraulic structures. In the west of the map, three subcatchments are clustered into the third cluster in yellow. The fewest hydraulic structures and dams have been constructed in this area. Exactly, after the Plochingen station (No. 427), the third cluster is started, where the main river has been mostly regulated.



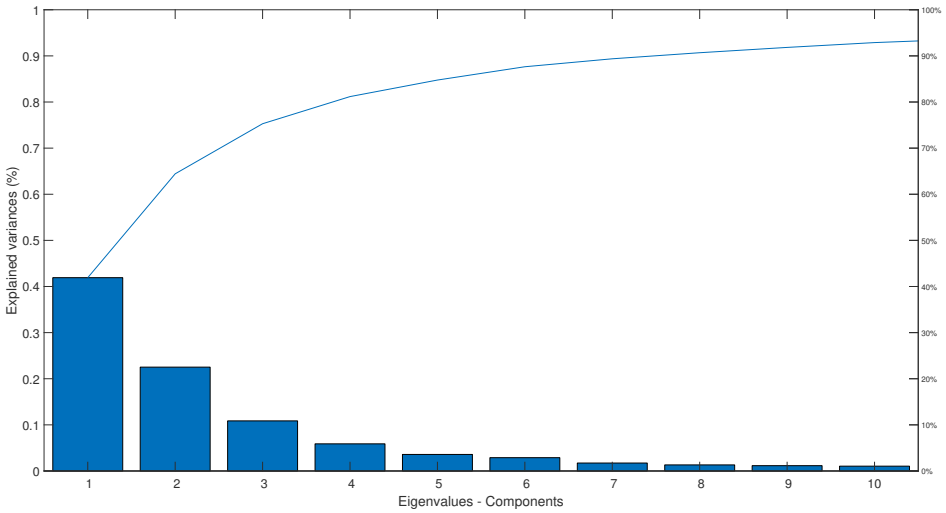
(a) Average linkage algorithm (b) Ward linkage algorithm

Figure 3.10: Silhouette coefficient of different hierarchical clusters

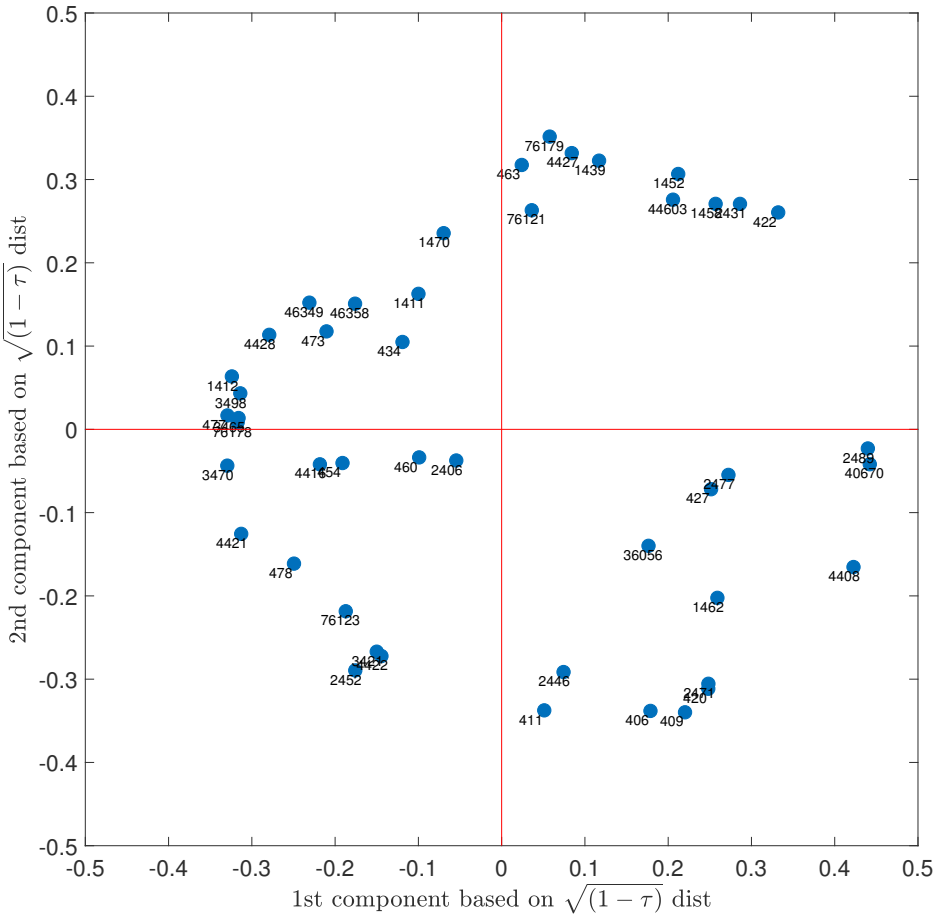
These algorithms with two different distance types are plotted to verify hierarchical clustering in the Neckar basin. By plotting Silhouette coefficient diagrams of the most sensible linkage methods in Figure 3.10, three out of four sub-figures except the lower left diagram (3.10a) show a unique problem in the third cluster, which appears in the last figure. Therefore, despite verifying the hierarchical cluster trees in the Neckar basin, two critical stations are still dubious. However, the Average linkage method employing Kendall distance reliably clustered the area.

3.3.3 Interpretation of AHCT using MDS

To visualize the multivariate data to the 2d plot, MDS is accomplished, and the result is revealed in Figure 3.11. Ten first eigenvalues are plotted in Figure 3.11a. These components explained more than 90 percent of the total variances of the dataset. In the next part of the plot, all the stations based on the Kendall distance of the correlation matrix are plotted. The nearer stations had a strong relationship with each other, and the farther had a weak connection. As it came to perceive, in this plot, separated groups are apparent.



(a) The percentage of explained variances by eigenvalues



(b) Scatter plot of two first principal components based on Kendall distance

Figure 3.11: Multidimensional scaling of the largest floods in the Neckar catchment

The first two components of the matrix of selected peaks contribute more than 65% to the

variance. By applying multidimensional scaling, interactions of the gauges are disclosed. This visualization shows that the first two eigenvalues capture the differences. As is clear from Figure 3.11b, the map might be divided into three concentration points. First is on the low right corner, second on the top right and one or two dots near them (Nr. 1470 or Nr. 1411), and third the remaining spots. Consequently, MDS and hierarchical tree clustering with the selected linkage methods are compared on the same map.

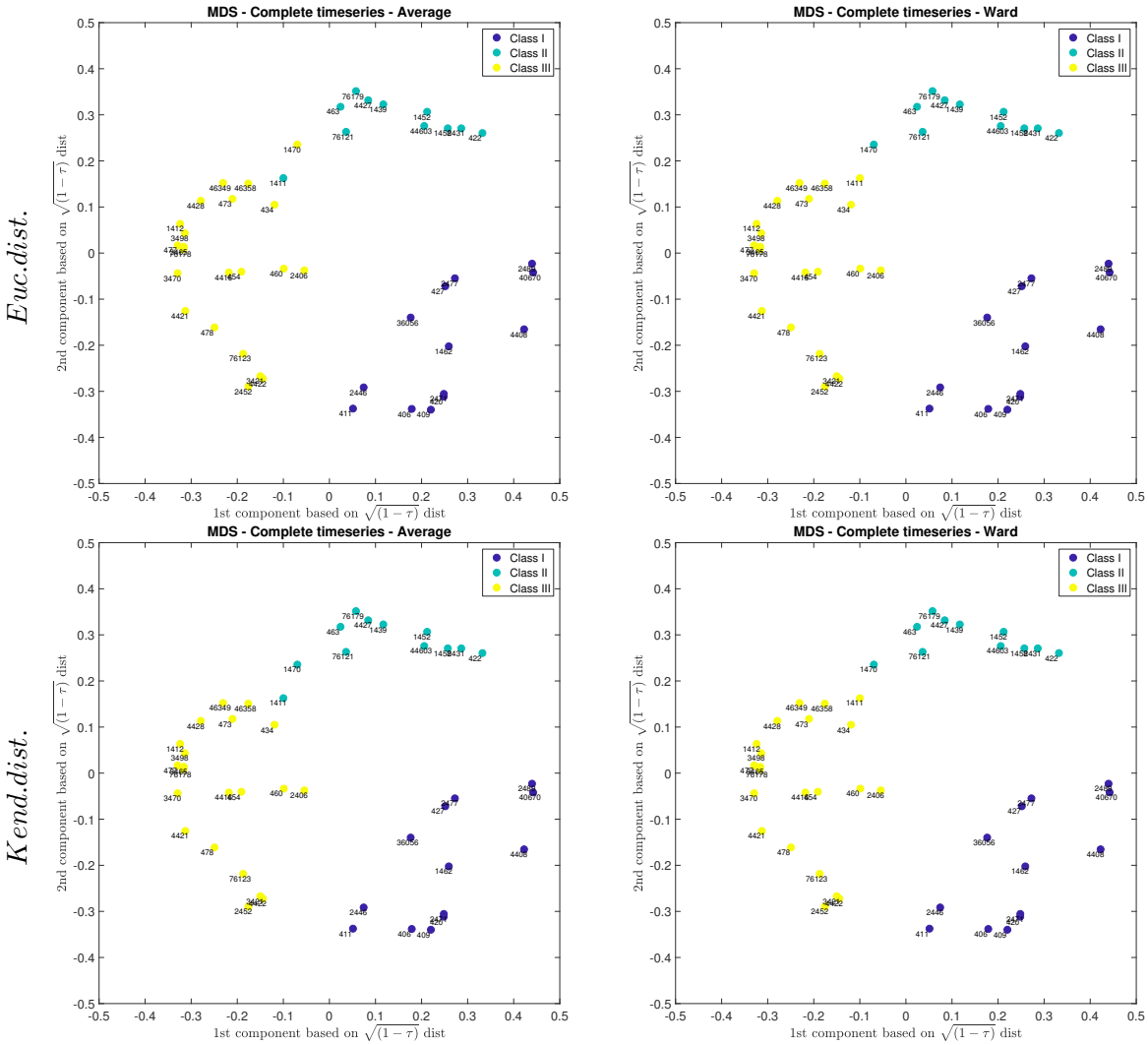


Figure 3.12: Comparison between multidimensional scaling and hierarchical clustering

Three determined classes are transformed into the scatter plot and in two-dimensional space. Each group is drawn with a specific color that is recognizable and comparable with the others. Regarding the above Figure (3.12), Ward linkage works very similar with both distance matrices and shows well-defined clusters. This algorithm illustrates consistent results that are not altered by changing the calculation of the dissimilarity matrix. In the first row of the above figure, this comparison is plotted by employing Euclidean distance, and the second row belongs to Kendall tau-based distance. The determined clusters are accepted by hierarchical

clustering, multidimensional scaling, and verified with the goodness of fit methods. Therefore, multidimensional scaling is verified entirely with the hierarchical clustering in both distance matrices. The small differences within the appointed manner appear here on the upper left side of Figure 3.12 between the light blue and yellow clusters and precisely for the measurement gauges of 1411 and 1470 on the southeast of the basin.

3.4 Conclusions

Floods destruction in vulnerable areas of river plains is critical for the community due to the high potential of casualties. Extreme events analysis has high uncertainty and is not well described in multidimensional space so far. High-resolution data and novel concepts for quantifying interactions undertake a primary step to achieve new results. However yet, the simultaneous occurrence of flood events has not been taken into account seriously. Identifying typical clusters of vulnerable flood areas requires a classification of concurrent flood patterns, which spatially shows to what extent floods happen simultaneously. In this thesis, the flood classification objective is formulated to determine concurrent flooding patterns merged by complementary and common patterns. These different classifications are compared to figure out how much the pattern is responsible for coincidence and to what extent this is a random effect as an individual pattern.

The clustering methods are used to estimate better multiple flood events evolution, which depend on the contribution of many sub-catchments. Also, clustering is a new method in multivariate analysis of the simultaneous occurrence of floods. For this reason, the hierarchical cluster tree and multidimensional scaling are applied. Neither of these methods needs initial assumptions, and they act independently of additional presumptions. The trees are formed based on rank correlation matrices of the highest occurrence of the floods in 55 years, and various linkage methods evaluated them. The Average group method appeared to produce the best values for two verification statistics, and the Ward algorithm is in the second rank. After Silhouette verification, which is the best integrated method for clustering validation, the results show that the Average (UPGMA) and Ward linkage methods are well-matched clustered in the Neckar basin. Therefore, these two linkage methods are selected for the following steps. Furthermore, The Neckar is split and mapped, and despite the small differences between the algorithms, the same pattern of clustering emerged. Then, these maps are compared by the outcome of multidimensional scaling. The results of MDS and Ward linkage in Agglomerative Hierarchical Cluster Tree (AHCT) perfectly matched each other, and consequently, the figure resulting from these two methods is selected as the final map.

The results show the simultaneous occurrences of high discharges operating as a function of the basin's topology and the precipitation's seasonality as the primary input of hydrological analysis. To conclude, the Neckar is divided into three major clusters: the first one is around the western part of the upper Neckar catchment and seized by the Black Forest and Swabian Alps, the second cluster is primarily located in the eastern region of the upper Neckar, with smaller sub-catchments and the karstic geology features. These two clusters are locating in the high elevation areas (see Figures 2.1 and 2.2b). The third part is the remaining area of the Neckar basin, which is a lowland area compared to the other parts. It can be mentioned that a reason for some clustering mismatches might be due to the anthropogenic alterations

in this area. Besides, the difference among clusters is observed in the southeast of the Neckar basin, where a dipping of formation characterizes the geological structure. It could be the reason for the disagreements. The turning point in this region is in the Plochingen, where the third cluster begins on the main river. After this point, the main river is mainly regulated (see Figure 2.8). Therefore, there is not a unique cluster on the main river. Most of the high elevated headwater subcatchments are located in the south and southwest of the Neckar (see Table 2.2), which are clustered into first and second clusters. These two clusters are in the upper Neckar region. In addition, there are three subcatchments in the west of Neckar, two of which are categorized as headwater regions. They clustered in the third cluster; it might be due to the small number of water reservoirs and dams in this part of the basin on Nagold river. The second cluster is mainly located in the upper Neckar, including subcatchments not on the main river, and they are on the Fils and Rems tributaries.

4 PCA-based clustering of the simultaneous occurrence of floods

4.1 Introduction

Extreme floods management is considered one of the most complex challenges in hydrology. Lack of sufficient knowledge of these events can lead to cataclysmic natural disasters. Floods have different types due to the distinct circulation patterns, seasons, and forms of precipitation. Also, the mechanism of triggering extremes could be different. Synchronous occurring floods can generally make a convergence of peaks in different catchment regions that can cause unpredictable losses and corresponding destruction downstream. Thus, investigating the mechanism of simultaneous floods is one of the newest challenges for hydrologists. Due to synchronized flood events in the Rhine, Main, and Neckar rivers in the last two decades, enormous losses near the urban regions took place (McPhillips et al., 2018; Kundzewicz, 2019; Modiri, 2021). After the flood in 1995 on the Meuse river in the Netherland, the government decided to evacuate around two hundred thousand people living in the flood plain areas (Geertsema et al., 2018). Past research has briefly pointed to the destructive nature of simultaneous floods in different parts of the World (Yinkang, 1996; Ahmad, 2003; Prohaska et al., 2008; Tsivtsivadze et al., 2019). These experiences showed concurrent extremes might deteriorate a get worse influence, leading to massive socio-economical impacts than a particular incidence form of an extreme (Meade et al., 1991; Leonard et al., 2014; Hao et al., 2018; McPhillips et al., 2018). Besides, flood defenses are critical components when it comes to flood perils. Therefore, it is substantial to deal accurately with their effects and consider the simultaneous flood occurrences.

An underlying problem in the flood analysis is that traditionally risk is estimated using one-dimensional extreme value statistics and ignoring multidimensional floods' interplay (Zhou et al., 2019). Different spatiotemporal precipitation events and the resulting outflows cause various combinations that can afford undercurrents to extreme floods (Blöschl et al., 2017, 2019). The risk arises when massive floods coincide in the mainstream and its tributaries. This enhanced risk is characterized by flood magnitude and occurrence date (Chen et al., 2012). Therefore, one of the major concerns of this chapter is to gain a better systemic understanding of multivariate analysis of such different flood-triggering regions to develop a methodology to simulate their behavior.

The discharge time series has various influential elements, including a spatiotemporal factor, like linear trend and seasonal terms (Ming et al., 2017; Milliner et al., 2018). Hydrological series have vivid periodic terms that are caused by seasons, geomorphological and meteorological patterns (Schmidt et al., 2008; Moeeni et al., 2017). Floods are events that rarely

occur and do not follow the primary behavior of the time series. In this chapter, initially, the Principal Component Analysis (PCA) is applied to the data set to capture the primary behavior of the discharge series over the Neckar. The first principal components that explain the high level of variances are chosen to reconstruct a new time series. Then, the difference between the original and reconstructed time series matrices is obtained as a residual time series. PCA is employed to separate the hydrological signal from temporally incoherent and spatially correlated noise (Milliner et al., 2018). In recent years, the residual time series has become a heated research topic in time series analysis (Yuan et al., 2018; Tan et al., 2020). The residual time series is computed by filtering the reconstructed time series from the original discharge time series. Effective extraction of residual time series is quite beneficial in catching pure extreme magnitudes out of seasonal trends and linear relations.

Therefore, this study considers the spatiotemporal multivariate analysis and quantitative characteristics of synchronous floods of the Neckar River and its tributaries to provide new insights to flood protection and risk analysis. Finally, the proposed chapter aims to investigate the coincidence occurrence of flood events in different clusters. These clusters incorporate components that can be used independently for developing models or integrated hydrological modeling. The difference between this chapter and the last one is the application of PCA on clustering analysis. Thus, the clustering maps can be a valuable decision aid for flood protection authorities and water management. Moreover, obtained clusters will be evaluated by different cluster evaluation methods to determine the best data grouping.

4.2 Methodology

In this chapter, the combination of statistical and dimension reduction methods is used to cluster the simultaneous occurrence of flood events. The PCA residual clustering technique, a hybrid method, aims to cluster data to capture similarities between the dependency of magnitudes of the biggest flood events. PCA is used to calculate the statistical residuals from the original series and reconstruct the time series based on selected principal components. In the next step, we compare the maps with the adopted clustering presented in the previous chapter and its relevant published researches (Modiri and Bárdossy, 2018, 2019b, 2021), intending to compare the current results following the last findings. The following section will outline the theoretical background of PCA and hierarchical clustering.

4.2.1 Principal Component Analysis (PCA)

PCA has been previously used to quantify the importance of the parameters that affect extreme events (Hudson and Colditz, 2003; Kourgialas et al., 2015). One of the main applications of the PCA technique is for identifying patterns and consequently clustering in multivariate data sets. This method can help express the data to highlight the similarities and dissimilarities, which may be challenging to investigate otherwise. The principal components are identified as physically independent processes controlling the variance in the hydrological time series' essential parameters. The relative importance of each variable is defined based on the amount of variance explained by each of the principal components. One of the PCA's key advantages is its low noise sensitivity (Karamizadeh et al., 2013). PCA should

be used mainly for variables which are strongly correlated. It is widely used to understand better the role of the nexus of climate variability (Thornton et al., 2006; Hudson et al., 2019; Koradia et al., 2019; Bahrami, 2019; Martínez et al., 2020; Rahman and Rahman, 2020). It is exclusively substantial when the variables vary simultaneously in multiple spatiotemporal dimensions (King and Jackson, 1999). The principal component analysis performed in this work aims to establish a process that neglects natural periodic terms and their effect on discharge time series in the spatial patterns and over the temporal changes. In other words, floods usually do not follow the main pattern of the time series, and they rarely occur. Thus, removing the underlying behavior of the hydrological time series lets us be one step near to the actual magnitude of floods. Consequently, it may improve the understanding of flood generating mechanisms in simultaneous occurrences.

PCA is defined as an orthogonal-linear transformation that transforms the data into a new coordinate system. The greatest variance by some scalar projection of data lies on the first coordinate as a first principal component. Subsequently, the second greatest variance lies in the second coordinate, continuing to the last PC (Jolliffe, 2002). Consider an $n * p$ data matrix, X, with standardized values. Each of the n rows expresses a different time step, and each of the p columns gives the measurement in a gauge. The transformation is mathematically defined by a set of size l of p-dimensional vectors of coefficients $w_{(j)}$ that map each row vector $x_{(i)}$ of X to a new vector of the principal component as a score $t_{(i)}$, given by:

$$t_{j\,(i)} = x_{(i)} \cdot w_{(j)} \qquad \text{for} \qquad i = 1,\ldots,n \qquad j = 1,\ldots,l \tag{4.1}$$

where

$$w_{(j)} = (w_1,\ldots,w_p)_{(j)}$$
$$t_{(i)} = (t_1,\ldots,t_l)_{(i)}$$

As follows that the individual variables $t_1,\ldots,t_l$ of t considered over the data set consecutively acquire the maximum possible variance from X, with each coefficient vector w bounded to be a unit vector, where l is usually selected to be less than p to reduce dimensionality.

4.2.1.1 Components computations

The first component

In order to maximize the variance, the first coefficient vector $w_{(1)}$ has to satisfy:

$$w_{(1)} = \underset{\|w\|=1}{\arg\max} \left\{ \sum_i (t_1)^2_{(i)} \right\} = \underset{\|w\|=1}{\arg\max} \left\{ \sum_i \left(x_{(i)} \cdot w \right)^2 \right\} \tag{4.2}$$

Equivalently, writing this in the matrix form gives:

$$w_{(1)} = \underset{\|w\|=1}{\arg\max} \{ \|Xw\|^2 \} = \underset{\|w\|=1}{\arg\max} \left\{ w^T X^T X w \right\} \tag{4.3}$$

Since $w_{(1)}$ has been specified to be a unit vector, it also equivalently serves:

$$w_{(1)} = \arg\max \left\{ \frac{w^T X^T X w}{w^T w} \right\} \tag{4.4}$$

The expected result for the matrix $X^T X$ is that the maximum possible value is the largest eigenvalue of the matrix, which arises when w is the commensurate eigenvector.

The first principal component of $x_{(i)}$ accepted as a score $t_{1(i)} = x_{(i)}.w_{(1)}$ in the transformed coordinates, or as the corresponding vector in the original variables, $\left\{x_{(i)}.w_{(1)}\right\} w_{(1)}$.

Further components

The jth component can be found by subtracting the first $j-1$ principal components from X:

$$\hat{X}_j = X - \sum_{s=1}^{j-1} X w_{(s)} w_{(s)}^{\mathrm{T}} \tag{4.5}$$

and then finding the coefficient vector, which extracts the maximum variance from this new data matrix:

$$w_{(j)} = \underset{\|w\|=1}{\arg\max} \left\{ \|\hat{X}_j w\|^2 \right\} = \arg\max \left\{ \frac{w^T \hat{X}_j^T \hat{X}_j w}{w^T w} \right\} \tag{4.6}$$

It reveals that this gives the remaining eigenvectors of $X^T X$, with the maximum values for the amount in braces given by their corresponding eigenvalues. Thus, the coefficient vectors are eigenvectors of $X^T X$.

The jth principal component of a data vector $x_{(i)}$ can, therefore, be given as a score $t_{j(i)} = x_{(i)}.w_{(j)}$ in the transformed coordinates, $\left\{x_{(i)}.w_{(j)}\right\} w_{(j)}$, where $w_{(j)}$ is the jth eigenvector of $X^T X$.

Therefore, the complete principal components decomposition of X can be given as:

$$T = XW \tag{4.7}$$

Where W is a $p*p$ matrix of coefficients whose columns are the eigenvectors of $X^T X$. Columns of W multiplied by the square root of corresponding eigenvalues, eigenvectors scaled up by the variances, are called "loadings" in PCA.

4.2.1.2 Singular Value Decomposition (SVD)

The principal components transformation in this research is associated with the Singular Value Decomposition (SVD) of X:

$$X = U \Sigma W^T \tag{4.8}$$

Here Σ is an $n*p$ rectangular diagonal matrix of positive numbers $\sigma_{(k)}$, called the singular values of X; U is an $n*n$ matrix, the left singular vectors of X is the columns with the length n orthogonal unit vectors, and W is a $p*p$ matrix whose columns are orthogonal unit vectors of length p and called the right singular vectors of X (see Figure 4.2a).

In terms of this factorization, the matrix $X^T X$ can be written:

$$\begin{aligned} X^T X &= W\Sigma^{\boldsymbol{T}} U^T U \Sigma W^T \\ &= W\Sigma^{\boldsymbol{T}} \Sigma W^T \\ &= W\hat{\Sigma}^2 W^T \end{aligned} \tag{4.9}$$

where $\hat{\Sigma}$ is the square diagonal matrix with the singular values of X and the excess zeros chopped off that satisfies $\hat{\Sigma}^2 = \Sigma^{\boldsymbol{T}}\Sigma$. Comparison with the eigenvector factorization of $X^T X$ establishes that the right singular vectors W of X are equivalent to the eigenvectors of $X^T X$, while the singular values $\sigma_{(k)}$ of X are equal to the square root of the eigenvalues $\lambda_{(k)}$ of $X^T X$.

The score matrix T by using the singular value decomposition can be written as:

$$\begin{aligned} T &= XW \\ &= U\Sigma W^T W \\ &= U\Sigma \end{aligned} \tag{4.10}$$

Each vector of T is given by one of the left singular vectors of X multiplied by the corresponding singular value. Therefore, computing the SVD is a standard way for calculating the principal components of a data matrix.

4.2.1.3 PCA and SVD properties

PCA finds directions of maximal variance of data. This technique determines mutually orthogonal directions. In other words, all the directions or the new features they find have an extensive global constraint, namely that they must be mutually orthogonal.

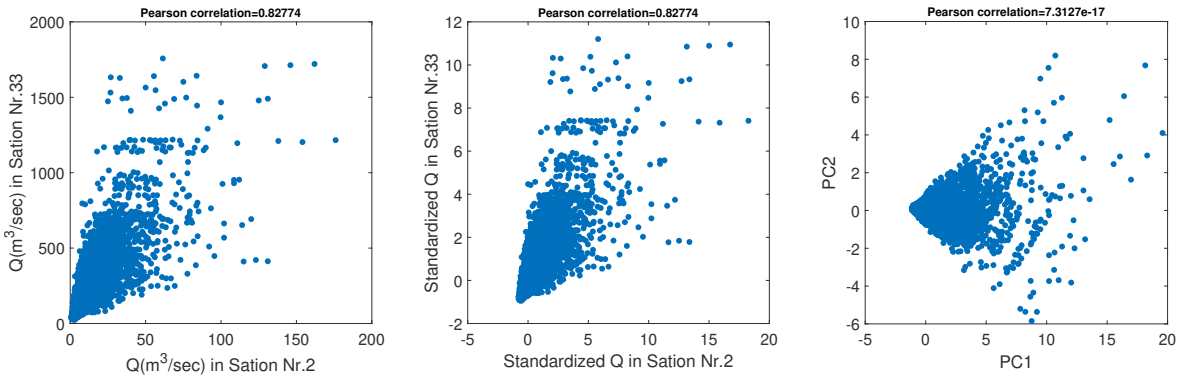
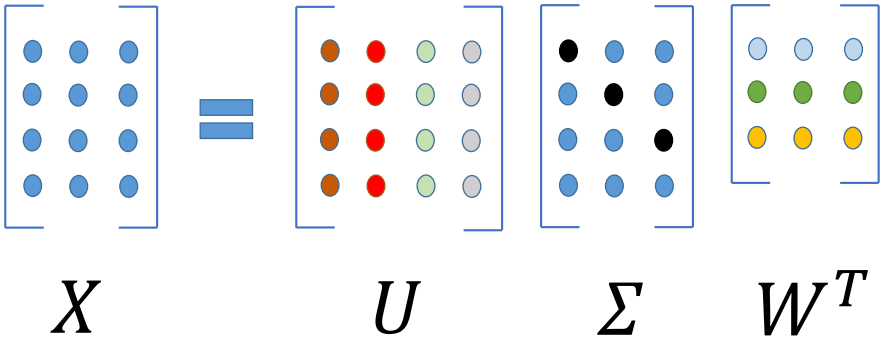


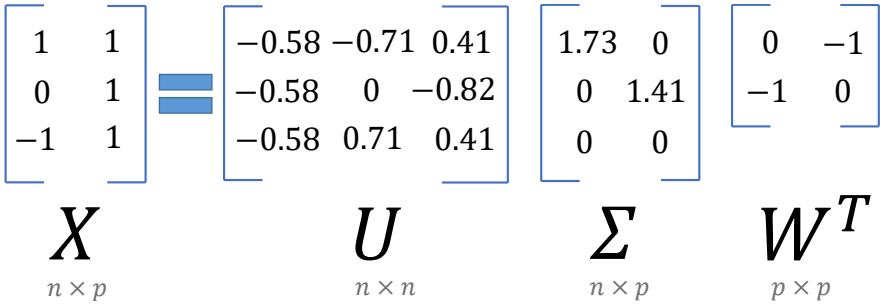
Figure 4.1: An example for the PCA's properties

In Figure 4.1, two measurement are randomly selected to show PCA's property. On the left, actual discharges are plotted so that they have a correlation of approximately 0.83. In the middle, the original data is standardized. Each series of data has an average of zero and a standard deviation of one. As a result, the correlation should not change, and it is equal to 0.83. On the right, the data is transformed to the principal component space. The two first PCs are illustrated, which correlate equally to zero. Consequently, regarding PCA's property, the correlation of all two PCs is equal to zero as its definition.

The principal component analysis is usually explained via an eigendecomposition of the covariance matrix. However, in the current research, it is performed via singular value decomposition of the data matrix X. Mathematically, there is no difference between whether PCA is calculated based on the data matrix directly or on its covariance matrix. The possible difference is solely due to numerical precision and complexity. In general, applying SVD directly to the data matrix is numerically more stable than to the covariance matrix.



(a) SVD's schematic matrices properties



(b) A simple mathematical example of SVD performance

Figure 4.2: SVD's properties

A schematic example of SVD matrices is visualized in Figure 4.2. SVD is more stable than typical eigenvalue decomposition procedures, especially for the highly collinear regressors in the machine learning approach. For example, 1000 random samples are generated for testing the stability of SVD and eigenvalue decomposition. In many cases, the eigenvalue decomposition method shows no small eigenvalues, which would lead to the singularity of the matrix. However, SVD is relatively twice more precise on a small eigenvalue determination, which may be essential depending on the problem. It is always possible to decompose a real matrix X into the SVD's components. These components are unique. In this chapter, PCA

is applied using SVD directly on the original standardized matrix.

4.2.1.4 Dimension reduction

The truncation of a matrix T using a truncated singular value decomposition generates a truncated matrix with the nearest possible matrix of rank L to the original matrix, known as the Eckart–Young theorem (Eckart and Young, 1936; Johnson, 1963). By having the eigendecomposition, a truncated $n * L$ score matrix T_L can be obtained by holding only the first L principal components that are produced by using the first L largest singular values and corresponding singular vectors, as follows:

$$T_L = U_L \Sigma_L = X W_L \tag{4.11}$$

where the matrix T_L has only L columns and n rows. Namely, PCA acquires skill in a linear transformation, addressed by:

$$t = W^T x, \qquad x \in R^p, t \in R^L, \tag{4.12}$$

where columns of $p * L$ matrix W configure an orthogonal structure for the L uncorrelated components of representation t. Construction of the transformed data matrices using L columns, this scoring matrix maximizes the variance in the original data, minimizing the total squared reconstruction errors.

$$\|TW^T - T_L W_L^T\|_2^2 = \|X - X_L\|_2^2 \tag{4.13}$$

The reduction of dimensions can be appropriate when the signals of variables within a data set are noisy. Dimensionality reduction can be a necessary step for visualizing and processing high-dimensional data sets while retaining as much variance in the data set as possible. Thus, PCA can concentrate the most detail of the signal into the first few principal components, which can usefully be captured by dimensional reduction.

4.2.1.5 Time series reconstruction

The flow discharge time series consists of a linear trend, periodic terms such as annual and semi-annual signals, offsets, and noise. Usually, hydrological models cannot correctly simulate the extreme components. Therefore, first, the underlying pattern among hydrological time series needs to be investigated, and consequently, new series will reconstruct. Then, the residual series should be calculated. The implemented method in this chapter is first introduced by (Modiri and Bárdossy, 2019a). Then, the reconstructed time series are computed using the SVD based PCA and as follows:

$$Rc_L = T_{1,\dots,L} * W'_{1,\dots,L} + \sigma_X + \mu_X \tag{4.14}$$

T is the score matrix and W' transposes the coefficient matrix for the first L components. σ_X and μ_X are the standard deviation and mean of the original matrix X, respectively.

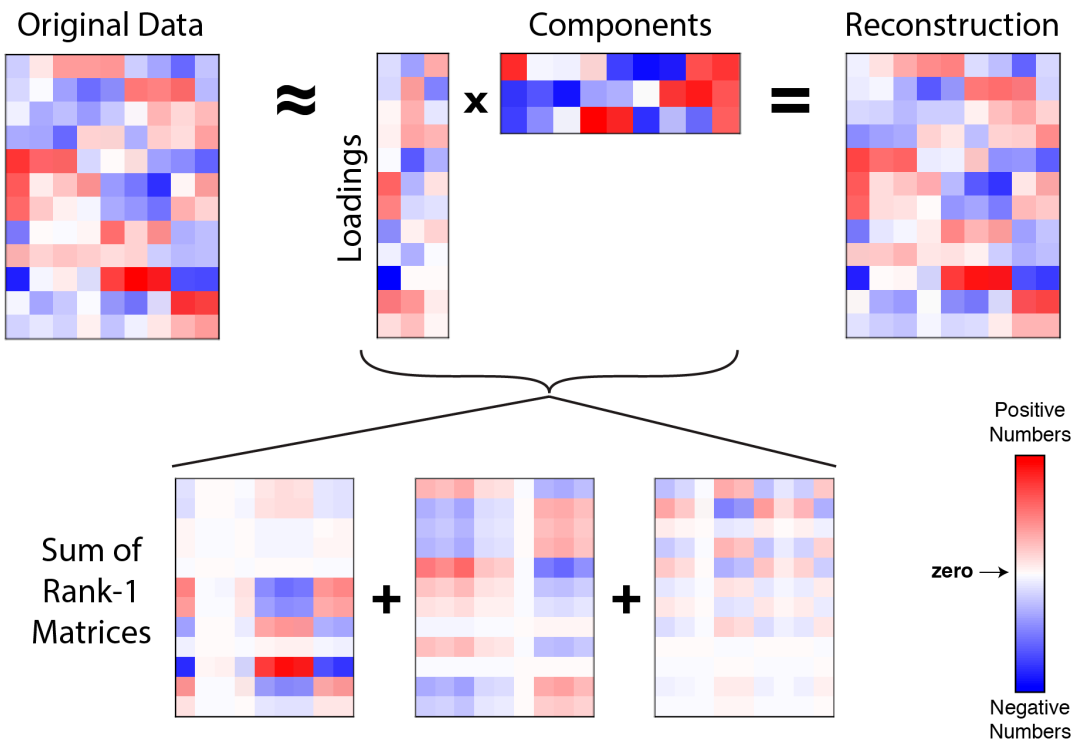


Figure 4.3: A schematic example of data reconstruction using three first PCs (Williams, 2016)

Figure 4.3 is an example for the reconstruction of data with the first three principal components. A data matrix on the left is approximated by the product of a $n * l$ matrix T and a $l * p$ transpose matrix W. This product is at most a rank-l matrix (in this case $l = 3$). Each paired column of loadings and row of components form an outer product, so the full reconstruction can also be thought of as a sum of l rank-one matrices. Finally, the mean and standard deviation have to be added to the reconstruction matrix to take the standardized value to the original form of data. The residual series represents the difference between the observations and the PCA's reconstructed series.

4.2.1.6 PCA cross-validation

It is always critical to determine the optimum number of principal components to reconstruct the time series. The leave one out cross-validation or so-called Jackknife can evaluate the quality of PCA to realize the appropriate number of dimensions by calculating Predicted sum of squares (PRESS) (Krzanowski, 1987; Diana and Tommasi, 2002; Abdi and Williams, 2010). Here, the Moore-Penrose pseudo-inverse method is applied to calculate the amount of PRESS and to correctly solve the problem of total least squares in the cross-validation (Penrose, 1955; Golub and Kahan, 1965; Barata and Hussein, 2012; Josse and Husson, 2012).

Let the data set consist of n points in d-dimensional space in Equation 4.15. First, perform PCA on the training set $\mathbf{X}^{(-i)}$ with an excluded single test data point $\mathbf{x}^{(i)}$ to compute reconstruction error. Then, take a certain number k of principal axes as columns of $\mathbf{U}^{(-i)}$,

and find the reconstruction error as follows:

$$\left\|\mathbf{x}^{(i)} - \hat{\mathbf{x}}^{(i)}\right\|^2 = \left\|\mathbf{x}^{(i)} - \mathbf{U}^{(-i)}\left[\mathbf{U}^{(-i)}\right]^\top \mathbf{x}^{(i)}\right\|^2, \quad \mathbf{x}^{(i)} \in \mathbb{R}^d, i = 1 \dots n. \tag{4.15}$$

Next step is to leave out one data point $\mathbf{x}^{(i)}$ at a time, compute PCA on the training data, but then to loop over dimensions of $\mathbf{x}^{(i)}$, leave them out one at a time and compute a reconstruction error using the rest.

$$\text{PRESS} = \sum_{i=1}^{n}\sum_{j=1}^{d}\left| x_j^{(i)} - \left[\mathbf{U}^{(-i)}\left[\mathbf{U}_{-j}^{(-i)}\right]^+ \mathbf{x}_{-j}^{(i)}\right]_j\right|^2 \tag{4.16}$$

Considering the inner loop, one point $\mathbf{x}^{(i)}$ is left out and computed k principal components on the training data, $\mathbf{U}^{(-i)}$. Then, keep each value $x_j^{(i)}$ as the test and employ the remaining dimensions $\mathbf{x}_{-j}^{(i)} \in \mathbb{R}^{d-1}$ to perform the prediction. The prediction $\hat{x}_j^{(i)}$ is the j-th coordinate of the projection of $\mathbf{x}_{-j}^{(i)}$ (i.e., in terms if the least squares) onto subspace spanned by $\mathbf{U}^{(-i)}$. Find a point $\hat{\mathbf{z}}$ in the principal component space $\mathbb{R}^k$, which is closest to $\mathbf{x}_{-j}^{(i)}$ by computing following equation:

$$\hat{\mathbf{z}} = \left[\mathbf{U}_{-j}^{(-i)}\right]^+ \mathbf{x}_{-j}^{(i)} \in \mathbb{R}^k, \tag{4.17}$$

where $\mathbf{U}_{-j}^{(-i)}$ is $\mathbf{U}^{(-i)}$ with j-th row kicked out, and $[.]^+$ stands for pseudoinverse. Finally, map $\hat{\mathbf{z}}$ back to the original space as follows:

$$\mathbf{U}^{(-i)}\left[\mathbf{U}_{-j}^{(-i)}\right]^+ \mathbf{x}_{-j}^{(i)} \tag{4.18}$$

Then take its j-th coordinate $[\cdot]_j$.

4.2.2 Flood events identification

In this chapter, the flood detection method is similar to the methods described in Section 3.2.1 processes. The difference is that the two biggest events per year in the absolute residual series for 55 years are selected for each gauge.

4.2.3 PCA-based hierarchical clustering of concurrent floods

In the following steps, the concurrency of peaks is recognized (see Section 3.2.2). Further the pairwise rank correlations of coinciding extremes are taken into account and different distance matrices are calculated (see sections 3.2.3 and 3.2.4).

Figure 4.4 illustrates the applied algorithms in this chapter.

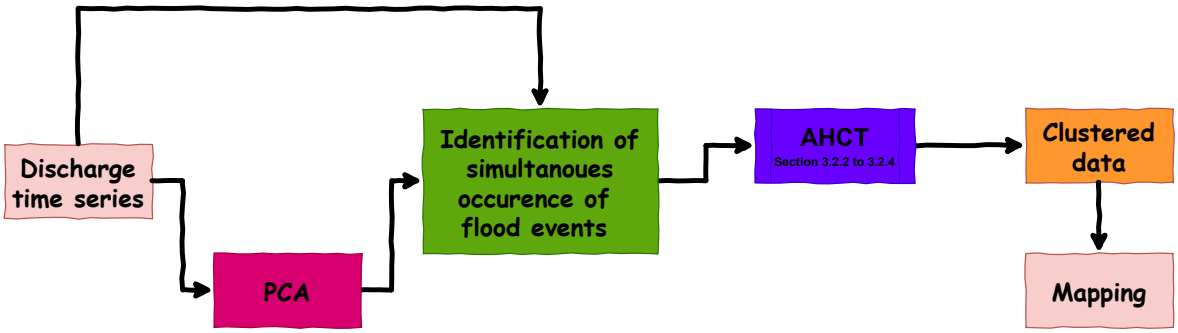


Figure 4.4: The applied algorithms in PCA-AHCT clustering

Also, the fast Fourier transform (FFT) periodicity detection is implemented for determining the most prominent cycles of the flood series (Lohre et al., 2003; Jalón-Rojas et al., 2016; Stojković et al., 2017). These cycles mainly affect the discharge time series.

4.2.4 Clusters similarity

4.2.4.1 Rand index

A mathematical coefficient can translate and interpret non-tangible dendrograms. The Rand index (RI) is a measurement index for the similarity between two data clusters (Rand, 1971).

If C is a ground truth cluster assignment and K the clustering, a and b can be defined as:

a: the number of pairs of elements that are in the same set in C and in the same set in K

b: the number of pairs of elements that are in different sets in C and in different sets in K The unadjusted Rand index is then given by:

$$\mathrm{RI} = \frac{a+b}{C_2^{n_{\mathrm{samples}}}} \tag{4.19}$$

where $C_2^{n_{\mathrm{samples}}}$ is the total number of possible pairs in the dataset. It does not matter if the calculation is performed on ordered or unordered pairs as long as it is performed consistently. However, the Rand index does not guarantee that random label assignments will get a value close to zero (especially if the number of clusters is in the same order of magnitude as the number of samples).

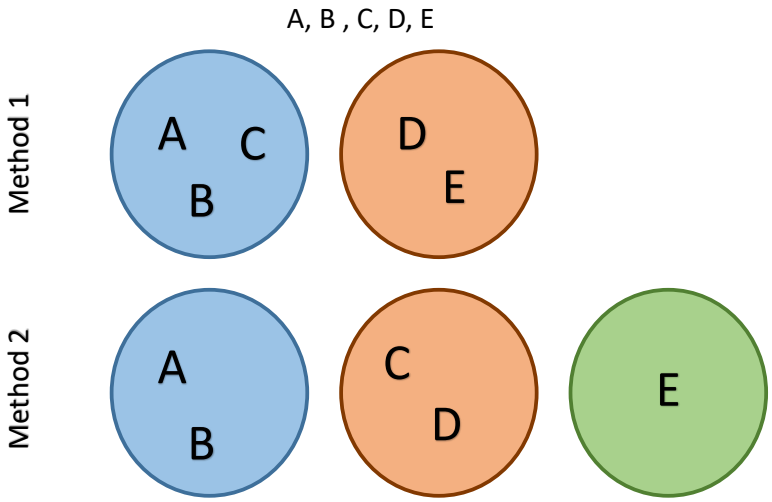


Figure 4.5: The schematic example of Rand Index calculation

As an example, we have a five elements dataset of $\{A,B,C,D,E,F\}$. It is supposed to use two clustering methods that place each element in the Figure 4.5. First, it is needed to write all possible unordered pair in the dataset. Therefore, the unordered pairs are: $\{A,B\},\{A,C\},\{A,D\},\{A,E\},\{B,C\},\{B,D\},\{B,E\},\{C,D\},\{C,E\},\{D,E\}$.

There are 10 pairs in this example. Then, the next step is to calculate a. In this case, $a=1$, which is equal to $\{A,B\}$.

Next, It is demanded to calculate b, which is the neumber of unordered pairs belong to different clusters across two clustering methods: These are: $\{A,D\},\{A,E\},\{B,D\},\{B,E\},\{C,E\}$.

Finally, the Rand Index is $(1+5)/10=0.6$.

The ARI is the normalized difference of the RI. It assumes a generalized hypergeometric distribution. It means the null hypothesis is that the two clusterings are grouped randomly with a fixed number of clusters and the number of elements in each cluster (Hubert and Arabie, 1985). It is possible to discount the expected RI $E[\mathrm{RI}]$ of random labelings by defining the adjusted Rand index to counter this effect as follows:

$$\mathrm{ARI}=\frac{\mathrm{RI}-E[\mathrm{RI}]}{\max(\mathrm{RI})-E[\mathrm{RI}]} \tag{4.20}$$

The ARI may result in a value between 0 for completely different and 1 for identical clusters, except when the RI is less than the expected Rand index. In this situation, the ARI can produce negative values (Meilă, 2003; Wagner and Wagner, 2007).

For example, U and V are two random partitions with multiple cluster inside. n_{ij} is the number of objects that are in both cluster u_i and v_j. Let n_i and n_j be the number of objects or elements in cluster u_i and cluster v_i, respectively. The notations are illustrated in the following table (Table 4.1).

Table 4.1: The Adjusted Rand Index contingency table

$X\backslash Y$	Y_1	Y_2	$\cdots$	Y_s	sum
X_1	n_{11}	n_{12}	$\cdots$	n_{1s}	a_1
X_2	n_{21}	n_{22}	$\cdots$	n_{2s}	a_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
X_r	n_{r1}	n_{r2}	$\cdots$	n_{rs}	a_r
sum	b_1	b_2	$\cdots$	b_s	

The adjusted rand index using the contingency table can be written as:

$$ARI = \frac{\sum_{ij}\binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right]\binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right]/\binom{n}{2}} \tag{4.21}$$

where n_{ij}, a_i, b_j are values from the contingency table. The n_{11} would be the number of times an element in the first cluster of X and the first cluster of Y. a refers to the sum of row value and b refers to the sum of the column value.

Assume that $X = \{1,2,3,3,2,1,1,3,3,1,2,2\}$, and $Y = \{3,2,3,2,2,1,1,2,3,1,3,1\}$ are the dataset clustered into three partitions. Each value represents each cluster. Also, the index is play an essential roll to calculate ARI. It is necessary to build the contingency table by counting the coincidences. In Table 4.2 the contingency table are calculated.

Table 4.2: The contingency table of the test dataset

$X\backslash Y$	Y_1	Y_2	Y_3	sum
X_1	3	0	1	4
X_2	1	2	1	4
X_3	0	2	2	4
sum	4	4	4	

As a result the adjusted Rand Index is equal to:

$$ARI = \frac{6 - [18*18]\binom{12}{2}}{\frac{1}{2}[18+18] - [18*18]/\binom{12}{2}} = 0.083$$

4.2.4.2 Fowlkes–Mallows index

The FM is an external evaluation method that is used to measure the between two hierarchical clusterings (Fowlkes and Mallows, 1983).

The hierarchical trees C_1 and C_2 are cut to produce $k = 2,\ldots,n-1$ clusters for each tree. For each value of k we can label the clusters for C_1 and C_2 arbitrarily from 1 to k and then form the matrix

$$M = [m_{ij}], \quad i = 1,\ldots,k; \quad j = 1,\ldots,k$$

where the quantity m_{ij} is the number of objects in common between the i-th cluster of C_1 and the j-th cluster of C_2. The FM is defined as follow:

$$FM_k = T_k / \sqrt{P_k Q_k} \tag{4.22}$$

The measure of association (FM_k) is derived from the matching matrix (M), which is formed by cutting the two hierarchical trees and counting the number of matching entries in the k clusters. Where, n is the number of objects, and

$$\begin{aligned}
T_k &= \sum_{i=1}^{k}\sum_{j=1}^{k} {m_{ij}}^2 - n, \\
m_{i.} &= \sum_{j=1}^{k} m_{ij}, \\
m_{.j} &= \sum_{i=1}^{k} m_{ij}, \\
m_{..} &= n = \sum_{i=1}^{k}\sum_{j=1}^{k} m_{ij}, \\
P_k &= \sum_{i=1}^{k} {m_{i.}}^2 - n, \\
Q_k &= \sum_{i=1}^{k} {m_{.j}}^2 - n
\end{aligned} \tag{4.23}$$

The Equation 4.22 can be written in terms of contingency table definition as follow:

$$FM = \frac{TP}{\sqrt{(TP+FP)(TP+FN)}} \tag{4.24}$$

Where TP is the number of True Positive, the number of pairs of points belonging to the same clusters in both the true and predicted classes. FP is the number of False Positive, equal to the number of pairs of points that belong to the same clusters in the true classes and not in the predicted classes. Finally, FN is the number of False Negative, which is the number of pair of points that belongs in the same clusters in the predicted classes and not in the true classes.

A higher value for the FM indicates a more remarkable similarity between the hierarchical clusters. Fowlkes–Mallows index is not sensitive to the labeling numbers. The FM index is close to zero when labels are randomly (uniformly) assigned. Also, it is bounded at one as the upper bound. Further, values of exactly 0 and 1 indicate purely independent label assignments and equal labeling (with or without permutation). In addition, this index has no assumption to compare clustering algorithms. These points, as mentioned earlier, are some of the advantages of FM (Nemec and Brinkhurst, 1988; Wagner and Wagner, 2007).

For example, four stations in the Figure 3.8 are selected. These are stations with ID-Nr.

4416, 454, 509, 1411. The obtained cluster labels are $x_1 = [3,3,1,2]$ in Figure 3.8a. But in Figure 3.8b, the labels are as follows: $x_2 = [3,3,1,3]$. According to Equation 4.22, the FM index equals 0.577. If the labels are equal to $x_3 = [2,2,1,2]$ or $x_4 = [1,1,2,1]$, the FM would be again equal to 0.577.

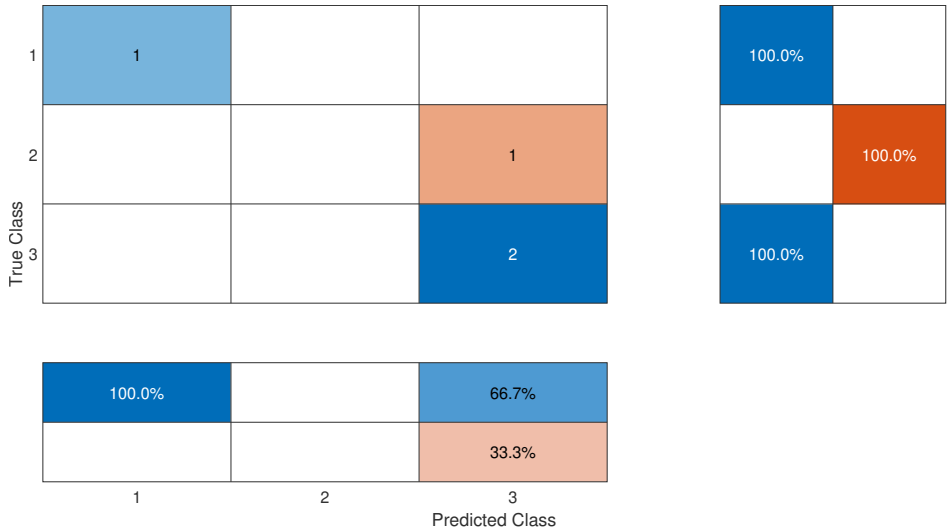


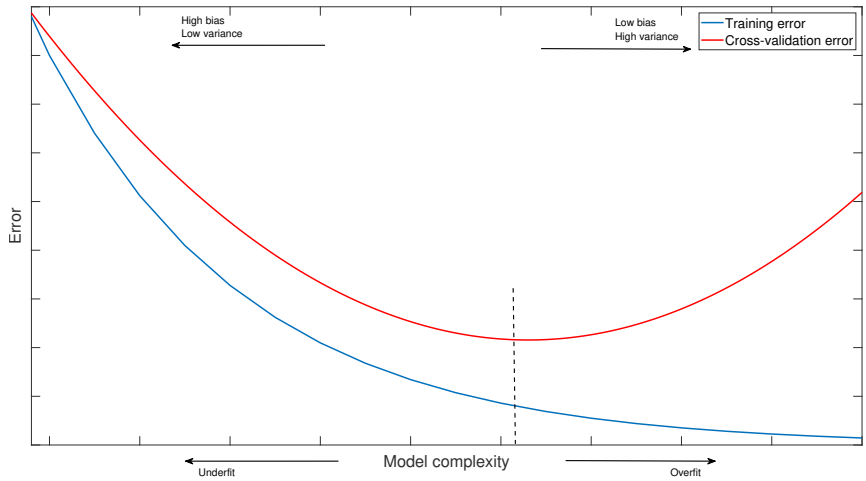
Figure 4.6: The confusion matrix of designed example

The confusion matrix in Figure 4.6, shows that class one is 100% correctly clustered, the second class does not exist in the second clusters. The third class label is 66.7% correct clustered and 33.3% wrong. This matrix is an input of the Fowlkes–Mallows index, where Equation 4.22 counts the number of elements in each row and column.

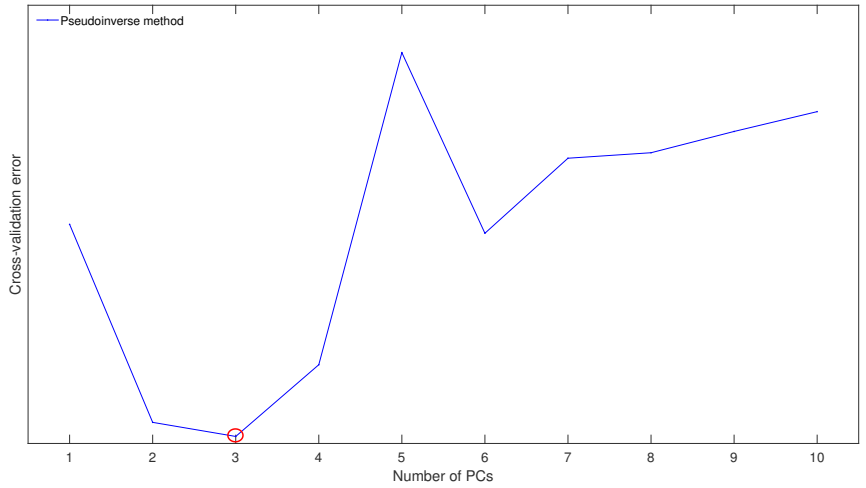
4.3 Results

4.3.1 Reconstructed time series using PCA

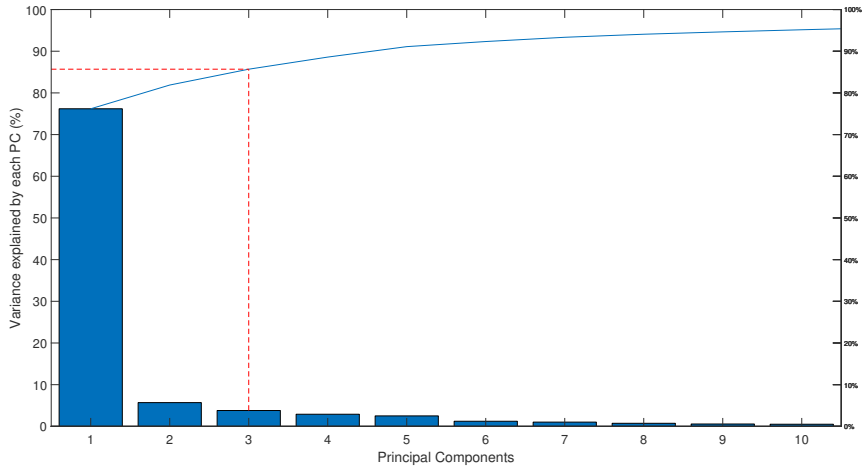
Since flood events rarely occur, their magnitudes and distributions do not directly follow the annual and seasonal signals. Therefore, these series do not have the main periodic terms, which is unnecessary in such cluster analysis. Instead, PCA is applied to catch the main and shared behavior of all dimensions. In the following figure, the error of the cross-validation technique in the principal component analysis is illustrated to select the appropriate number of PCs to reconstruct the new series according to the Section 4.2.1.6. Consequently, the percentage of variability explained by the selected number of principal components is calculated.



(a) Training error and cross-validation error, as a function of model complexity



(b) PCA cross-validation errors



(c) Explained variances by the first ten principal components

Figure 4.7: Schematic training and test samples error estimation according to the model complexity in 4.7a; determining the best number of clusters and the cut-off threshold for the explained variability of data using the pseudo-inverse method in 4.7b; and the variances explained by each PC for the first ten PCs in 4.7c.

In Figure 4.7, first it is illustrated that by adding PCs the complexity of model will increase and consequently variance will go high and bias low (Figure 4.7a). The training samples error will decrease; however, the cross-validation error after a turning point will increase. Therefore, it is demanded to find the optimal number of components to have the lowest cross-validation error (Ballabio, 2009). Adding and subtracting components will cause over or under-fitting. Then, the first ten principal components are illustrated. The Pseudo-inverse cross-validation errors (PRESS) are shown for the first ten principal components in Figure 4.7b. By adding one by one PCs, cross-validation error decreases from 1 to 3 PCs. In fact, the minimum point of the cross-validation error series is achieved on the third PC. Therefore, the best number of components to reconstruct the data set is determined. The scree plot[1] shows the data variances explained by each PC (Figure 4.7). This plot illustrates the first ten components that explain 95% of the total variance. The first PC explains more than 75% of the data set's variability, which denotes that the discharge time series are highly correlated. The blue line displays the accumulation of explained variances, which has an intersection point with the third principal component. Therefore, the cut-off threshold is plotted in the red dash line and it is equal to 85.7% of total variances. Another point that can recognize is that the gradient of explained variances is very low and almost steady after the third PC.

Based on equations 4.12 and 4.14, and by applying the first three PCs, the reconstructed time series for each measurement gauge is calculated.

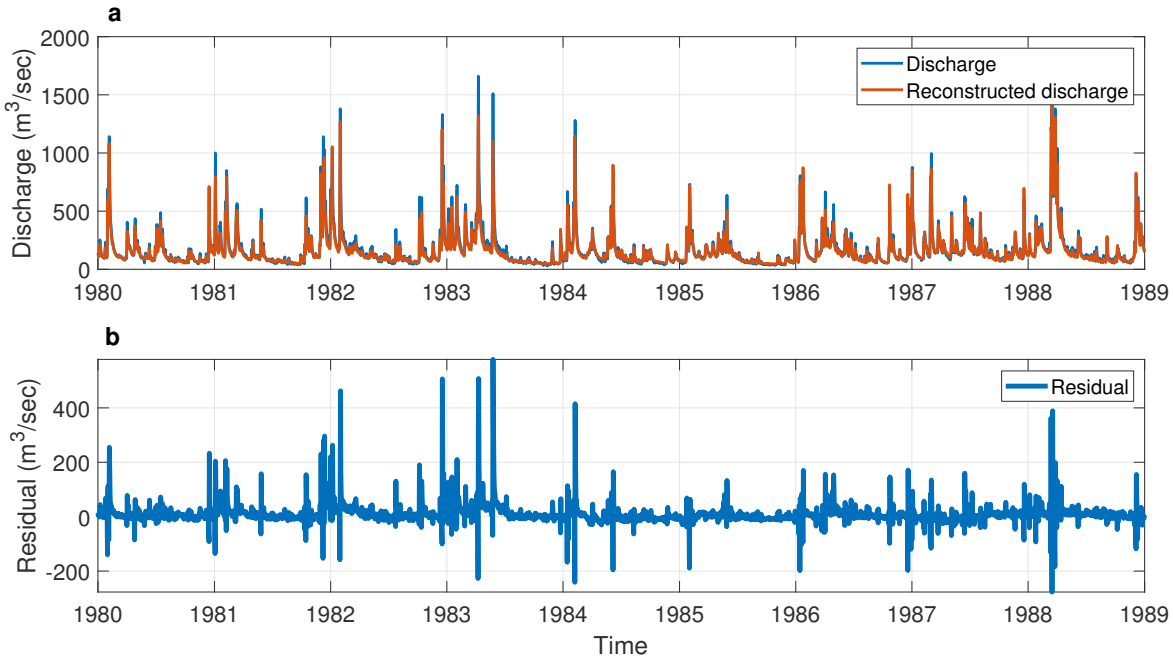


Figure 4.8: a) Reconstructed time series based on the first three PCs, and b) Residual time series

The first three PCs explain more than 85 percent of the discharge time series variability. Floods are infrequent events that usually do not follow the underlying behavior of the time

[1] A scree plot is a line or bar plot of the eigenvalues of factors or principal components on the x-axis in multivariate statistics. It determines the number of factors to retain in principal components to keep in a principal component analysis (Jackson, 1993).

series. Therefore, it is better to assess their spatial interactions when the time series does not include the primary behavior. PCA is a great tool to catch this underlying behavior of the study area. Extreme floods are caused and triggered by short waves, which are not inside the principal conduct of a catchment. Thus, instead of using the reconstructed series, the residual series are employed to concentrate on irregular series variations, i.e., mainly short waves. In part *a* of Figure 4.8, the original discharge flow time series for one station and the corresponding reconstructed time series are plotted. The two time series are highly correlated and the reconstructed one follows the original time series' behavior. The most significant errors are in the case of floods. Subsequently, Figure 4.8, part *b*, presents the residual series adopted by subtracting the two mentioned time series. By paying attention to the mentioned figure, it is evident that the indices of peaks in the residual and the original time series are almost identical. However, the amplitudes of the extremes are shrunk more than twice the size of the original peaks. The absolute value of the residual time series is calculated at this stage to prevent the loss of negatively sized peaks. This is because some extremes may have a delay of several days.

4.3.2 Correlation and distance matrices of concurrent floods

By implementing the mentioned method in section 3.2.2, simultaneous occurrences in the absolute residual time series are discerned and a new matrix of these extreme values is assembled. Subsequently, Kendall's Tau correlation of the residual extreme magnitudes is calculated between a couple of event sets. After that, every two sites had a specific correlation between the greatest contemporary cases. Figure 4.9 shows the relationship between different stations.

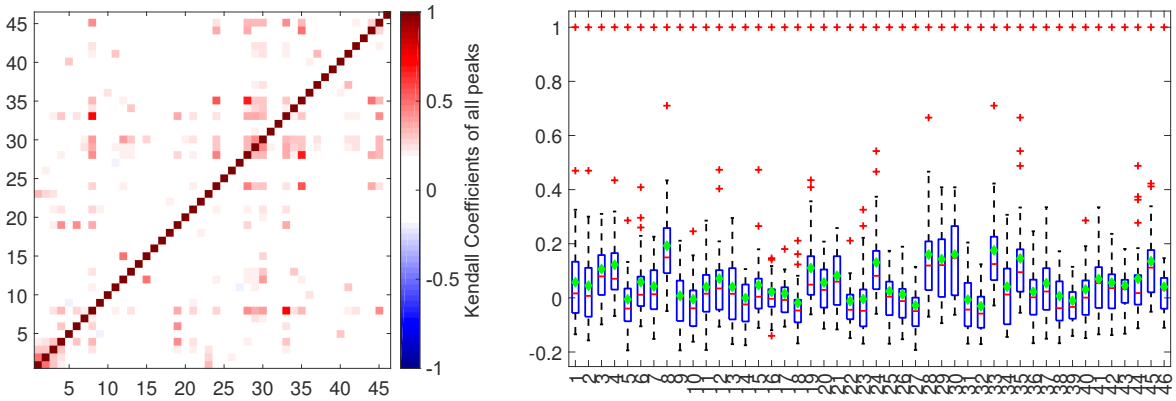


Figure 4.9: Rank correlation among all pair set of extremes

On the left side of Figure 4.9, the square rank correlation matrix is plotted, which shows a positive correlation among most stations. However, some of the small independent sub-catchments show a negative correlation. The dark red shows a stronger relationship between two points, whereas the pixels with near-zero values (white pixels) express no correlation. In the right part of Figure 4.9, the range of correlation changes is drawn. The red central mark indicates the median, the green lozenge is the mean, and the outliers are plotted using the '+' symbol. Large-scale catchments have a higher correlation with other catchments, while smaller catchments generally have a lower correlation. This figure shows that some

of the sub-catchments react differently. After PCA, the correlations among extremes are weaker than their correlation in the original coordinate system of the data set. It shows that the residual series, which has no underlying behavior of the total input dataset, has a low relationship with the farther subcatchments. The residual series is the part of the complete series, which primarily explains the extremes.

To transfer the non-tangible rank correlation plot to an explicit configuration, Figure 4.10 is mapped. Then, for each measurement gauge as a reference, the rank correlation is traced over the space. The meaning of applied colors here is the same as the previous figure.

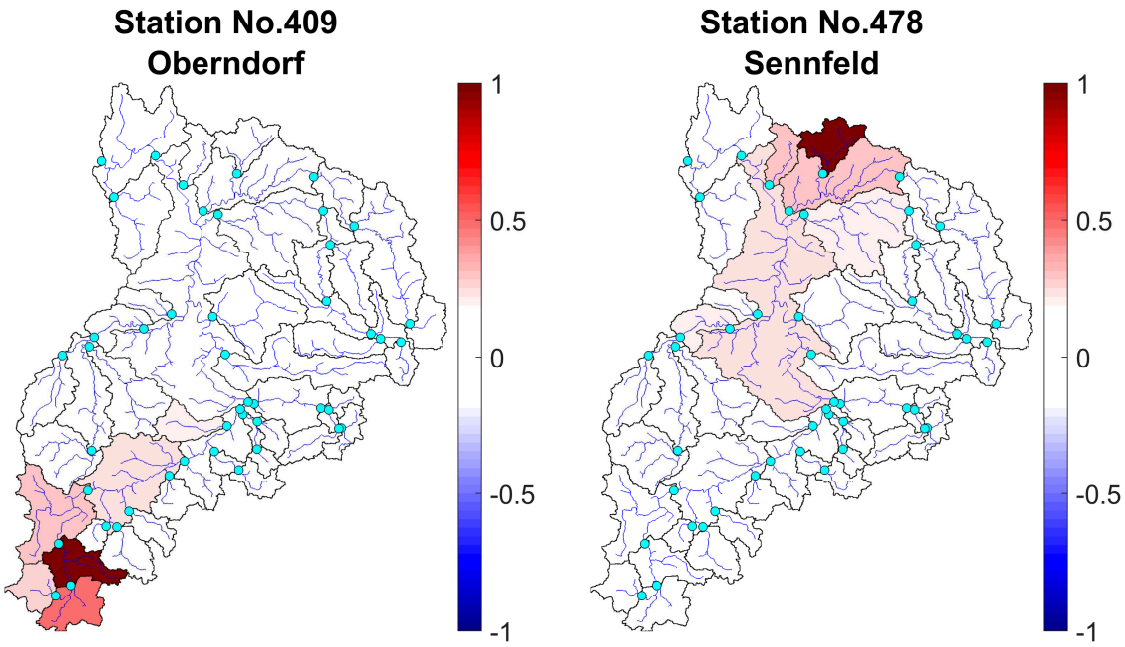


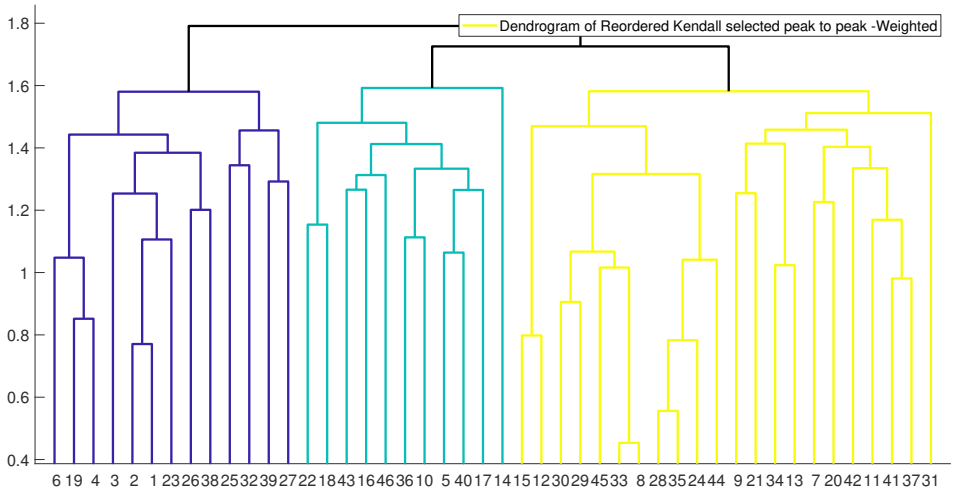
Figure 4.10: The simultaneous occurrence of the largest floods regarding the reference stations; Left: Oberndorf - Right: Sennfeld

Two sub-catchments are selected from two distinct parts of the region to show distinct spatial responses. In Figure 4.10, on the left, Oberndorf (ID-Nr. 409) as the upstream part is appointed; the darker red color presents a strong association with the reference. It shows that nearer and almost flow-connected sub-catchments react simultaneously and with the same power harmony in this part. Also, the station Sennfeld (ID-Nr. 478) in the north region of the Neckar has a different relationship with the other catchment. Thus, according to the number of catchments, there are distinct flood response regions. These two figures clearly show the purpose of performing PCA on data. In the figures mentioned earlier, only the areas exclusively related to the excess runoff are colored red. This means that the river's main flow has been removed from the time series after the application of PCA. Therefore, only values that are directly related to the flood have been analyzed. Hence, the behavior of the area is still almost vague. Accordingly, the hierarchical cluster analysis, by combining 46 dimensions, provides reasonable behavioral groups.

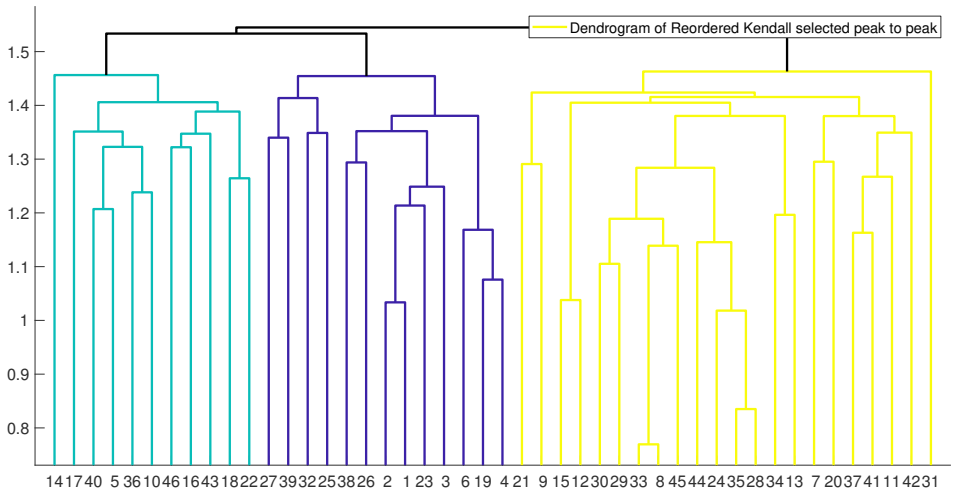
4.3.3 Cluster analyzing

4.3.3.1 Hierarchical tree and validation

Based on computed dissimilarity matrices, hierarchical cluster trees are calculated. This clustering designates the simultaneous behavior of extremes in different catchments. In Figure 4.11, two samples of the linkage method are plotted, which had the highest Silhouette value. The shape, structure and height of trees show distinguished groups.



(a) Weighted based cluster tree



(b) Average based cluster tree

Figure 4.11: Hierarchical cluster tree of the simultaneous occurrence of the extreme floods

The dendrogram presents similarities between the simultaneous peak discharge of the sampling sites that, noticeably, different groups can be distinguished from each other. Furthermore, the three main clusters in this figure are shown in three different colors. Inside each batch, some sub-clusters illustrate the interconnection of sub-basins in a specific area. The next step is to test the goodness of clustering to decide which hierarchical tree is more relevant to the aim of this research. As Figure 4.11 shows, it is not easy to recognize the similarities

between dendrograms, i.e., the class numbers and orders are not always the same. Therefore, it is demanded to apply different indices to compare the clustering results in this multidimensional space. Table 4.3 is expressed the quality of clustering using different distance metrics and linkage methods.

Table 4.3: The Silhouette coefficient of various linkage methods based on different distance matrices

	Complete	**Average**	**Weighted**	**Ward**	**Single**
Euclidean distance	0.173	0.297	0.302	0.297	-0.016
Kendall distance	0.280	0.302	0.297	0.297	-0.016

The results are presented in Table 4.3. In the hierarchical cluster tree based on Euclidean distance, the Weighted linkage method shows the highest score, and the Ward and Average methods are both in the second rank. Also, the Average linkage algorithm has the top score according to Kendall's distance, and the Ward and Weighted methods are in the following ranks. As it is clear, the Single linkage method is in offset for both distance methods. It is a sign to have several negative values. Therefore, it is not recommended to highlight this method for the following steps.

The class numbering is different for each of which stations. Now, the investigation is faced with a question. How much different dendrograms are similar to each other?. The ten possible combinations of the two applied distances and linkage methods for hierarchical clustering are calculated and reported in Tables 4.4 and 4.5. These results show that the similarity among clustering algorithms and make easy the interpretation of dendrogram trees, i.e., instead of plotting ten dendrograms to visually compare the similarity of members inside each cluster, here ARI and FM, as the two external indices to measure the similarity between two hierarchical clustering, count one by one the distribution of a member in a cluster in other clusters. The process is not sensitive to the number of clusters and the way of numbering each group of data. As a result, it is possible to find the most similar clusters by having the higher indices.

Table 4.4: Contingency table of two applied distances using ARI

	$Complete_K$	$Average_K$	$Weighted_K$	$Ward_K$	$Single_K$
$Complete_E$	0.167	0.212	0.202	0.202	0.056
$Average_E$	0.857	0.933	1	1	0.012
$Weighted_E$	0.796	1	0.933	0.933	0.045
$Ward_E$	0.857	0.933	1	1	0.011
$Single_E$	-0.001	0.045	0.012	0.011	1

The results of the adjusted Rand index in Table 4.4 show the highest similarity is between Ward linkage with Euclidean distance and Average linkage with Kendall distance. And after that for the Complete linkage method using Kendall distance and Average and Ward methods using Euclidean distance. Here, the Complete method with Euclidean distance shows a totally

different clustering result in comparison to the others. Naturally, the Single method is also acted differently from all the others and has identical clusters for both distance methods.

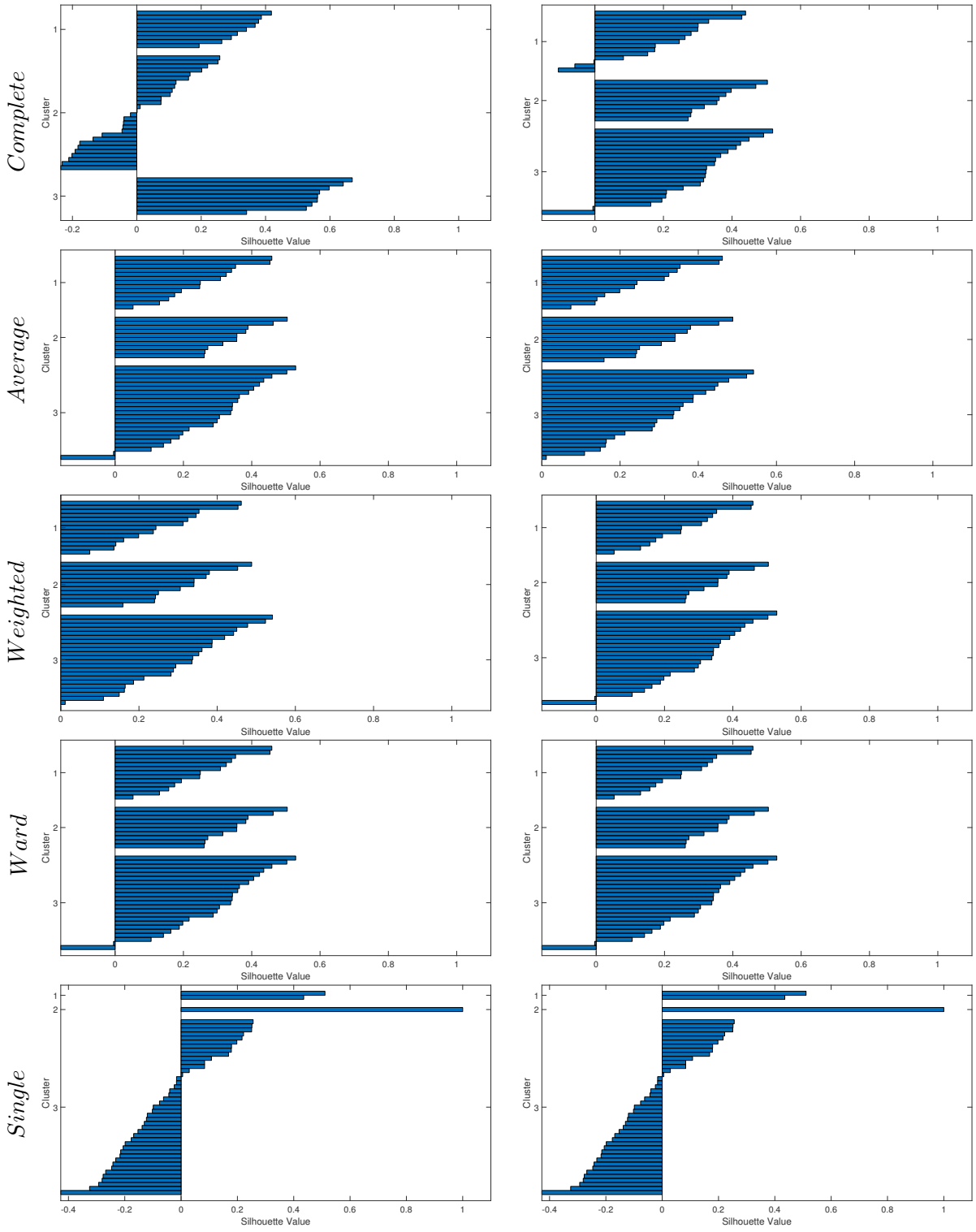
Table 4.5: Contingency table of two applied distances using FM

	$Complete_K$	$Average_K$	$Weighted_K$	$Ward_K$	$Single_K$
$Complete_E$	0.492	0.521	0.520	0.520	0.641
$Average_E$	0.901	0.957	1	1	0.569
$Weighted_E$	0.867	1	0.957	0.957	0.579
$Ward_E$	0.908	0.957	1	1	0.569
$Single_E$	0.550	0.579	0.569	0.569	1

The results of FM index in Table 4.5, express the same results from the adjusted Rand index, except the fact that the second most similar clustering is between Complete linkage with Kendall distance and Ward linkage with Euclidean distance. Here, the $Average_K$ and $Ward_E$ have almost identical results. Having one in the above tables is possible when the input is the obtained floods from the residual series, which does not include the primary behavior of the catchment.

In general, the results of the FM and ARI are mostly similar to each other. Due to the nature of the FM index, the comparison between the single linkage method and all the others showed a value equal to 0.6. The ARI coefficients for the single method are always near zero. The ARI correctly shows that the differences between the Complete linkage method and the rest. In the first row of the Tables 4.4 & 4.5, the difference between the mentioned two indices has appeared. The clustering results for the two Complete methods are not similar to each other. The ARI score for $Complete_E$ Vs $SingleK$ equals 0.056 and for $Complete_E$ Vs $CompleteK$ equals 0.167. Nevertheless, the FM scores are equal to 0.0492 and 0.641 for the mentioned comparisons, respectively. According to the Silhouette coefficient, the Single methods have many negatives and wrong clustering. Therefore, ARI showed better results than FM for the two methods of Complete and Single. For the others, they are mostly similar.

In general, after implementing PCA, clustering is less sensitive to the use of different distances and linkage methods. This is a prominent point for applying this method in future studies.



(a) Euclidean dist. (b) Kendall dist.

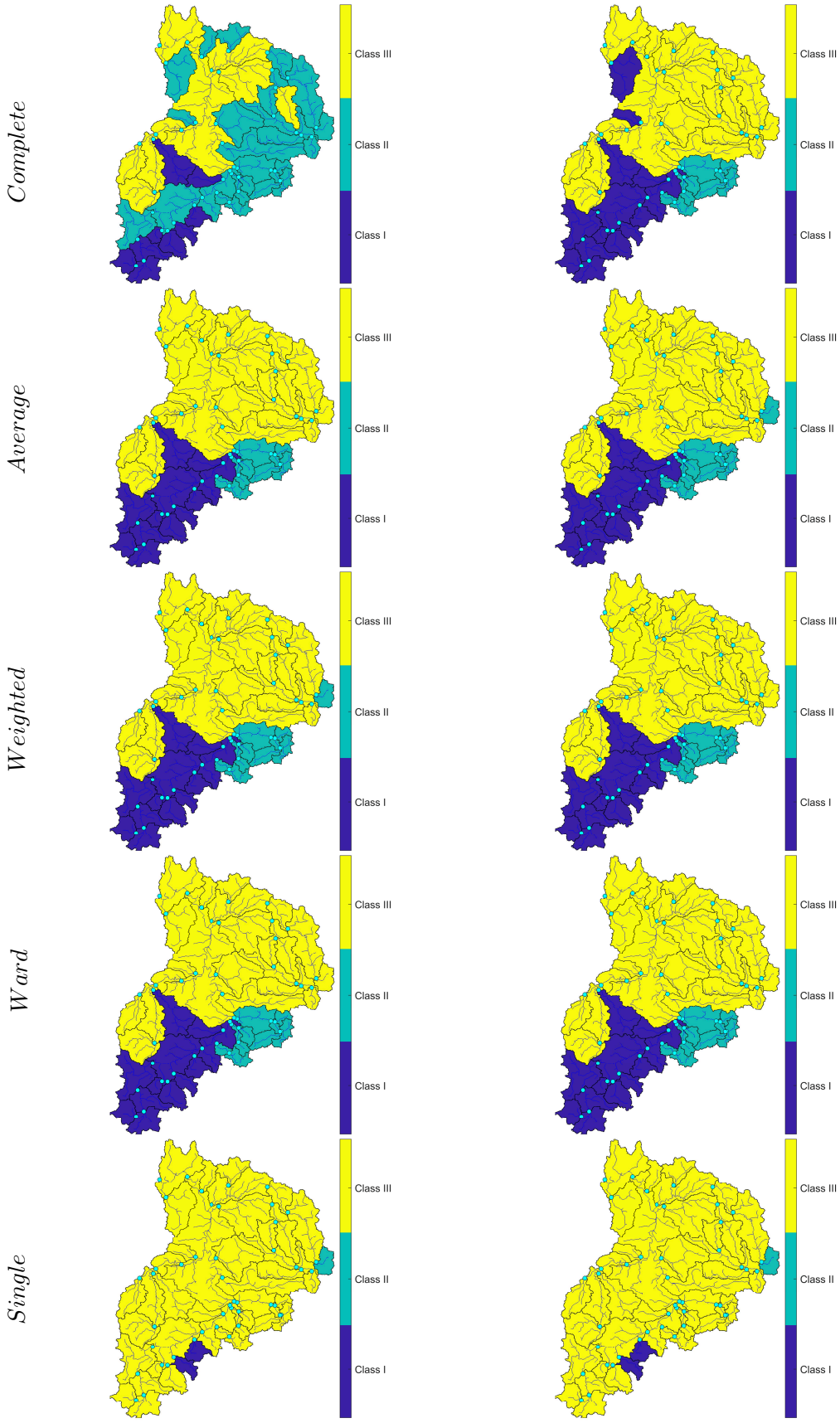
Figure 4.12: Silhouette coefficient of different hierarchical trees

By plotting Silhouette coefficient diagrams, the most reasonable linkage method can be cho-

sen. The outcomes of Figure 4.12, in addition to the information gained from Table 4.3, are determined as the most appropriate way to cluster simultaneous flood occurrences. According to Figure 4.12, the Complete, and Single methods represent many negative values, which shows an unreliable cluster tree. The negative Silhouette value is a sign of a wrong placement of a station in a given cluster. Therefore, concerning the above plot, the two best possible ways to cluster the sub-basin are chosen to be Average and Weighted linkage algorithms.

4.3.3.2 Mapping clusters

Regardless of the results of the verification coefficients, the clusters are mapped in Figure 4.13. These maps illustrate differences of employing different distance and linkage methods. However, they are mainly similar to each other and have some small differences except in the Complete method. Then, by comparing verification coefficient and the following maps, the best clustering map may determine. The first expectation is that, these maps should not be very different from the last chapter. Because the main behavior of the streamflow is captured by PCA and extracted from the time series. The identified extreme floods occurred at about the same time as the floods in the original time series. Nevertheless, the magnitude of the residual time series floods is lower than the primary series.



(a) Euclidean dist. (b) Kendall dist.

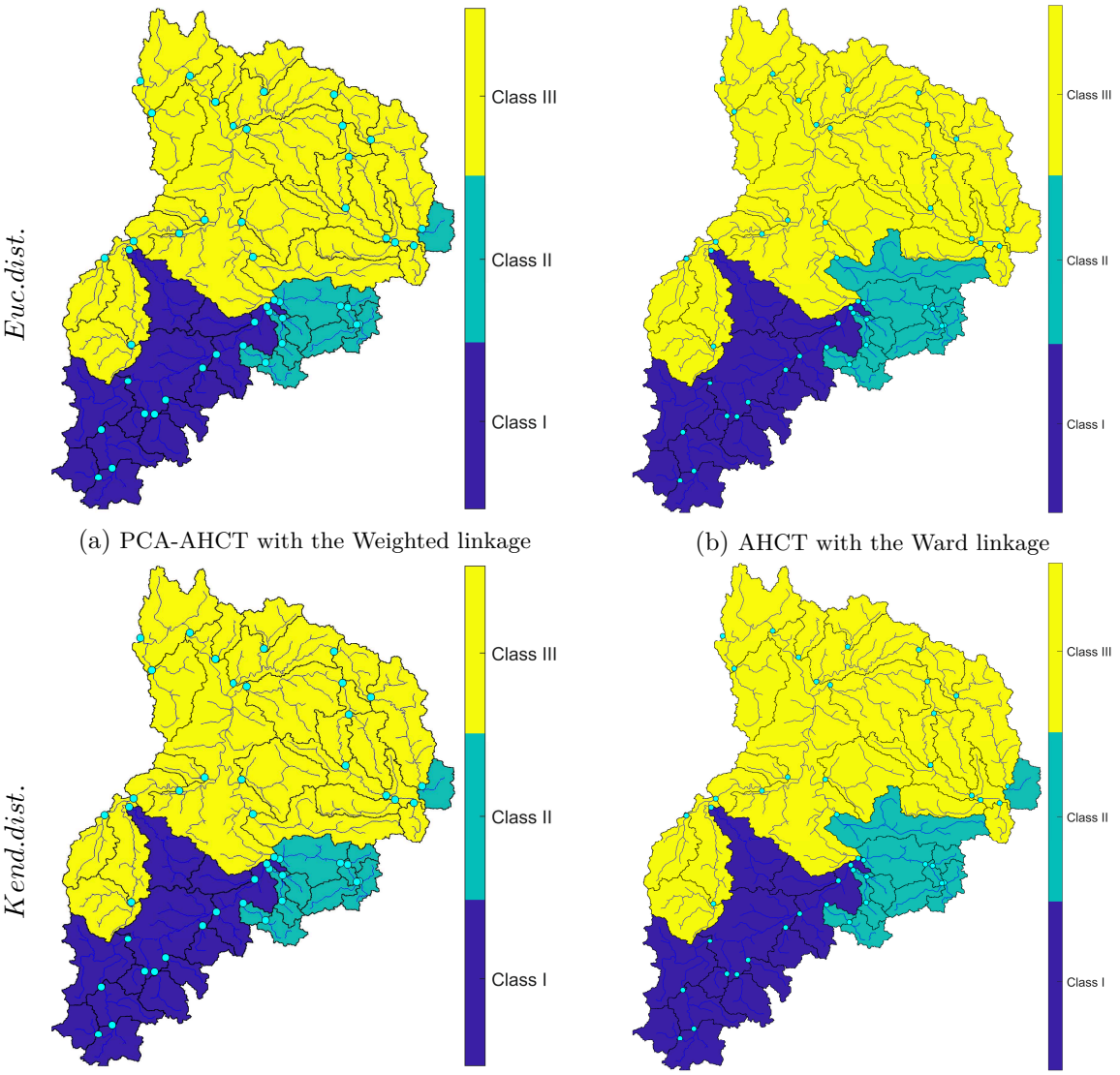
Figure 4.13: Spatial mapping of different hierarchical clustering in PCA-AHCT

Figure 4.13 shows that the application of each linkage method could make a diverse pattern. However, the most appropriate way to interpret clustering is selected concerning the research objectives and the verification result of the Silhouette value. The Silhouette coefficient shows the validity of each cluster. Maps with the four linkage methods (the Average, Weighted, Ward, and Single) are mainly the same in applying distance methods, except for one sub-catchment. However, the Single method is not successful according to the Silhouette examination. The ARI and FM also prove that the maps obtained from the Single linkage methods should not show the right way of clustering for floods. Also, the Complete algorithm illustrates massive differences in clustering by employing two distinct distance methods.

The maps for the Average, Weighted, and Ward linkage methods all follow the fact that the third cluster (Class III) starts from the outlet of the upper Neckar catchment in Plochingen (see Table 2.2. It can be because of the massive river regulation after this point on the Neckar river (please see Figure 2.8. Also, obviously, the three subcatchments in the west of the basin are separated from the other in the yellow color. They are Nagold and Neuenbürg as headwater subcatchments, Pforzheim on Enz river with (ID-Nr. 2452, 76123, 4422). In this region, there are not many dams and reservoirs (Figure 2.7a). The Würm river near the Pforzheim (ID-Nr. 36056) is always clustered as a part of class I in the dark blue, and it is a headwater subcatchment. It has two reasons; the first is that this part of the catchments is vividly separated from the rest in terms of the topographical division line around this sub-basin. Also, there are some dams and hydraulic structures, Which distinguishes this area from the western part of the catchment. Also, there is a sub-basin in the Average map with the defined Kendall distance at Schwasberg (ID-Nr. 1411) on the origin of the Jagst River, which is in class II. This headwater area has a high concentration of dams and reservoirs, which might be the reason for having it in the different clusters. In the upper Neckar basin, two areas are broken apart. One is on the main river Neckar and the Class II areas on the Fils river with Karsric geological feature.

4.3.4 Comparison between PCA-AHCT and AHCT clustering methods

In this part, the two chosen clusters based on PCA-AHCT are compared with the AHCT cluster maps presented in the preceding chapter and by (Modiri and Bárdossy, 2019b, 2021).



(a) PCA-AHCT with the Weighted linkage

(b) AHCT with the Ward linkage

(c) PCA-AHCT with the Average linkage

(d) AHCT with the Average linkage

Figure 4.14: Comparison between PCA-AHCT and AHCT

In Figure 4.14, three determined classes are plotted. (*a*) and (*c*) are the two selected algorithms in this section, while (*b*) and (*d*) are clustering methods without applying the PCA and reconstructed time series (normal AHCT). According to the Silhouette score presented in Table 4.3 and Figure 4.12, the Weighted method with Euclidean distance and the Average linkage with Kendall distance worked very well and showed well-defined clusters. Furthermore, the mentioned PCA-AHCT algorithm illustrates consistent results that are not altered by changing the distance matrix. The AHCT method showed mostly similar clusters in this region except for an area in the east part of Neckar in the second cluster (green color). This sub-catchment is a station in Schwabsberg (Nr. 1411). However, this area is located in the second cluster in all the other selected clusters in this plot.

The Rems river in the AHCT is clustered as class II, but after implementation of PCA on

the dataset, this area is clustered in class III, which has more meaning in terms of river regulation and geomorphology of the catchment. In general, the region is divided into the upstream and downstream, wherein the highland region, the catchment is shown two distinct clusters, one on the main river and another in the Fils river (in green color). The discharge time series after starting point of the third cluster at Plochingen (ID-Nr. 427) has the highest slope of changes, which can have some impacts on extreme as well (see Figure 2.12). Also, all the measurement gauges along the main river are always showed a significant trend in different time resolutions. However, floods are not a function of the leading regime of a river (see Figure 2.10). There is another reason that between Rockenau and Plochingen (stations 454 and 427), there are not any gauges. In between, some main tributaries of the Neckar (mentioned in Table 2.1) are connected to the main river. Therefore, it is reasonable to have the mentioned region as a separate cluster.

Therefore, results represent three major regions that have a well-defined simultaneous extreme flood mechanism. Despite a bit of difference in clustering by the AHCT method, the PCA-AHCT shows the same synchronic occurrences of flood events. The west part of the upper Neckar catchment plus Pforzheim are utterly separated from the other sub-catchments in both methods. The difference between these plots is within clusters two and three. Despite changing the distance algorithm, the clustered regions are mostly identical by using the mentioned linkage methods.

4.4 Conclusions

Simultaneous flood occurrences enhance the vulnerability in river plains. Additionally, extreme events analysis is not well understood in multidimensional space so far. High-resolution data and novel concepts for quantifying interactions undertake a primary step to achieve new results. Identifying typical clusters of vulnerable flood areas requires a classification of contemporary flood patterns, which spatially shows to what extent floods respond simultaneously. In this thesis, the flood classification objective function is formulated to find out contemporary flooding patterns. Different clustering approaches are compared to detect convergence patterns of the flood flow mechanism for magnitudes of coincidence floods. Its applicability is generally tied to flood protection, insurance, and river training.

Extreme discharge events of the river Neckar and its 46 sub-catchments are studied, and we introduced a new approach of analysis based on the PCA residual series. The PCA algorithm offers a pure series of discharges that are out of periodic term signals like the linear trend, annual and semi-annual signals. Therefore, the PCA method is firstly applied to acquire the primary behavior of the signals in the basin. After that, a new time series so-called residual is constructed based on the first three principal components. Finally, the leave one out cross-validation methods is applied to calculate the least predicted sum of squares to catch the optimal number of principal components to have a better reconstructed series. These three PCs explain more than 85% of input data's variability.

The clustering methods are used to estimate better multiple flood events evolution, which depends on the contribution of many sub-catchments. Also, clustering is a new metric in multivariate analysis of the simultaneous occurrence of floods. For this reason, PCA residual

and the hierarchical cluster tree are applied. Neither of these methods needs initial assumptions, and they are free of additional presumptions. The trees are computed based on rank correlation matrices of the highest occurrence of the absolute residual floods in the period of 55 years and are evaluated by various linkage ways. Mainly the Average, Weighted, and Ward methods have the same clusters. Then, the Silhouette score as a verification way of cluster tree is employed to all possible clusters. Therefore, the Neckar is grouped and mapped, and despite the small difference between the algorithms, an identical clustering pattern emerged. Then, these maps are compared to the outcome of the AHCT method. The result in the PCA-AHCT shows similar maps, but the AHCT has partly different behavior. Two critical stations are dubious in clustering with the AHCT method in the previous study despite verifying hierarchical cluster trees in the Neckar basin. This study determines that the station Schwabsberg (Nr. 1411) in Jagst tributary should be in the second cluster and station Neustadt (Nr. 1470) in Rems' tributary should be in the thirds cluster to have a higher evaluation coefficient.

The ARI and FM indices also show the similarity between different clustering maps; however, the FM measure has reported unfavorable properties for small numbers of clusters. On the other hand, the index can be very high for independent clusterings. In this research study, the ARI contingency table shows more reliable results. However, the results of both methods show similar outcomes.

The results show the simultaneous occurrences of high discharges operating as a function of the basin's topology and geology. To conclude, the Neckar is divided into three major clusters: the first is around the western part of the upper Neckar catchment and seized with the Black Forest and Swabian Alps. The second cluster is primarily located in the eastern region of the upper Neckar, which includes smaller sub-catchments and karstic geologic features. These two clusters are in the farthest Euclidean distance to the catchment outlet as well (see fig. 2.4a). Finally, the third part is the remaining area of the Neckar basin lowland area. It can be mentioned that a reason for some clustering mismatches between PCA-AHCT and AHCT might be due to the anthropogenic alterations in this area or the low magnitude discharge flow in the small catchments such as the Pforzheim. Besides, the observed difference among clusters in the southeast of the Neckar basin can be due to the geological characteristics of a dipping formation.

The main Neckar river is massively regulated after the Plochingen (Nr. 427), and it can be a reason to have a distinct cluster after this point, which is separated the Neckar into two regions: the upper Neckar and the rest in downstream regions. Also, the number of dams and reservoirs play a significant role in having the western subcatchments of the Neckar into the first cluster (class I) in the dark blue (please see Figure 2.7a). Also, maybe due to the high concentration of hydraulic structures on the Rems river, it is possible to have this area (Nr. 1470) in different clusters. The PCA-AHCT advantage is highlighted, especially in this region when the Rems is clustered as class III. As a result, being headwater subcatchment does not play a vital role in the simultaneous flood clustering. There are several headwater areas in the third cluster near the outlet and near the border of the Neckar. However, indirectly, the upper Neckar with a high concentration of small-sized sub-basin and many headwater catchments react differently from the study area's downstream.

Furthermore, the lack of a discharge flow measuring station along the main river Neckar between Plochingen and Rockenau, as well as the connection of four tributaries, Rems, Enz, Kocher, and Jagst, to the main flow path in between, might cause a difference between the upstream and downstream in the resulted clusters in both chapter three and four.

In total, clustering is less sensitive to employing different distances and linkage methods after implementing PCA. Therefore, it can be concluded that this is an efficient approach that can be utilized in further clustering studies on extremes in hydrology.

5 Distribution-based clustering of general floods

5.1 Introduction

One of the most prominent challenges in hydrology is the reliable estimation of extreme floods, which are crucial for designing and operating flood control structures and flood defenses. Regional flood frequency analysis (RFFA) is a general approach for estimating the magnitude of floods with various return periods in homogeneous regions (England Jr et al., 2019). Let assume floods are stochastic processes, and their magnitudes and frequencies can be predicted using certain probability distributions. RFFA provides flood quantile estimates, which are then used as the basis for designing flood defense infrastructure. The regionalization process's goal is to identify catchments that are similar enough to corroborate the combination of extreme flow information on all sites in the region. Also, RFFA tries to reduce sampling uncertainties in calculations of extreme flood events (Leščešen and Dolinaj, 2019). Thus, it is necessary to combine streamflow data records from different sites in a geographical region with similar flood characteristics. Similar regions can provide the basis for the computation of extreme flood events at ungauged basins within the same geographical region.

This chapter aims to detect whether the region can be considered hydrologically homogeneous and determine the distribution function that best fits the observed extreme floods (Cunnane, 1988). There are dozen of RFFA methods such as station year methods (Rosbjerg and Madsen, 1995), Dalrymple's method (Dalrymple, 1960), Methods based on dimensionless moments (Stedinger, 1983; Saf, 2009) Methods based on order statistics and record extension (Kajambeu et al., 2020). As well as Bayesian methods (Reis Jr and Stedinger, 2005), and methods based on standardised probability-weighted moments (Greis and Wood, 1981; Ahmad et al., 1988; Deng and Pandey, 2009). Moreover, several different methods, which are not in the context of this study. Most RFFA methods assume that some type of regional homogeneity of flood frequency behavior exists, except the Bayesian approach and record extension methods. The distribution and probability of exceedance floods are the basis of each flood frequency analysis. Therefore, it is essential to investigate the regions with similar flood behavior.

Unlike the last two chapters, clustering similar floods as the term "general floods" in the Neckar catchment is investigated to determine possible accompaniment among floods. Thus, in this chapter, the main concentration is not on the simultaneous occurrences of floods. Instead, the largest floods are identified, regardless of their synchronicity. However, some simultaneous floods may be categorized as general floods, while many floods occurred independently.

Many existing clustering methods are assumed the linearity of the series. However, linear hypotheses can often not be maintained (Zhang and Chen, 2018). In this chapter, the problem of clustering non-linear time series and distribution-based clustering are considered. We propose implementing a two-sample Kolmogorov-Smirnov test (KS) statistic as a distance measurement of two series by measuring the discriminatory power of a reference series and the affinity of non-linear serial dependence structures. This statistic allows unsupervised clustering on sets of distributions and detects the clusters in those sets.

In this part, the Cumulative Distribution Function (CDF) is the basis of clustering to group similar CDFs into sets as realized clusters. The smaller the difference between two CDFs, the more likely that two sets of data are in the same cluster. Furthermore, the farther apart the two CDFs are, they can be classified into different clusters. This test can determine differences in position and dispersion of CDFs (Mora-López and Mora, 2015). This type of clustering is done in a multidimensional space by a hierarchical clustering tree without considering any assumptions (Pauwels and Frederix, 2000; Modiri and Bárdossy, 2020). So far, KS statistics applied in combination with K-means and K-medoid techniques (Wang et al., 2019; Zhu et al., 2021) and have not been applied in hydrology. Therefore, the CDF-based clustering on general floods is the investigated approach in Chapter 5.

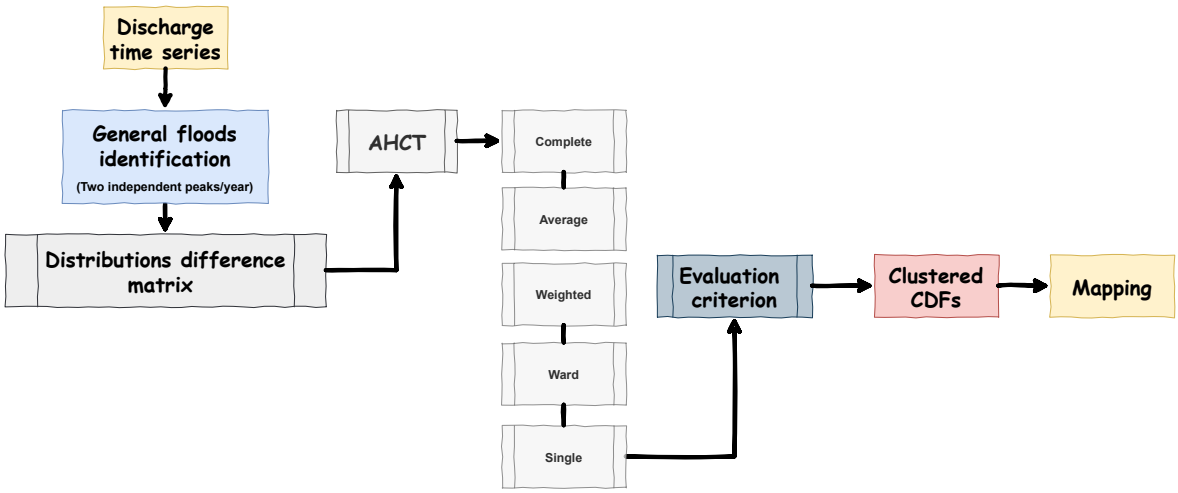


Figure 5.1: The applied algorithms in the distribution-based clustering

In Figure 5.1, the stepwise procedure of this chapter is illustrated to clarify the structural outline. First, the general floods are identified so that we have two floods every year. This helps to maintain continuity over time in a way that we still have a flood time series. Then, empirical cumulative distribution functions are computed and plotted to calculate the dissimilarity matrix of the flood series. Finally, the AHCT clustering is done by implementing five forms of linkage connections, and different criteria to evaluate the results. Further, the return period is calculated as the basis for flood frequency analysis for the catchment outlet, and the Flow duration curve (FDC)s are plotted for all areas. In the end, the obtained results are illustrated as maps. Then, the resulting clusters are interpreted in terms of high and low flood occurrences.

5.2 Methodology

5.2.1 General floods identification

The general floods in this section are identified based on Section 3.2.1. Thus, two large independent floods per year are selected as general floods in the Neckar region. In total, concerning 55 years of data, each station has 110 peaks as a time series of extreme floods. Regarding the coordinates of each sub-catchment and locating the stations upstream or downstream, the magnitudes of floods are not comparable. The selected peak matrix is normalized to have the same scale data properties such as skewness and kurtosis. This process lets the comparison of two or more data sets with different magnitudes and units.

For a random variable X with μ and σ as the mean and standard deviation, the z coefficient of a value x is:

$$z = \frac{(x-\mu)}{\sigma} \tag{5.1}$$

For sample data with mean $\bar{X}$ and standard deviation S, the z coefficient of a data point x is:

$$z = \frac{(x-\bar{X})}{S} \tag{5.2}$$

z scores measure the distance of a data point from the mean in terms of the standard deviation. This is also called standardization of data. The standardized data set has a mean of 0 and a standard deviation of 1, keeping the original data's shape properties.

5.2.2 Return period estimation

A recurrence interval, also known as a return period, is a statistical measurement typically based on historical data over an extended period. It is the first step of each flood frequency analysis. The following analysis assumes that the probability of the event occurring does not vary over time and is independent of past events (Kang et al., 2019). There are different way to estimate the return period; here, the inverse Weibull equation is used as follow (Selaman et al., 2007):

$$T_r = \frac{1+n}{m}, \tag{5.3}$$

where m is the rank of flood series and n is the number of years in the dataset. There are two approaches to estimate the return period of the flood series. The first one is the annual maximum series, which is the traditional way to calculate the return period. The second method utilizes the partial duration series when we have more information about the peaks inside the input time series. Depending on the return period approaches, the m factor can be changed. It means, by adding other peaks, first the floods have to be sorted, and then the rank of each element will be calculated.

There are several methods to calculate the average of recurrence exceedance interval (T_r) explained in Selaman et al. (2007); Zhang et al. (2015). Nevertheless, the concentration of the current research is not on this issue.

5.2.3 Flow duration curve

The FDC is a stochastic representation of the variability of runoff, which arises from the rainfall and snowmelt transformation all over a given catchment area (Yokoo and Sivapalan, 2011). It is a representation of the frequency distribution of streamflow at a specific time. The temporal resolution can be very fine to a coarse-scale (Searcy, 1959; Vogel and Fennessey, 1995). Also, FDC can play an alternative delegation of the CDF of discharge flows.

A flow duration curve represents the association between the frequency and magnitude of a time interval streamflow for a particular river basin, which provides an estimate of the percentage of time a specific flow was exceeded over the historical time series (Vogel and Fennessey, 1994). An FDC is the complement of the cumulative distribution function of discharge. Each value of discharge flow (Q) has a corresponding exceedance probability (p). The FDC is simply a layout of Q_p, the pth quantile or percentile of given resolution streamflow against exceedance probability p, where p is defined as follow:

$$\begin{aligned} p &= 1 - P\{Q \leq q\} \\ &= 1 - F_Q(q) \end{aligned} \tag{5.4}$$

The quantile Q_P is a function of the measured streamflows. Due to the dependency of this function to the empirical observations, it is often entitled the empirical quantile function (to have detailed information of mathematical equations, please see Vogel and Fennessey (1994)).

5.2.4 Non-parametric distribution similarity

In this chapter, the dissimilarity among empirical CDFs is assessed to group floods with a close distance between their distributions. The Kolmogorov–Smirnov test (KS test) is a nonparametric test of the equality of distributions that can be used to compare two samples. Here, the selected peaks for each discharge measurement station are a sample vector. Thus, the two-sample KS test evaluates the difference between the CDFs of the two sample data vectors over the range of x per data set. This test quantifies a distance between the empirical distribution of two samples.

The cumulative distribution function $F(x)$ of a random variable X, is:

$$F(x) = \mathbb{P}(X \leq x), \qquad x \in \mathbb{R} \tag{5.5}$$

The cumulative distribution function distinctively characterizes a probability distribution. Given observations $x_1, \ldots, x_n$, the empirical distribution function $F_{\text{obs}}(x)$ gives the proportion of the data that lies below x, i.e. if the ordered observations $y_1 \leq y_2 \leq \cdots \leq y_n$, then:

$$F_{\text{obs}}(y_i) = \frac{i}{n} \tag{5.6}$$

The null hypothesis that data in vectors X and Y comes from two independent random samples from continuous distributions (populations with the same distribution) with distribution functions F and G, respectively. Subsequently, the difference among CDFs has to be tested.

Thus, the KS two-sample can be applied to test the null hypothesis,

$$\mathrm{H_0 : F(x) = G(x),\ for\ every\ x,}$$

against the general alternative

$$\mathrm{H_1 : F(x) \neq G(x),\ for\ at\ least\ one\ x,}$$

which utilizes the test statistic as follows (Massey Jr, 1951):

$$D_{n,m} = \sup_{-\infty<x<\infty} |F_n(x) - G_m(x)| = \sup_{1\leq i\leq n+m} \left|F_n\left(z_{(i)}\right) - G_m\left(z_{(i)}\right)\right| \tag{5.7}$$

where F_n and G_m denote the empirical distribution functions for the X, and Y samples, respectively. The $Z_{(i)}$ are the order statistics of the combined mentioned samples. In general, the two-sample KS test uses the maximum absolute difference between the CDFs of the distributions of the two data vectors. This test is conventional and free over the class of all continuous distribution functions and consistent against any differences between F and G; however, it is biased (Katzenbeisser and Hackl, 1986). The decision to reject the null hypothesis in KS test is based on comparing the p-value with the significance level (α), not by comparing the test statistic with a critical value. The null hypothesis is rejected at level (α) for large samples, if

$$D_{n,m} > c(\alpha)\sqrt{\frac{n+m}{n\cdot m}} \tag{5.8}$$

where n and m are the sizes of first and second sample respectively, and $c(\alpha) = \sqrt{-\ln(\frac{\alpha}{2})\cdot\frac{1}{2}}$.

Moreover, there are some other classes of test statistics, which employ information from all and not only the largest deviation introduced by Lehmann (1951); Rosenblatt (1952); Kuiper (1960). Also, if the two samples have different sizes, it is possible to use another nonparametric two-sample test such as the introduced method in Gretton et al. (2012).

As a result, clustering can be done by accepting or rejecting the null hypothesis and asymptotic probability.

5.2.5 Agglomerative hierarchical cluster tree based on, KS statistics

To construct clusters, the KS statistics is the basis for clustering in this part of the thesis. This coefficient shows how near/far are two datasets' vectors. Therefore, the square matrix of KS statistics has to be calculated to show the maximum differences among CDFs. This matrix is assumed to be a sort of CDF-based distance matrix employed in cluster tree computations. The mentioned steps are similar to Section 3.2.4.

5.2.6 Clustering performance

To evaluate the clustering performance, three clustering evaluation methods are applied to evaluate different clustering efficiency approaches. Each of these criteria follows distinct mathematical concepts and can have different judgments. First, for the goodness of clustering, the Silhouette coefficient algorithm is performed as it is described in Section 3.2.5.

5.2.6.1 Calinski-Harabasz index

The Variance Ratio Criterion (VRC), so-called the Calinski-Harabasz (CH) index is defined as follows (Caliński and Harabasz, 1974):

$$CH_k = \frac{SS_B}{SS_W} \times \frac{(N-k)}{(k-1)}, \tag{5.9}$$

where SS_{B} is the total between-cluster dispersion, and SS_{w} is the overall inter-cluster dispersion, k is the number of clusters, and N is the number of observations.

The total between-cluster dispersion SS_B is defined as:

$$\begin{aligned} SS_B &= \sum_{i=1}^{k} n_i \left\| m_i - m \right\|^2 \\ &= \sum_{i=1}^{k} n_i (m_i - m)(m_i - m)^T \end{aligned} \tag{5.10}$$

where n_i is the number of observations in cluster i, m_i is the centroid of cluster i, m is the overall average of the sample data, and $\|m_i - m\|$ is the L^2 norm between the two vectors. The L^2 calculates the distance of the vector coordinate from the origin of the vector space. It is also called as the Euclidean norm as it is calculated the Euclidean distance from the origin.

The overall within-cluster dispersion SS_{w} is defined as follows:

$$\begin{aligned} SS_W &= \sum_{i=1}^{k} \sum_{x \in c_i} \left\| x - m_i \right\|^2 \\ &= \sum_{i=1}^{k} \sum_{x \in c_i} (x - m_i)(x - m_i)^T \end{aligned} \tag{5.11}$$

where x is a data point, c_i is the ith cluster, m_i is the centroid of cluster i, the same as the last equation, and $\|x - m_i\|$ is the L^2 norm between the two vectors.

The larger the CH_k ratio, the better the data division. Therefore, the best possible clusters have a large between-cluster variance (SS_B) and a small within-cluster variance (SS_W). As an advantage, this score is fast to compute and consider both within and between cluster variances (Łukasik et al., 2016; Kingrani et al., 2018)

5.2.6.2 Davies-Bouldin index

The Davies-Bouldin (DB) index is based on a ratio of within-cluster and between-cluster distances. Davies-Bouldin Index is defined as follows (Davies and Bouldin, 1979):

$$DB = \frac{1}{k} \sum_{i=1}^{k} \max_{j \neq i} \{ D_{i,j} \}, \tag{5.12}$$

where k is the number of clusters, and $D_{i,j}$ is the within-to-between cluster distance ratio for the ith and jth clusters.

$$D_{i,j} = \frac{\left(\bar{d}_i + \bar{d}_j\right)}{d_{i,j}} \tag{5.13}$$

$\bar{d}_i$ and $\bar{d}_j$ are the average distance between each point in the ith and the jth clusters and the centroid of their cluster, also known as "cluster diameter". $d_{i,j}$ is the Euclidean distance between the centroids of the ith and jth clusters. A lower Davies-Bouldin index is associated with a model with better separation between the clusters. The difference between CH and DB indices applies differently within/between -cluster variances and distances (Ünlü and Xanthopoulos, 2019). The computation of the Davies-Bouldin index is much less complex than the computation of the Silhouette index, which is a significant advantage regarding eventual real-time operation (Petrovic, 2006).

5.3 Results

In this part, the extreme floods' return period is estimated to determine the possible differences between the annual and partial approaches. The flow duration curves for all the measurement gauges for the daily discharge time series are plotted and computed separately for the flood series. Then, the Kolmogorov-Smirnov test's statistics are calculated, and empirical CDF of extreme floods is plotted. Later the polar form of the hierarchical cluster tree is illustrated to determine the final clusters. Finally, clustered stations are shown to present their differences in the empirical cumulative distribution functions. In addition, their differences are illustrated in the FDC form. The resulting clusters are also mapped in space to recognize a better group of stations. The chapter ends with a conclusion.

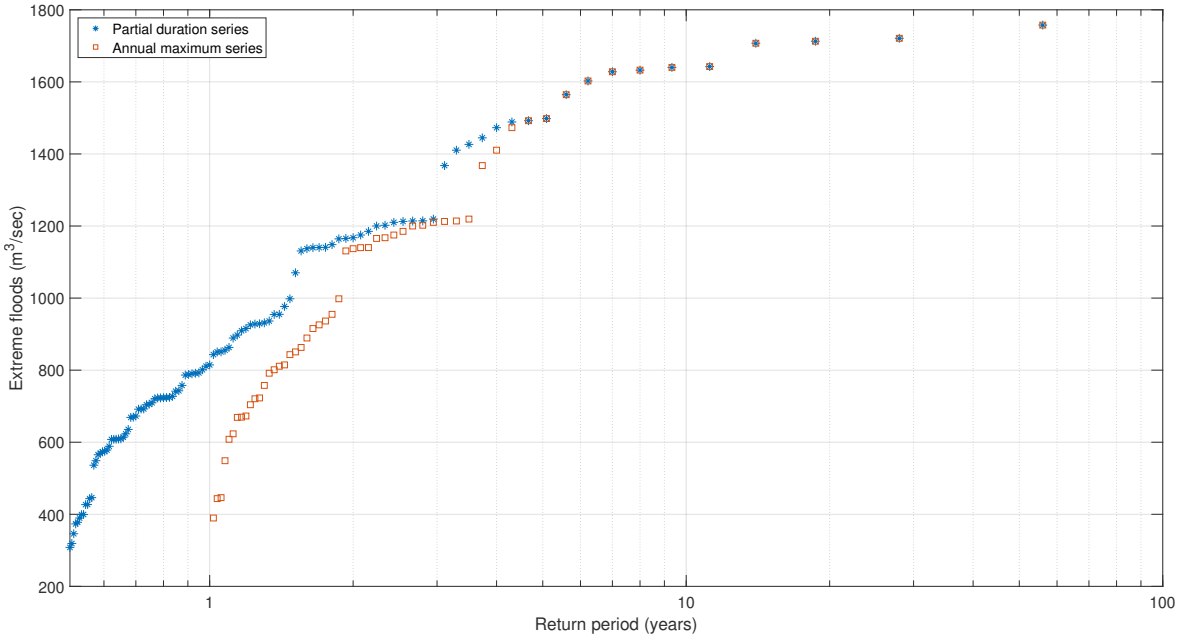
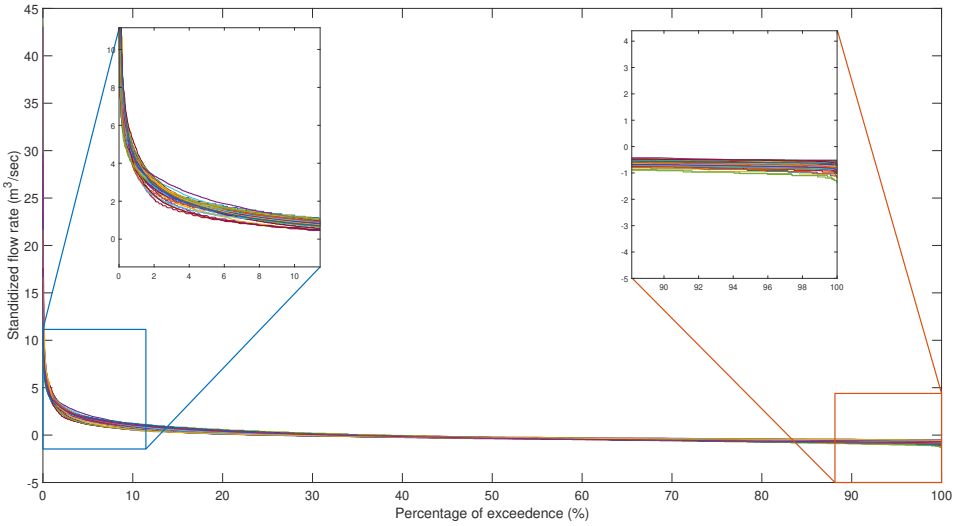


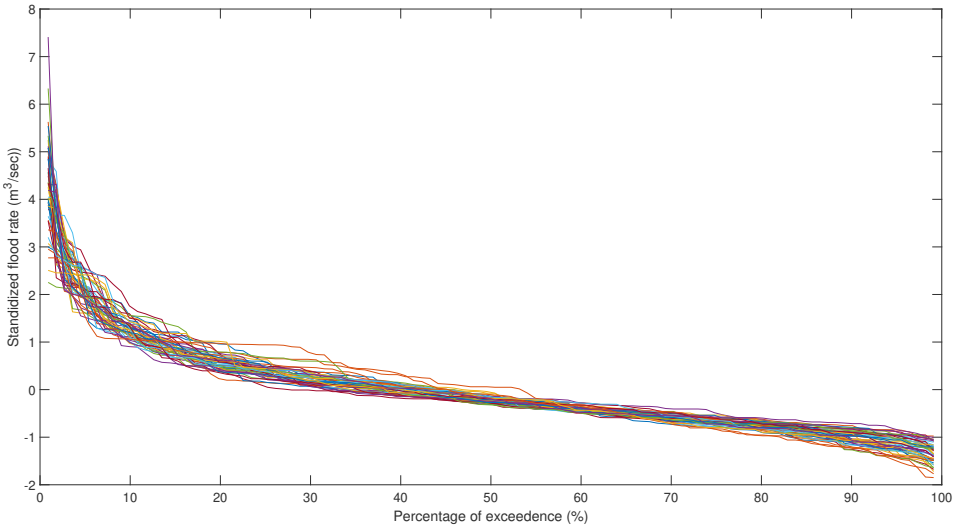
Figure 5.2: The return period of extreme floods based on two approaches of annual maximum and partial duration series at the outlet of the Neckar basin

In Figure 5.2, the partial duration series displayed as extreme flow versus the return period with blue stars for the extremes at the outlet of the study area. The x-axis is shown with logarithmic scales. Also, the annual maximum series is plotted in red squares for comparison between these two approaches. Note that the partial duration series has many more data points and quite a few with less than a one-year return period. It shows that these peak flows are occurring with a greater frequency than once per year. Further, the annual and partial duration series tend to diverge for return periods below five years.

The range of changes between the highest extreme floods and the lowest is massive and around 1450 m^3/sec. Therefore, it is possible to have different results for high and low volume floods.



(a) Flow duration curve for standardized the discharge flows



(b) Flow duration curve for standardized the extreme floods

Figure 5.3: Empirical flow duration curve for the discharge and biggest floods time series

Figure 5.3 shows the FDC for all the measurement gauges in the Neckar catchments. In general, there are two parts of the figure, which are distinguished the difference in FDCs,

in Figure 5.3a. First, the exceedance probability between 1 to 15% and then the last 15% of the x-axis for the exceedance probability. The FCDs for the standardized discharge flows are almost the same in between 35 to 75% of the probability. It shows that massive and low volume discharges are the distinguishing factor of the flow. All the curves have a steep slope, which intimates a highly variable stream whose flow is considerably due to the quick runoff of rainfall to the stream. Figure 5.3b, only illustrates the FDCs for the standardized extreme floods. Here, the difference is higher than the discharge time series. However, between 55 to 70 percent of the time, flood in the streams is likely to equal or exceed similar volumes. The highest difference belongs to less than 25% of exceedance of floods.

5.3.1 KS test decisions and statistics

Here, the acceptation and rejection of the null hypothesis of the KS test are investigated to determine the similarity of distribution type among empirical CDFs. Then, the absolute maximum difference between CDFs is calculated and shown.

Figure 5.4 shows an overview of the similarity of the distribution of the extreme events. In part 5.4c, the differences among CDFs are partly distinctive, especially in between 45% and 80%. Also, the differences among empirical cumulative distributions are high in the first and the last percent of the figure; however, it is difficult to catch this point easily by checking a single figure. Due to having more information in the KS test statistics 5.4b than the acceptation/rejection decision in 5.4a, the results of the KS statistics are more concentrated in this chapter. The KS statistics show the maximum distance between empirical CDF of extreme floods, and based on this factor, clustered maps are schemed.

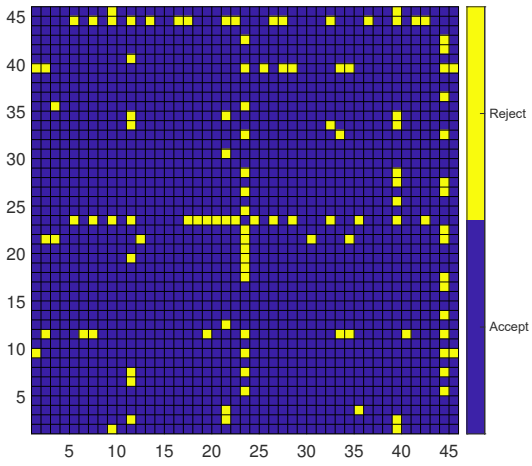
5.3.2 Clustering of the general floods

5.3.2.1 Polar hierarchical cluster

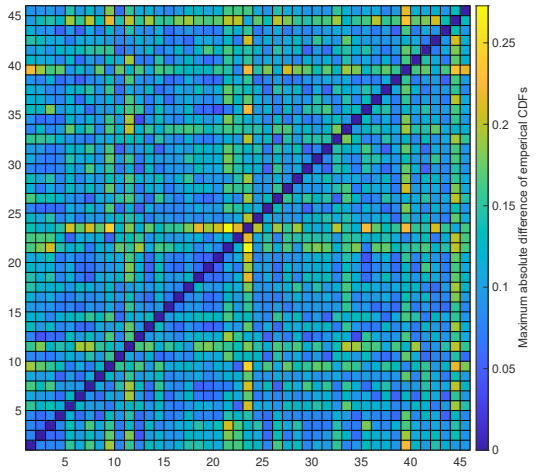
The polar form of AHCT is offered to present better performance for understanding cluster trees. The agglomerative hierarchical cluster tree is computed and plotted in Figure 5.5. The figure is divided into two parts; the upper part is for the KS statistics for Ward on the left and the Average on the right. Different linkage methods present partly different clusters. Also, here, clustering with KS statistics has better efficiency than the KS test decision, due to the binary nature of the decision matrix and the existence of more information in the KS statistics matrix. In this figure, station number 23 on the Eschach river and 44 in Neunbürg on the Enz river, usually reacted similarly and in a group. It is easy to determine two-by-two relationships between stations, but we still need some more interpretation.

5.3.2.2 Similarity of the general floods distributions, based on KS statistics

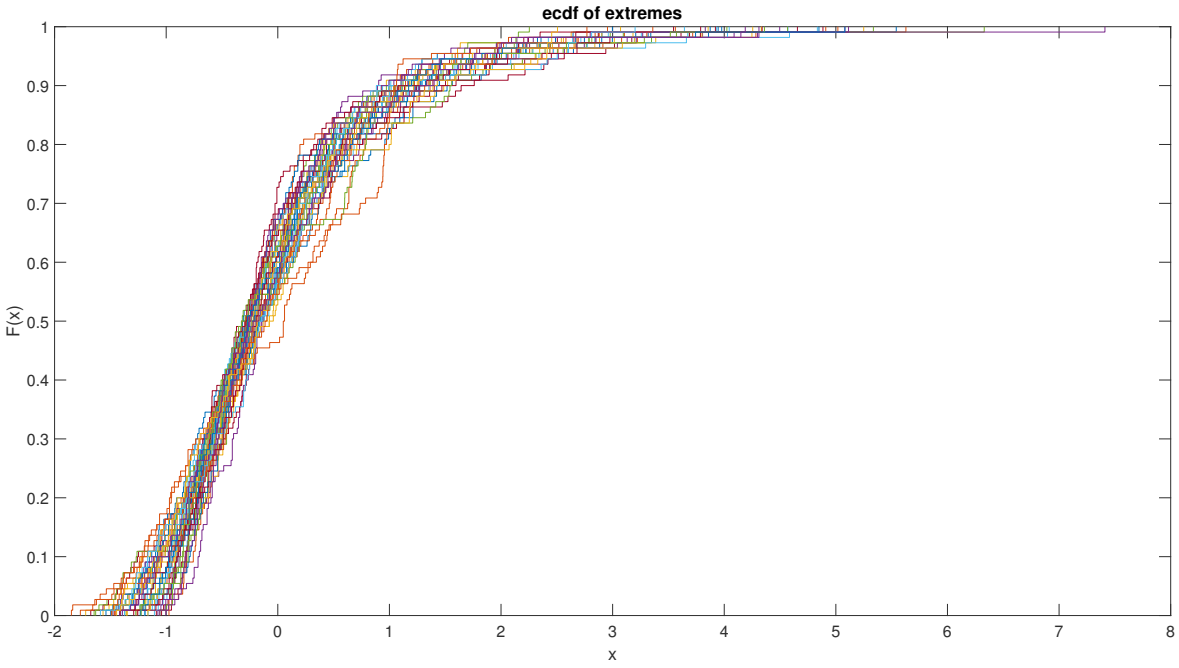
Figure 5.6, shows possible grouping ways in hierarchical clustering in CDF form. In Figure 5.6a, Class II is in between Class I and III. The blue and yellow colored clusters are separated from the beginning of the CDF and change their reaction around 0.4. One of the dark blue CDFs is acted differently after 0.82. It is a sign that, this station can be in a different cluster for high volume floods. More consistency in the orientation of CDFs is visible in Figure 5.6b. The two green and yellow colored clusters are started at almost similar x-axis points, around



(a) KS test decision for the null hypothesis



(b) KS test statistics



(c) Empirical CDFs of standardized general floods

Figure 5.4: In (a), the yellow pixels are the duo stations that are not from the same continuous distribution. In (b), the maximum absolute differences among CDFs are shown in detail. The empirical CDFs of the general floods for all stations are illustrated over the Neckar catchment, in (c).

0.3, and they shifted up and down. Between 0.65 to 0.8, they have the maximum distance with the dark blue clusters. In Figure 5.6c, despite almost perfect separation in the low values less than 0.1, clustering does not illustrate reasonable groups. It means the disagreement is high in Class III with yellow color. The Ward clustering in Figure 5.6d is relatively better than the previous clusters due to the stable separation of CDFs for the floods less than 0.8. In Figure 5.6e, the grouping is reasonable for the yellow color, but the dark blue cannot be

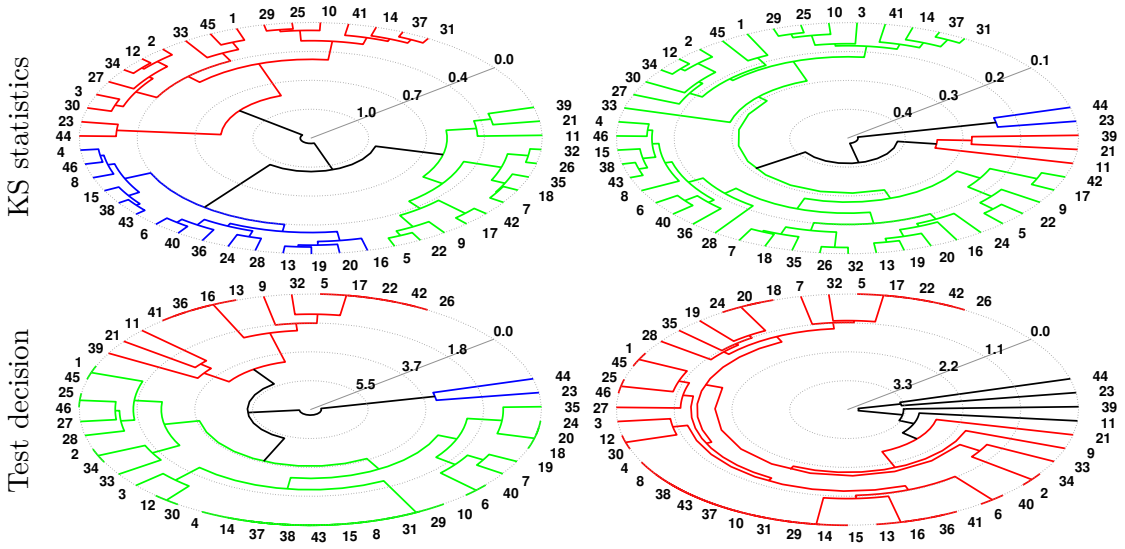


Figure 5.5: Polar AHCT Comparison between KS decision and Statistics for Ward and Average linkage methods

in a separate group for less than 0.2.

In Figure 5.7, the high exceedance probability floods are distinctively clusters in all parts. It means low volume floods in between 80 to 100% had completely different behaviors. In all subplots, some headwater subcatchments are shown illustrious flow duration curves up to 50% percentage of exceedance. For the extreme floods in the window of 0 to 25 %, there is not any pattern in FDCs. Here, all the lines are twisted into each other. Especially in the Ward method in Figure 5.7d, it is impossible to catch any meaning for this window. However, class II has higher standardized floods than class I and III in between 70 to 100% probability. Therefore, it is possible to cluster low volume floods spatially, but the extreme floods did not follow any layout in terms of flow duration curves.

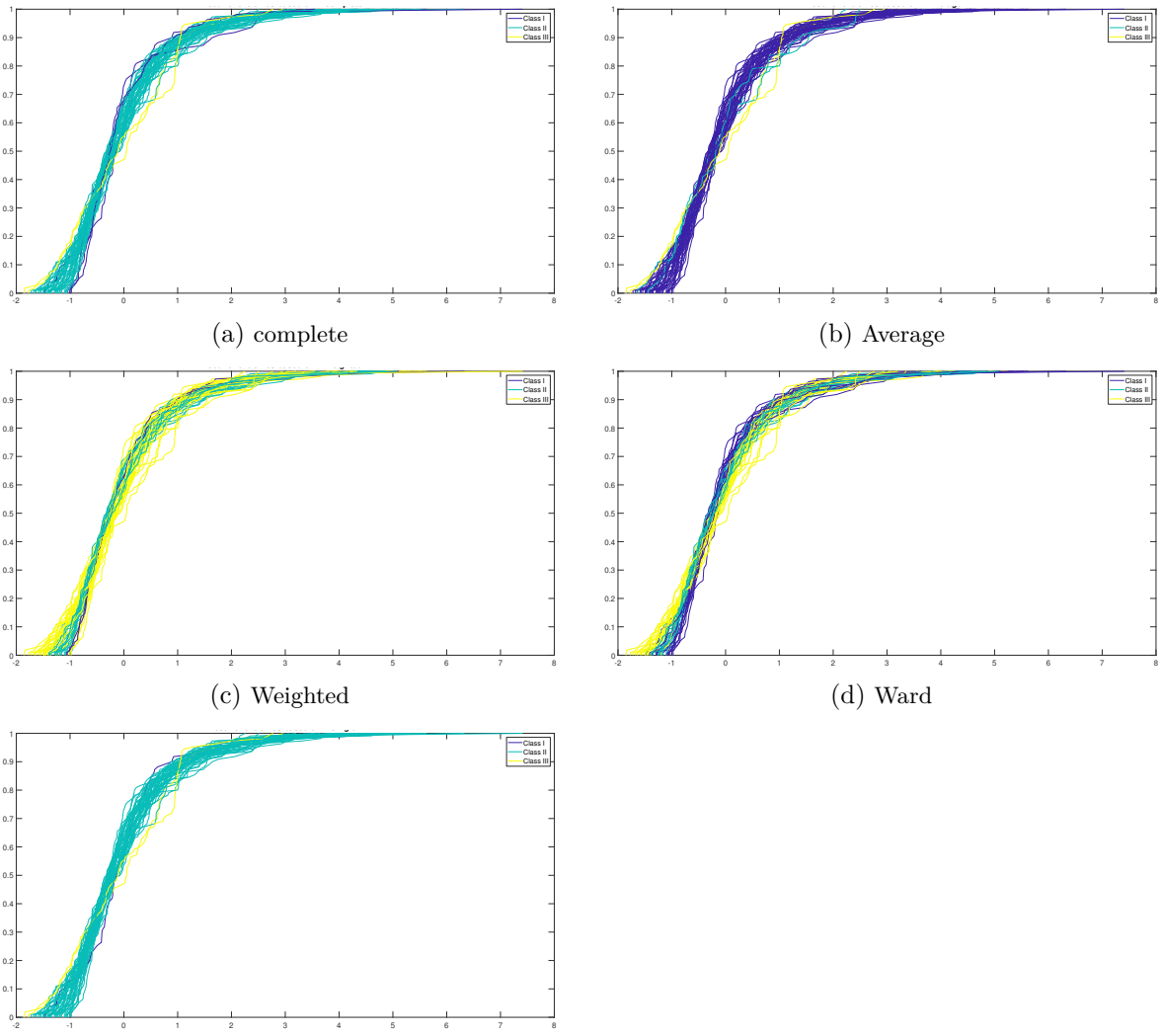
5.3.2.3 Clustering validation

The obtained Clusters by Complete, Average, Weighted, Ward, and Single methods are mostly different due to the different nature and essence of the linkage methods. Therefore, to evaluate the performance of different clustering algorithms, the resulting clusters in the previous steps are investigated using three described clustering evaluation methods in Section 5.2.6. The results are reported in Table 5.1.

Table 5.1: Different clustering evaluation coefficients for various linkage methods

	Complete	**Average**	**Weighted**	**Ward**	**Single**
Silhouette	**0.516**	**0.513**	0.386	0.322	0.454
Calinski-Harabasz	10.540	9.017	14.354	**16.925**	5.886
Davies-Bouldin	0.954	0.984	1.124	1.299	**0.649**

According to the clustering evaluation, the Silhouette coefficient shows that both Complete



(a) complete (b) Average

(c) Weighted (d) Ward

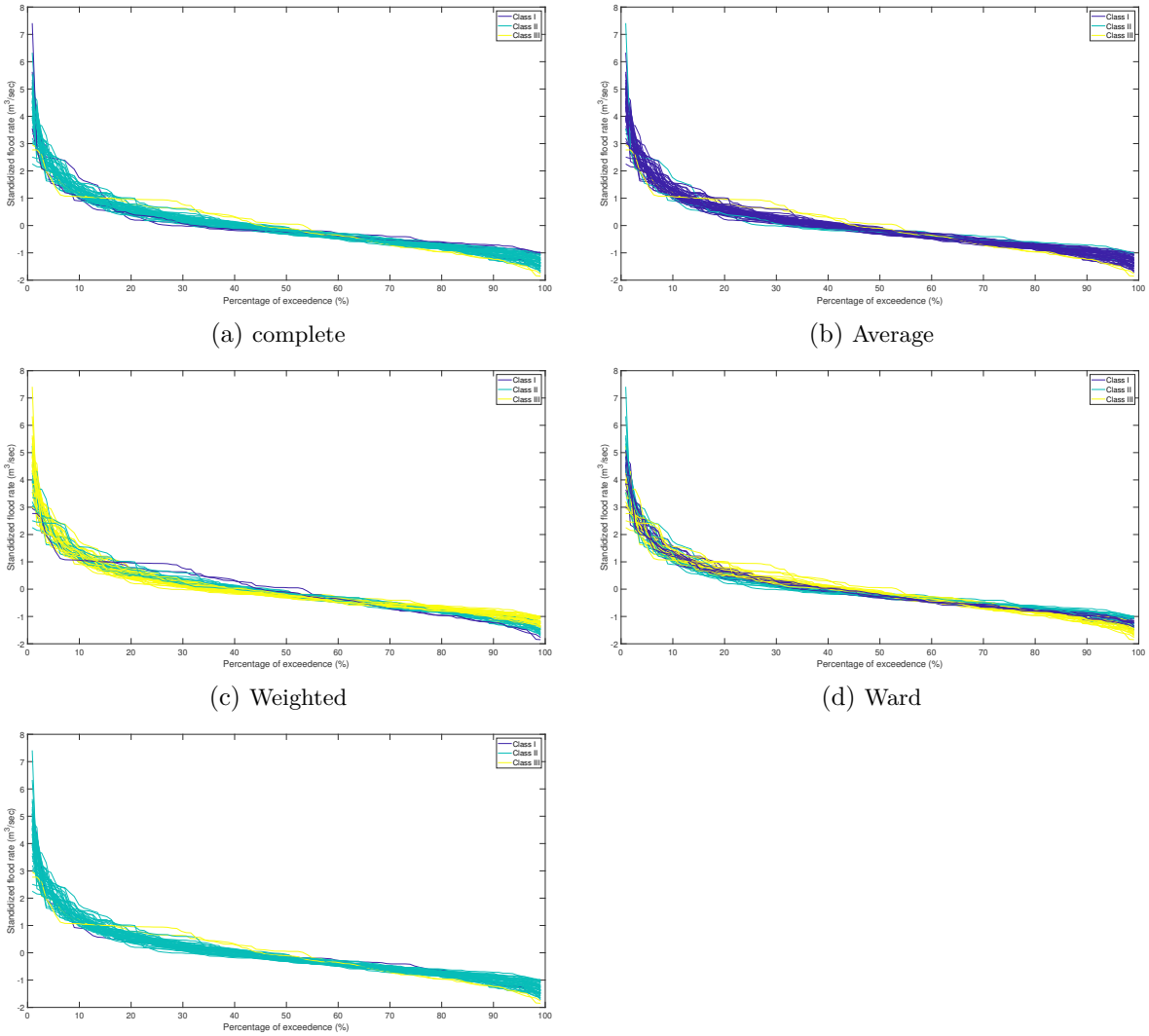
(e) Single

Figure 5.6: Resulting clusters of the standardized general floods in empirical CDF form

and Average linkage methods result in the best clusters. Nevertheless, the CH criterion shows that the Ward method has the highest coefficient value. The DB index expresses the Single method is the best way to cluster data, then the Complete and Average methods are the most trustable algorithms for grouping general floods. We already know that the Single linkage method is based on the shortest distance, and it is an appropriate way to determine the most distinct areas or sort of outlier stations.

5.3.2.4 Optimum number of clusters

By changing the number of clusters, the evaluation coefficients of clustering are changed. The best performance of clustering, i.e., the optimum number of clusters is shown in Table 5.2. Regarding the Silhouette coefficient, clustering into three groups has the highest rank with both Complete and Average methods. However, the CH criterion shows better grouping



(a) complete (b) Average

(c) Weighted (d) Ward

(e) Single

Figure 5.7: Resulting clusters of the standardized general floods in empirical FDC form

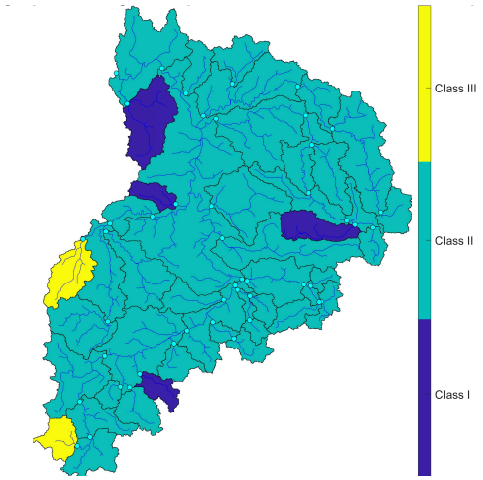
Table 5.2: The best number of clusters regarding evaluation criterion and linkage methods

Nr. of clusters	Silhouette	Calinski-Harabasz	Davies-Bouldin
3	**0.516 (Comp.)**	**16.925 (Ward)**	**0.649 (Sing.)**
4	0.463 (Sing.)	**17.851 (Ward)**	0.490 (Sing.)
5	0.406 (Ave.)	16.692 (Ave.)	0.491 (Sing.)

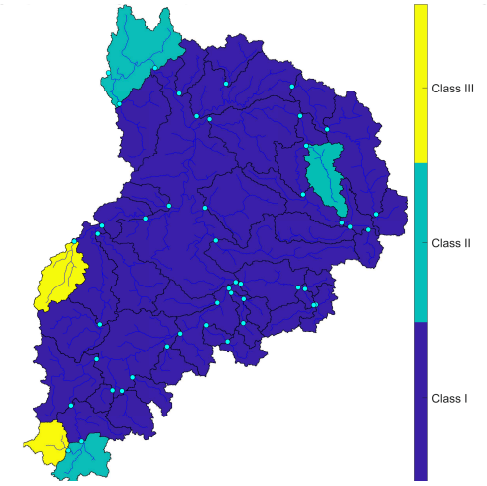
when data are clustered into four groups. Then the clustering with three clusters is in the second rank. Davies-Bouldin also introduces the best number of clusters into three groups. Apart from the Single linkage method in the DB criterion, the Average linkage method got a score of 0.742 with four clusters.

5.3.2.5 Mapping clusters

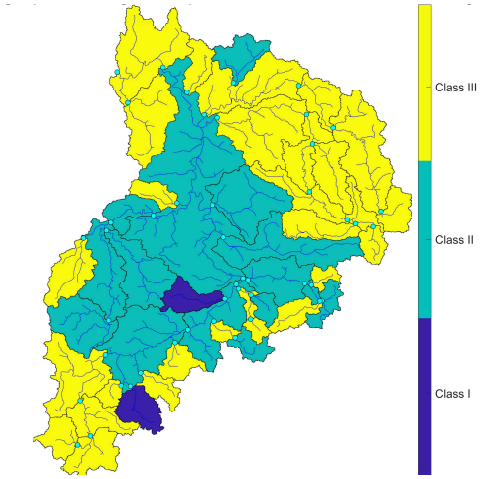
The understanding of Figure 5.6 may be difficult; therefore, Figure 5.8 can illustrate similar groups with the same color. Therefore, the resulting clusters are plotted in Figure 5.8. This figure shows how different sub-catchments are near/far from each other regarding the magnitude of general floods. According to the result of the clustering evaluation, more attention should be paid to the Complete and Ward method maps. In Figure 5.8b, almost all the Neckar sub-catchments are grouped in the same cluster with the dark blue color and some small sub-catchments, mainly in the border of the basin, are in other clusters. Figure 5.8d has more details and each of the clusters has more sub-catchments inside. However, the concept of this map and the previously mentioned maps are similar. Here, west of the upper Neckar and parts of the Jagst and Kocher in the north of the Neckar catchment are clustered in the same color. Some small sub-catchments in the dark blue, especially in the upper Neckar and the Murr and Rems river, are in another group. Also, the mainstream's sub-catchments after Plochingen and the east part of upper Neckar are in Class II. The Single linkage method showed the sub-catchments that acted differently from all the others. It means, in these upstream regions, floods are way different from the rest of the catchments. In general, the small sub-catchments and headwaters are reacted differently from the other parts of the Neckar basin. In all the clustering maps except the Ward map, the subscatchments along the main river are hydrologically similar to each other. However, in the Weighted methods, the upstream and small sub-basins are recognized differently. The geologic Karstic catchments are mainly the same as subcatchments along the Neckar river. The Jagst and Kocher, which are separated in Figure 5.8c and 5.8d have the lowest number of hydraulic structures. It may be a reason to have distinct behavior. The most number of hydraulic structures are built in the center of the map on the Murr and Rems tributary (station Nr. 434 and 1470, see Table 2.2 and Figure 2.7b), which is in class I in Figure 5.8d. The Murr is a headwater sub-catchment in the middle of the case study. Usually, these kinds of catchments are on the border of a catchment with high elevation. Also, the southwest to the east edge of the Neckar with the Jura lithospheric feature with karstic limestones cause differences for some small subcatchments upstream of the Neckar.



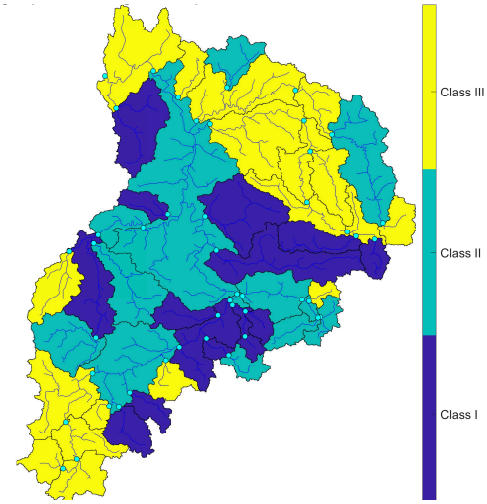
(a) complete



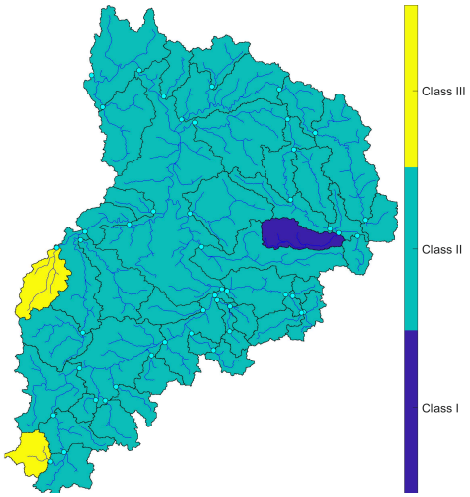
(b) Average



(c) Weighted



(d) Ward



(e) Single

Figure 5.8: Mapping clusters of different linkage methods

5.4 Conclusion

In this chapter, the existing and traditional methods of the regional flood frequency analysis are briefly introduced. Then, the concepts of the recurrence interval and the differences between the two approaches of annual maxima and partial maximum series are investigated. Further, the flow duration curve (FDC) diagrams for the whole time series and the time series of the maximums are then examined. This chapter uses a hybrid distribution-based hierarchical clustering with five different linkage methods to provide the best clustering performance for the general extreme floods. The clustering maps are presented with different clusters due to the distinct nature of linkage methods. However, in all maps, the sub-catchments in the mainstream before the outlet are in the same cluster; the central part of the Neckar catchment except some small upstream sub-catchments acted similarly.

Depending on clustering algorithms, the resulting clusters are better grouped in low volume or extreme volume flood. The polar dendrogram technique provided a new perspective on how to group data. These graphs may help readers to better catch the differences between groups in comparison to the normal dendrogram. Clustering is calculated for both the KS decision matrix and the KS two sample statistics matrix. The final clusters of KS statistics are rational due to the more information in KS statistics. Afterward, the dendrogram output on the CDF diagrams showed the distance and proximity of the clusters to each other.

The resulting clusters are compared using three clustering evaluation methods. Different clustering criteria show different best possible linkage methods. However, the Davies-Bouldin criterion shows the Single methods as the best performance, and it presents a grouping in a way that, some sub-catchments reacted entirely different from all others. Therefore, the next lower DB index is the Complete and Average, which is similar to the Silhouette coefficient results. Depending on the research goals, the clustering evaluation must be selected to determine the best possible data groups in the multidimensional space. In conclusion, the Silhouette coefficients with the Complete and Average linkage methods are recommended to evaluate the hierarchical clustering of general floods. In addition, the Calinski-Harabasz index is appropriate when the objective is to minimize the variance employing Ward methods. The reason why the results of the DB index differ from the other two indices is that the most sensitive method to distances is the Single method. This evaluation criterion is based on a ratio of within-cluster and between-cluster distances. Thus, the Single linkage is the shortest distance and is suitable to determine utterly different behavior among stations. Moreover, it can be helpful to find outliers in clustering, which mainly belong to small sub-catchments.

Due to investigating the optimal clustering evaluation criterion, the optimal number of clusters in the Neckar catchment and for general extreme floods is three clusters. Therefore, the Silhouette and Calinski-Harabasz as two clustering evaluation criteria are recommended that performed better in combination with agglomerative hierarchical clustering.

Moreover, FDCs expressed that the catchment differences are bold when the low and high flows are investigated. Also, for floods, the regions with low volume floods are easy to cluster into different groups. However, the extreme floods did not follow any pattern on the FDCs.

To conclude, the resulting cluster maps mostly agreed that the general floods occurred similarly in the catchment and mainly have the same clustering on the mainstream. However,

the Ward map has more information in terms of the complexity of clustering in this region. All the clustering maps show that the area between Horb in the middle of upper Neckar to Rockenau, one station before the outlet of catchments, reacted similarly. The steepness of the slope is different in between 50 and 100 km to the outlet of the Neckar catchment (see Figure 2.6a). It can be a reason that why the last sub-catchment is differently clustered from one station before the outlet gauge. The first three stations of the upper Neckar in the southwest of the catchment, also have similar distribution similarities. In addition, they have a high hillslope gradient that distinguishes them from the rest of the catchment (see Figure 2.6c). Moreover, the Jagst River in the north of this region shows different behavior, especially in Weighted and Ward maps. This pattern is similar to the monthly and yearly time interval trend maps (see Figure 2.10). The headwaters and small catchments on upstream of the Neckar are clustered into different classes in Figure 5.8. The highly anthropogenic areas in the Murr and Rems plus Elsenz with the high number of dams, reservoirs, and hydraulic structures are the regions that are determined as different clusters. In the Weighted and Ward maps in the mentioned figure, the Jagst and Kocher tributaries are taken apart from their neighboring areas. It can be because of a few man-made structures in this region. It is the same for the origin of the Enz at Neunbürg (station Nr. 76123). Therefore, the anthropogenic changes are highlighted clearly in the class III of Figures 5.8c and 5.8d.

Almost all the regions along the main river show the same pattern. However, the origin of the Neckar in Rottweil and Horger are always in the different cluster (see Table 2.2 and Figure 5.8). Usually, some regions which had a negative slope of change (see Figure 2.12) for the whole time series are clustered differently in Weighted and Ward maps. Also, the upper Jurassic with karstic limestones and crystalline rocks in the south edge of the Neckar basin may be the reason for having different clusters than other sub-basins along the main river. In general, headwaters, small size catchments, and anthropogenic changes are the factors that highly impact the resulting clusters. In addition, the existence of more than 4500 hydraulic structures, dams, and reservoirs has an enormous impact on the hydrological investigation in this basin.

6 Simulated annealing clustering of general floods

6.1 Introduction

The spatial evaluation of flood hazards is of ultimate concern for urban developments, agricultural and land use management, and infrastructure planning. The advancement of the mitigation compositions may optimally reduce the devastation loss (Hosseini et al., 2020). Improvement of the novel methods and continued promotion in developing the methods for hazard mapping are particularly indispensable for floods risk mitigation (Bui et al., 2020; Kalantar et al., 2021).

A new robust clustering technique is introduced and performed by implementing an optimization scheme in this chapter. In other words, finding similar flood behavior in different parts of a catchment is the goal of an optimization problem. So far, the traditional clustering algorithms did not fully consider clustering as an optimization problem. Like the previous chapter, the general floods are investigated to determine the best possible data groups.

Every clustering method initially demands to have the dissimilarity matrix as an input and basis of calculations. The only similarity between this method and the applied method in Chapter 5 is employing the same similarity matrix as the clustering input. The distance among flood time series is calculated with the two-sample KS test. The Simulated Annealing (SA) is the global optimization manipulated to determine and group similar sets of flood time series. Simulated annealing is a probabilistic technique to find a global optimum of a specified function, i.e., it is a meta-heuristic optimization algorithm to approximate global optimization in an ample solution space.

Some researchers implemented the SA algorithm to solve the clustering problem (Klein and Dubes, 1989; Selim and Alsultan, 1991; Bandyopadhyay et al., 2001; Maulik and Bandyopadhyay, 2002; Seifollahi et al., 2019). They used hybrid algorithms, which combined clustering methods with optimization approaches. In contrast, the applied method in this section is a robust technique that solves clustering problems by manipulating simulated annealing in different optimization levels. The term “level” means the twisted SA loops inside an optimization operation (Modiri and Bárdossy, 2020). The initial idea of solving problems employing an optimization scheme for automatic clustering has been presented recently (Brown and Huntley, 1992; Dey et al., 2019; Modiri and Bárdossy, 2020). This algorithm aims to group flood series regardless of the initial defaults of clustering problems. Traditionally, clustering is constrained by the number of clusters and the size of points inside each data group. The robust and automated simulated annealing overcome such assumptions to determine the best possible answer to the research question.

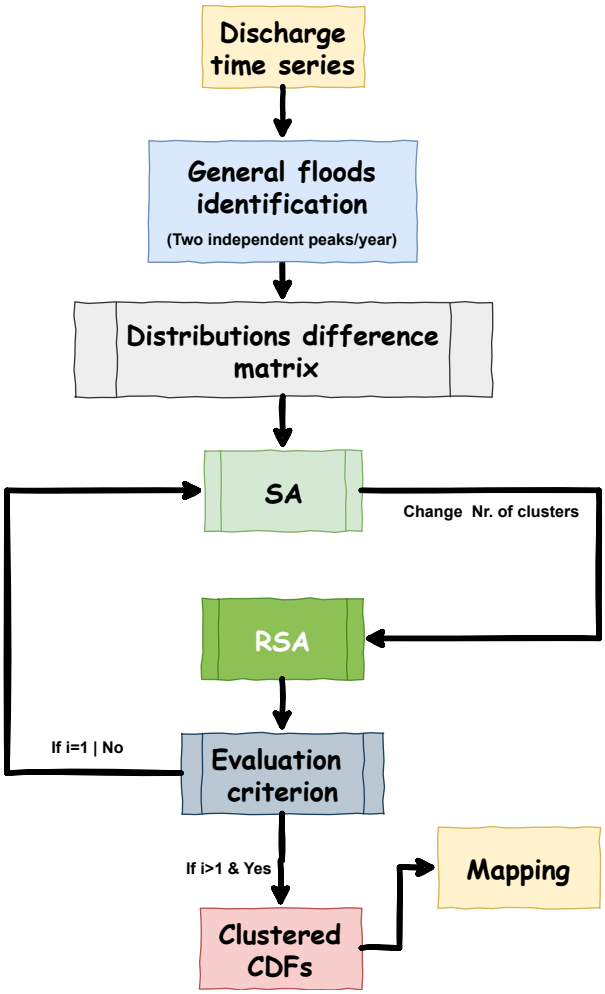


Figure 6.1: The optimization scheme clustering, based on distribution similarity

In Figure 6.1, the flowchart of this chapter is illustrated to clarify the structural overview. The purpose of this study is extreme events analysis. The same as preceding chapters, extracting floods from the discharge time series is the first step. Then the maximum absolute differences of CDFs are calculated to run the inner loop SAs. To solve and find the optimum answer of clustering possibilities, the outer SA is considered to make a dynamic clustering algorithm that can change the size of clusters inside the method. The factor that has to be minimized in this algorithm is an evaluation criterion like the Silhouette coefficient of each iteration. In the end, the clustered CDFs are determined and mapped to better illustrate the resulting clusters.

Moreover, an event-based clustering is performed by assuming the outlet of the Neckar basin near the Heidelberg as the reference. This investigation will show that to what extend the extreme floods are reacted similarly comparing the biggest events in this measuring gauge.

6.2 Methodology

6.2.1 Similarities between the cumulative distribution functions of general floods

The first input for the clustering is always the preparation of the similarity/dissimilarity matrix. Regarding the methods described in Section 3.2.1, the general floods are identified, and then dissimilarity among CDFs of flood series are computed based on the applied method in Section 5.2.4. Then the calculated matrix is used as a distance matrix in the simulated annealing to find the objective function's global optima. The optimization needs some initial assumptions like the number of clusters or the limiting factor for the number of stations in a cluster, and other optimization parameters, including the number of iterations and temperature. A newly added feature to the standard algorithm in this dissertation neglects some of the basic assumptions. SA intertwined loops are an idea accompanied to this section. This means, after performing a complete cycle of the SA method and finding the optimal data and clusters, the calculations are repeated by changing the number of clusters. This process is repeated until the best appropriate clusters and points inside clusters are identified. It is an automatic and robust simulated annealing. In the following parts, the compositions of this procedure are described, and the objective function is explained.

6.2.2 Simulated annealing optimization (SA)

In total, the optimization can be divided into two parts (Palop et al., 2010). Heuristic algorithms can succeed in such shortcomings of mathematical techniques and deliver distinguished outcomes when applied to particular problems; however, they remain inappropriate in a wide range of conditions. The meta-heuristic optimization techniques based on iterative simulations have been proposed to address the mentioned problem. They provide appropriate answers, which can be found using limited memory and computation time without requiring complex derivatives (Kim et al., 2014). Many meta-heuristic algorithms that combine rules and randomness mimicking natural phenomena have been developed. The simulated annealing algorithm is a meta-heuristic optimization method that was initially designed to imitate the slow cooling of metals. A progressive particle movements' reduction characterizes this method until the lowest possible temperature/energy state is reached (Kirkpatrick et al., 1983).

Here, in the flood analysis, it is demanded first to determine the objective function to minimize. Therefore, we must first turn to the research question for this section. In this chapter, clustering is presented as an optimization problem. The question now is which cluster or clusters will be the best results and what will be the spatial distribution of areas with similar floods. By asking these questions, we have found the internal clustering criteria, which can be expressed as the optimum parameters for measuring the quality of clusters in the optimization problem. But, can the SA alone recognize the best answer? The initial answer to this question would be a no. In multivariate statistics and in multidimensional space, where dozens of stations in a basin are spatially distributed, it is impossible to find an optimum score without implementing nested optimization algorithms.

According to clustering results in the previous chapters, three clusters are assumed as the initial number of clusters. Therefore, all general flood series distributions are randomly

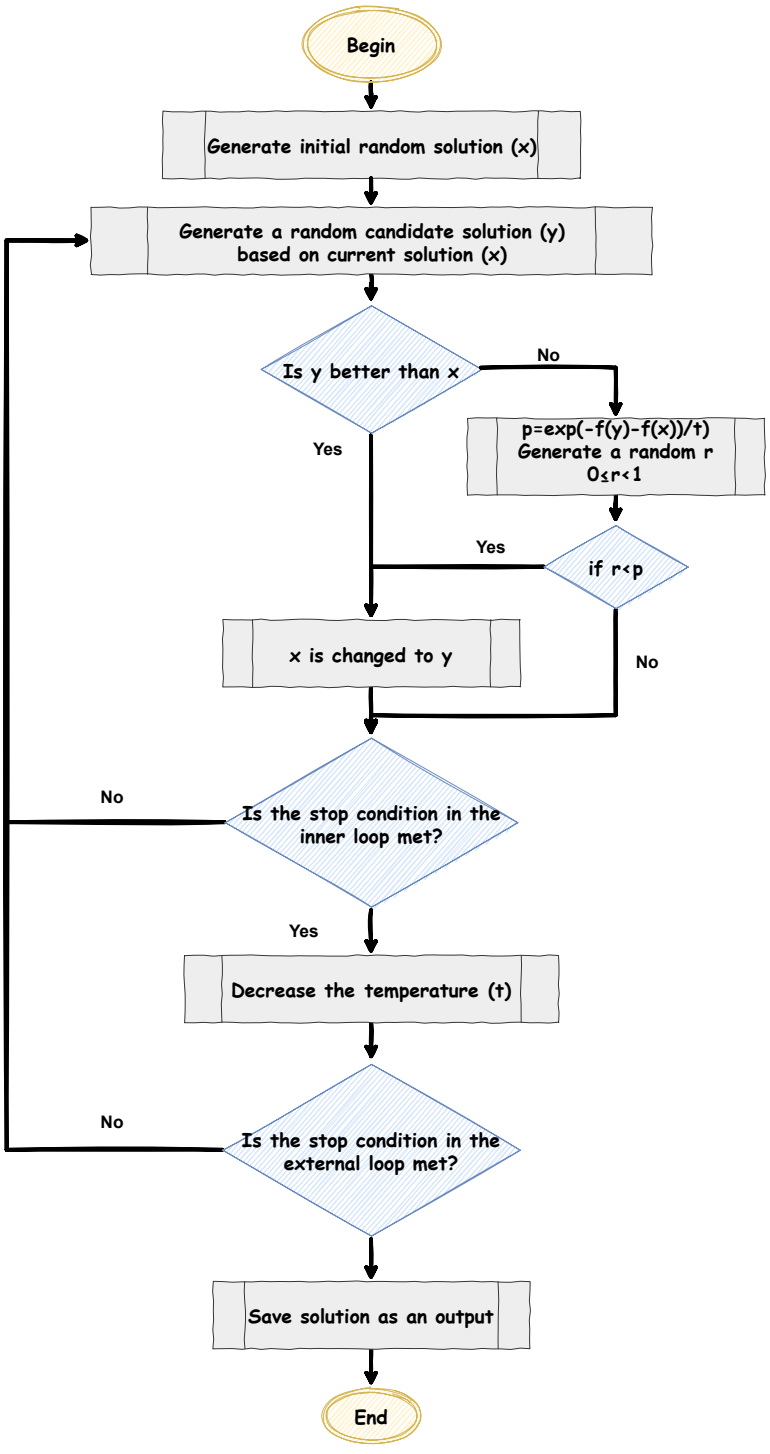


Figure 6.2: The concept of the simulated annealing algorithm

divided into three groups. Then, the simulated annealing is conducted to define the optimal solution in this step. In Figure 6.2, the stepwise procedure of the SA's concept is illustrated.

6.2.3 Robust Simulated Annealing (RSA)

The SA algorithm is a randomized search procedure that starts from an initial solution. A limiting factor or a control parameter is set to an initial 'temperature' value, which systematically decreases according to a cooling process (Hamzadayi and Yildiz, 2013). Before beginning the SA optimization algorithm, it is required to choose initial values for the parameters, the cooling process and the limiting factors as ceasing conditions in the loop (see SA parameters in Table 6.1). The robustness of simulated annealing comes into account when the optimal solution for initial grouping is determined. The number of points, i.e., stations, within clusters is changed to repeat the past SA step. Therefore, it is possible to check the dynamic size of clusters without considering initial assumptions. This step will continue until the optimal answer is obtained. Each time the number of flood series in each group changes, the temperature rises again, and consequently, the cooling process begins. Adding SA intertwined loops to each other makes it possible to eliminate other initial assumptions/values. After finding the best answer in the previous step, increasing or decreasing the number of clusters and then repeating the previous steps is possible. Depending on the number of iterations and the limiting factor in each loop, the time to find the optimal answer will also change.

Table 6.1: Simulated annealing parameters

Parameters	**Initial value**	**Sign**
Nr. of cycles in the RSA	100	n
Nr. of trials per cycle in the SA	10000	m
Probability of accepting the worst solution at the start	0.7	P_1
Probability of accepting the worst solution at the end	0.001	P_{end}
Initial temperature	$-1/log(P_1)$	t_1
Final temperature	$-1/log(P_{end})$	t_{end}
Cooling factor	$(t_{end}/t_1)^{1/(m-1)}$	c_f
Limited Nr. of stations in a cluster	5	lim

6.2.4 Objective function

Clustering validation has long been recognized as one of the vital issues essential to the success of clustering applications. Here, In general, clustering validation can be categorized into two classes: external and internal. The external validation methods consist in comparing the results of cluster analysis to an externally known result, such as externally provided class labels (Rendón et al., 2011a). Nevertheless, in the unsupervised learning and hydrological clustering of a catchment, the labels are not determined, and they are precisely the goal of clustering (Rendón et al., 2011b). Therefore, in this chapter, an internal clustering validation is applied as the objective function of an optimization scheme problem.

Several studies have been employed the Silhouette score as a more accurate internal validation than the others (Petrovic, 2006; Dudek, 2019; Lengyel and Botta-Dukát, 2019; Batool and Hennig, 2021; Shutaywi and Kachouie, 2021). However, nobody employed it as a direct tool for clustering in flood analysis. The implemented optimization algorithm tries to maximize

the Silhouette coefficient as the fitness ratio presented in Section 3.2.5. Therefore, here some advantages of utilizing this coefficient are described, and a schematic example is illustrated.

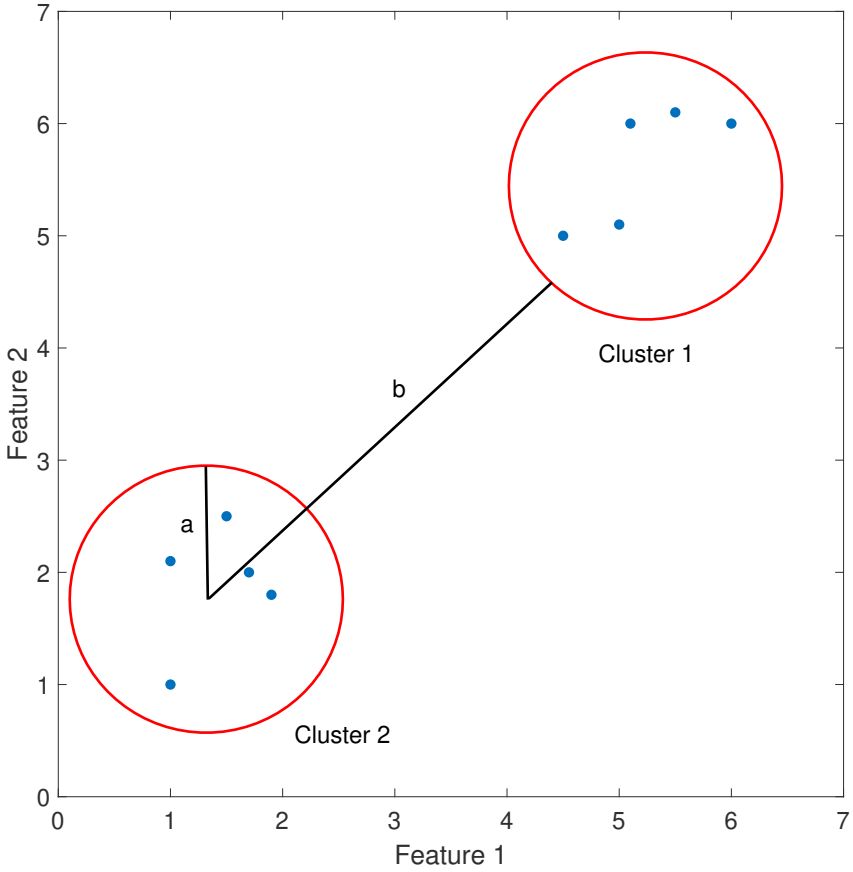


Figure 6.3: A schematic example for the Silhouette score

Figure 6.3 shows the advantage of the Silhouette coefficient in employing the average of intracluster distance (a) and the average of inter-cluster distances (b), which is explained in Equation 3.14. This score uses inter-and intracluster distances in its function, while most other methods only use intra-cluster distances (such as the elbow method). Silhouette analysis also has added advantage to find the outliers if present in a cluster (Lodhi et al., 2019). Silhouette index for a set of sample data points is utilized to measure how well-separated the clusters are. The search space becomes much uneven when the objective function in an optimization problem like a clustering algorithm becomes more complex due to the multidimensional inputs. However, Silhouette is an appropriate objective function for the SA-based clustering; it is highly possible that the search algorithm does not converge as required. In comparison several studies on different internal validation like gap statistics Tibshirani et al. (2001), DB (Davies and Bouldin, 1979), CH (Caliński and Harabasz, 1974), elbow method based on distortion score (Thorndike, 1953), Dunn index (Dunn, 1973), and dozens of introduced indices in Liu et al. (2010), the Silhouette index is the one is more accurate than others as objective function (Petrovic, 2006).

The adopted distribution-based dissimilarity matrix of general floods is an input of this objective function. Each inner and outer loop of the SA algorithm reaches the optimum for

the Silhouette coefficient. Therefore, the clustered flood series are automatically assessed by evaluation criteria.

6.2.5 Flood events identification and rank order based clustering

The annual maxima are identified as the described method in Section 3.2.1. Then the selected peaks have to be sorted and given a rank order. The normalized magnitude of floods are calculated as follow to have a reasonable range of changes for the obtained extremes all over the basin:

$$n_i = \frac{x_i}{\bar{X}}, \tag{6.1}$$

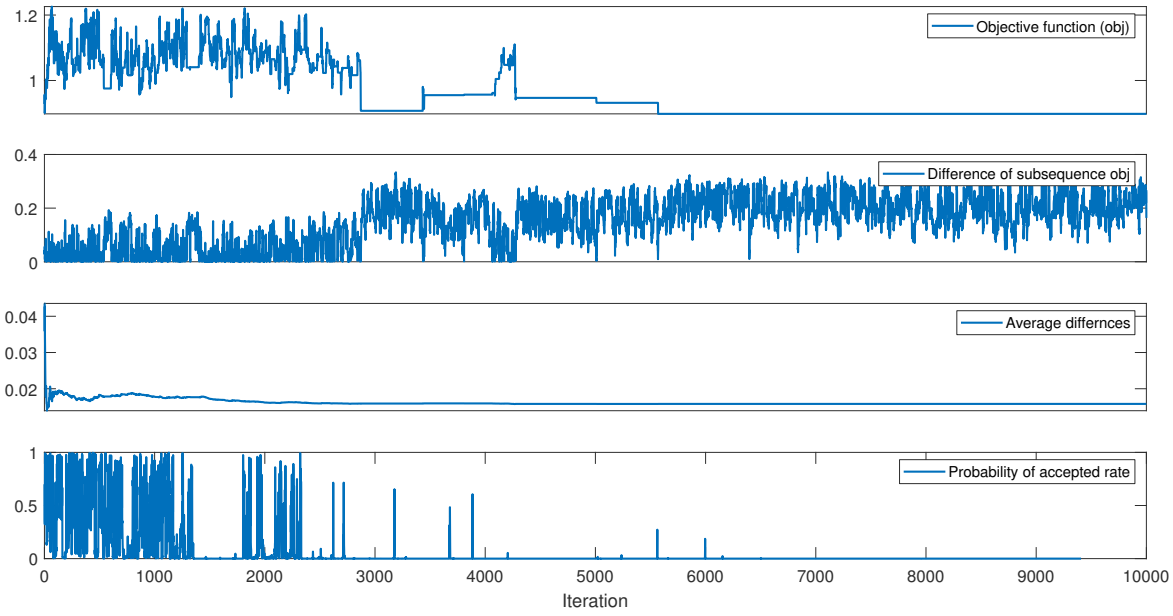
where x_i is the magnitude of ith flood and $\bar{X}$ is the mean of annual maxima for each measurement gauge. Therefore, the rank order and the normalized floods magnitudes are the two parameters that are used as the input of hierarchical clustering for the event-based analysis. The clustering in this section is performed as described in Section 3.2.4.

In addition, empirical CDFs of obtained flood series clusters are plotted and calculated to visualize the possible differences among the high and low volume floods. As a result, the normalized series of the floods magnitudes in the same cluster are gathered as a new series, and then empirical CDF is computed for these clusters.

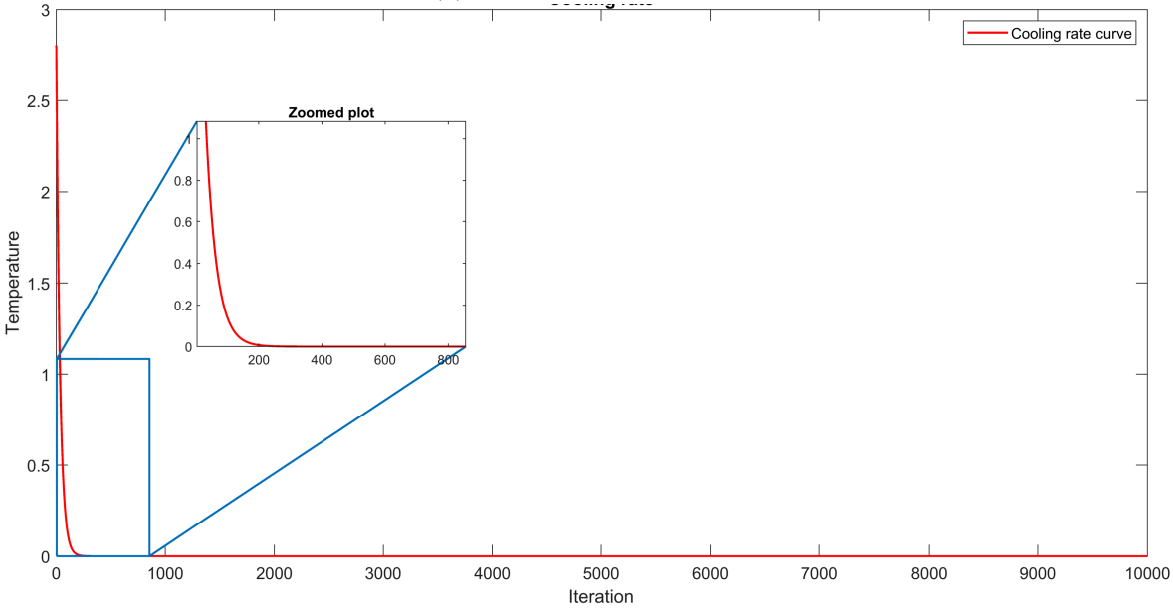
6.3 Results

In this part, first, the optimization parameters in the inner loop are reported, then the results in the outer loop for RSA are plotted. Subsequently, a map of the cluster area is illustrated to show similar behavior regions in the basin. Next, the objective functions, the difference of two subsequent objective functions, an average of differences, and the probability of acceptation are plotted in Figure 6.4a. Here, the objective function is tried to be minimized and it is equal to one minus average of Silhouette coefficient for each iteration. As a result, the objective function series show a reasonable performance of the optimization algorithm in the upper panel. Consequently, the best optimum result is mainly determined in the first 3000 iterations; however, the algorithm accepted some changes between 3000 to 5000. Finally, to reach the best performance of the method, regarding temperature, the cooling rate is drawn in Figure 6.4b. In the first stages, the temperature cools faster, and in the last stages, it cools down very slowly. Therefore, as shown in the figure, the temperature changes in the first 200 steps are considered to be severed. Therefore, determining optimums are correct for each outer loop.

The determined clusters and objective function inside each outer loop are plotted in Figure 6.5. The best points in a cluster, the way of their distribution and objective functions are illustrated in the upper panel and the current situation of each of which outer iteration is plotted in the bottom. Concerning the assumptions, the limited number of stations in each cluster is designed to be five in each group. It is also planned to have three clusters, the same as previous chapters; however, RSA can also add another loop to neglect this condition. In the upper left, the number of stations inside three clusters is drawn. There are eight and seven



(a) Optimization parameters



(b) Cooling rate

Figure 6.4: The optimization parameters in the inner loop of the applied simulated annealing

stations in the second and third clusters. The station number (ID) is recognized in the upper middle. After $100 * 10000$ calculation rounds, the most optimal answer for the clustering problem is obtained according to the presumptions made for the designated optimization scheme. Also, the best objective function is equal to 0.752, which is occurred in the twentieth iteration in changing the size of clusters.

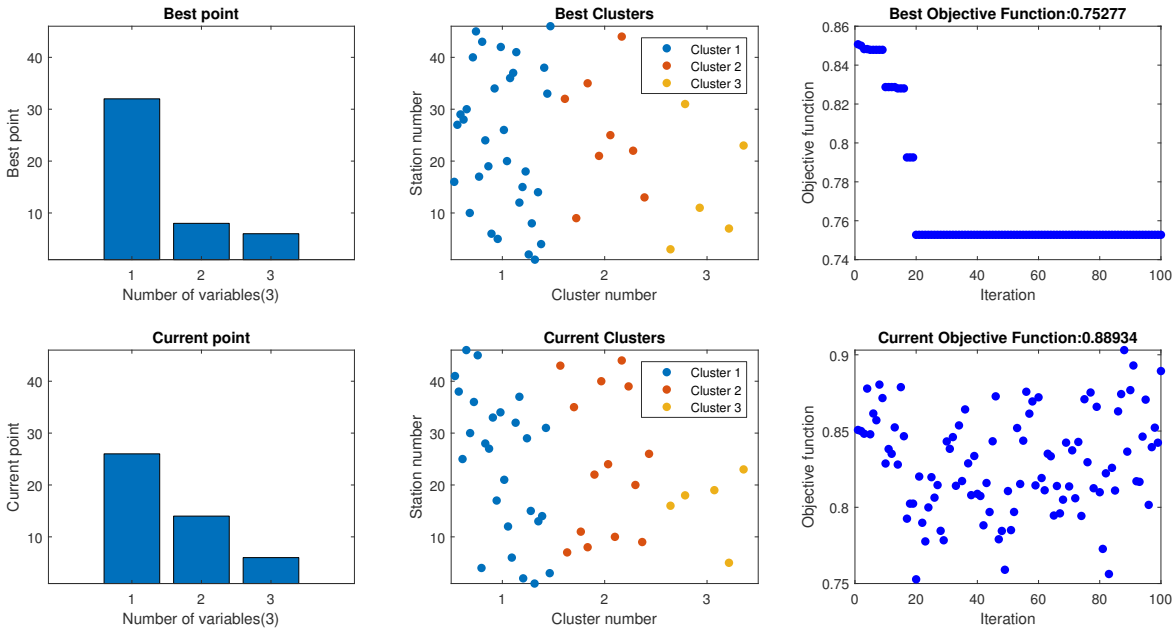


Figure 6.5: The results of robust simulated annealing optimization

6.3.1 Clustering of the general floods, based on RSA

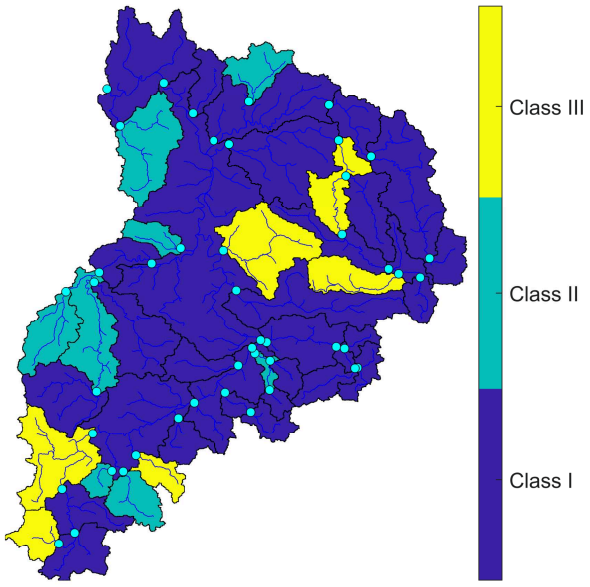


Figure 6.6: Mapping general floods clustering using RSA

The computed results of RSA are mapped in Figure 6.6. The results indicate three major clusters, which illustrate a particular pattern for general flood occurrences except in some small sub-catchments. These sub-catchments show agreement with each other in each cluster. Most sub-catchments reacted similarly in the dark blue. Some small areas on the west border of the Neckar catchment, plus two other small ones in the upper Neckar, are in the second cluster. The upstream of the Neckar and some parts of the Murr and the Kocher are in the

same group. This figure shows that the mainstream sub-catchments had the same reactions when the general flood happened. Two neighboring colors in the clustering map can be merged. If the green and yellow areas merge, we can see that floods generally acted similarly in all areas except upstream sub-catchments in the west (Nagold) and southwest, plus the Murr and Kocher. Here, the headwater subcatchments play a significant role in the resulting maps. Most of class II and III areas are classified as the two mentioned headwaters in Table 2.2.

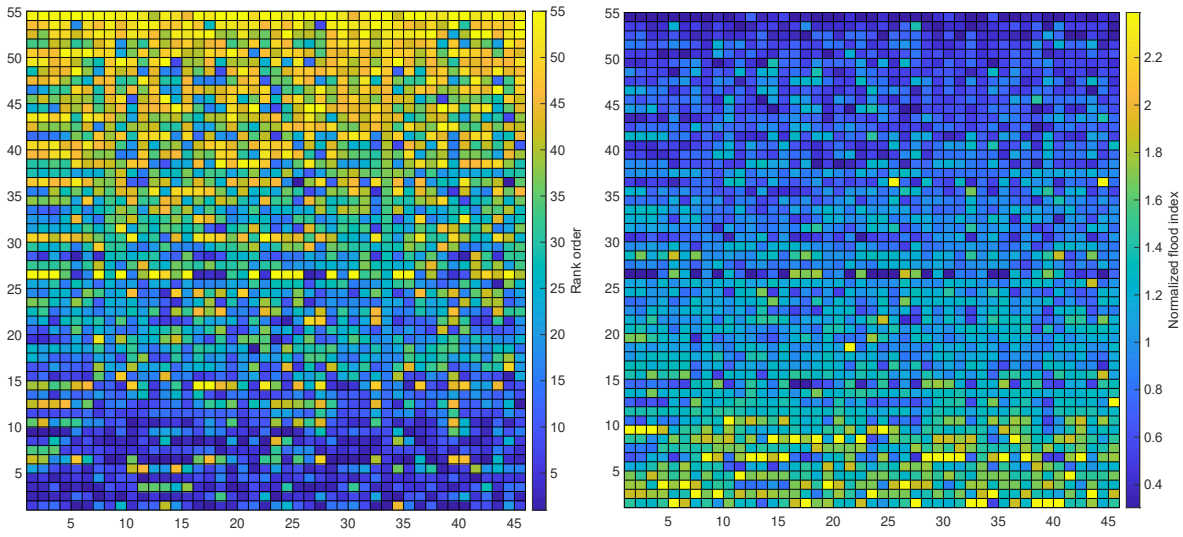
6.3.2 Comparison between RSA and KS-based clustering methods

The distribution-based clustering and optimization scheme clustering have two distinct points of view in the clustering approach. To compare the result of this chapter and the previous chapter, Figures 6.6 and 5.8 are investigated. The RSA has more similarities with the Ward cluster map. However, they have some discontinuity. Both of the methods divided the area into the following form. Some four to five upstream sub-catchments of the upper Neckar are in a separate group with some parts of Kocher. East of upper Neckar and the Murr, plus some small regions in the west of the catchment, are in another cluster. Due to some preset assumptions of the RSA method, the east part of the upper Neckar reacted similarly with the other catchments on the mainstream.

The result of Figure 6.6 is partly similar to the previous clustering map in the last chapter. Nevertheless, both results show that the headwaters are the areas that reacted differently from the others. Also, the origin of the main river due to the high elevation in the region, which is relatively always showed an unconventional reaction than others comparing all the other clustering maps. Here the two stations of 7612 at Neuenbürg and 4422 at Pforzheim on the Enz and Nagold rivers are shown plainly, which can be due to the red sandstone geological feature.

6.3.3 Event based clustering

In this part, by selecting the outlet of the Neckar as a reference point, the magnitudes orders of floods are investigated. It means if the enormous flood happened at this measurement gauge, what was the rank or magnitude of other floods all over the basin. Therefore, first, the functionality of floods is calculated.



(a) Biggest floods rank order

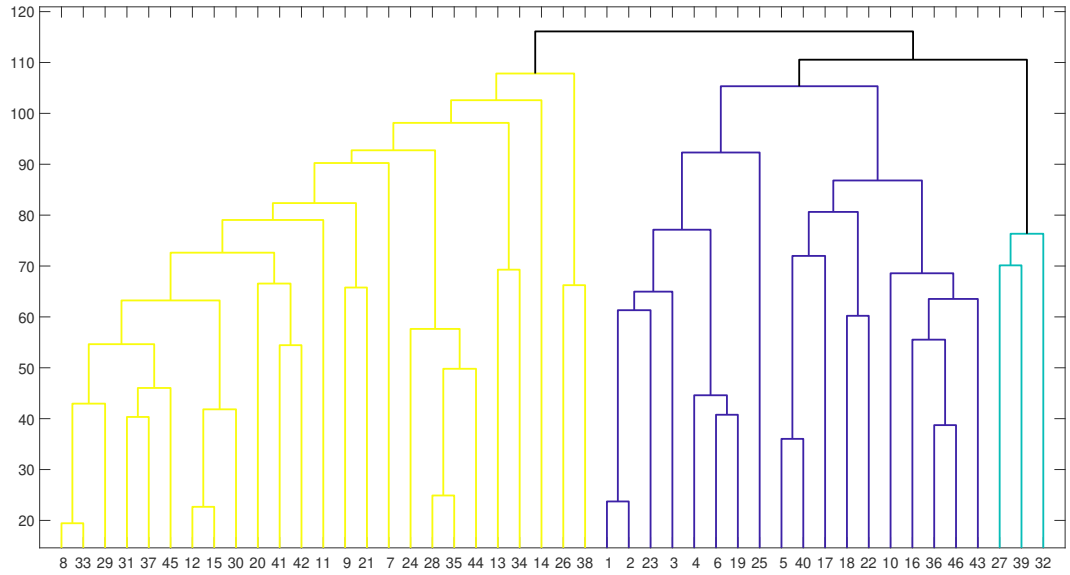
(b) Biggest normalized floods magnitudes

Figure 6.7: The functionality of floods in all subcatchments in comparison the outlet near the Heidelberg

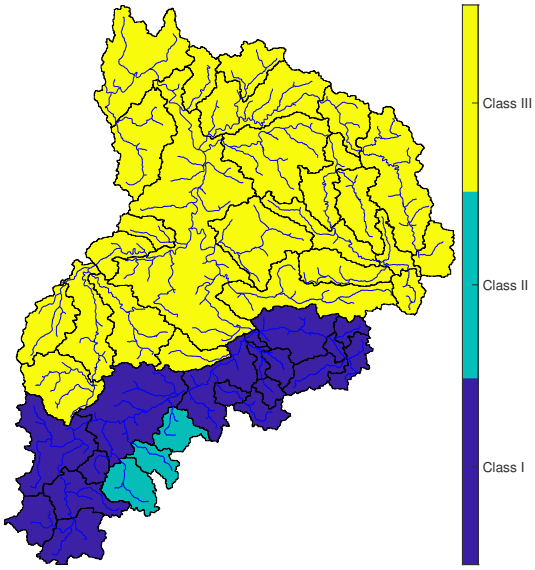
Figure 6.7 illustrates the simultaneity of the largest floods and their magnitude. i.e, in Figure 6.7a the rows are the annual maxima. The first row is the rank of floods comparing the biggest flood in Heidelberg (reference gauge - station Nr. 4416). The dark blue color is the biggest event, and the light yellow is the smallest event between annual maxima. If all the regions brought floods together with similar rank, the map has to be exactly like the color bar behind the map. The differences happened when a sub-catchment did not bring the same or similar order of floods compared to the rest of the case study. For example, in the first row, station 34 has a bold yellow color. It is a sign that for the biggest event in the Neckar catchment, this area did not follow the main pattern in the basin.

Figure 6.7b, evaluates the floods using normalized flood magnitudes explained in Equation 6.1. Here, instead of the rank order, the pixel color is the normalized flood magnitudes. The highest value is the high occurred flood magnitudes for each column (station). Here, the same as Figure 6.7a, the first row, column 34, showed a low normalize index which is a sign of differentiation of this sub-basin in the biggest flood in the region. These are the two metrics that are used in the following analysis.

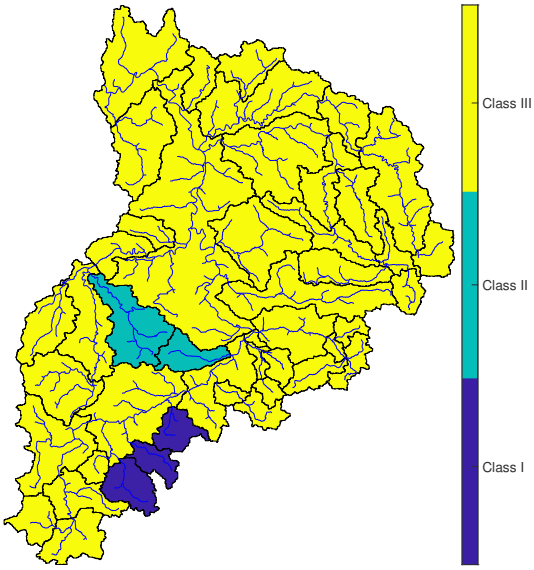
Figure 6.8 is a summary of the clustering results based on flood rankings compared to the magnitude of floods in Heidelberg. The same as the clustering method and results in the previous chapters, the dendrogram is the basis of the mapping in Figure 6.8a. Here, three clusters are recognized in different colors. Figures 6.8a and 6.8b are plotted based on employing only first 45 rank orders to show the possible differences when the full orders are used. Figure 6.8c is the clustering map that all the annual maxima's ranks are employed to achieve this map. The difference between Figures 6.8b and 6.8c is on two parts. First, the upper Neckar is illustrated as class I (dark blue) and fully separated from the rest of the Neckar catchment in Figure 6.8b, when only 45 biggest ranks are analyzed. However, using all the annual maxima's order does not recognize this area as a separate cluster. While



(a) The raw dendrogram of hierarchical clustering for the first 45 biggest annual events



(b) Event bases clustering map of the biggest only 45 annual maxima



(c) Event bases clustering map of the biggest annual maxima for the whole time series

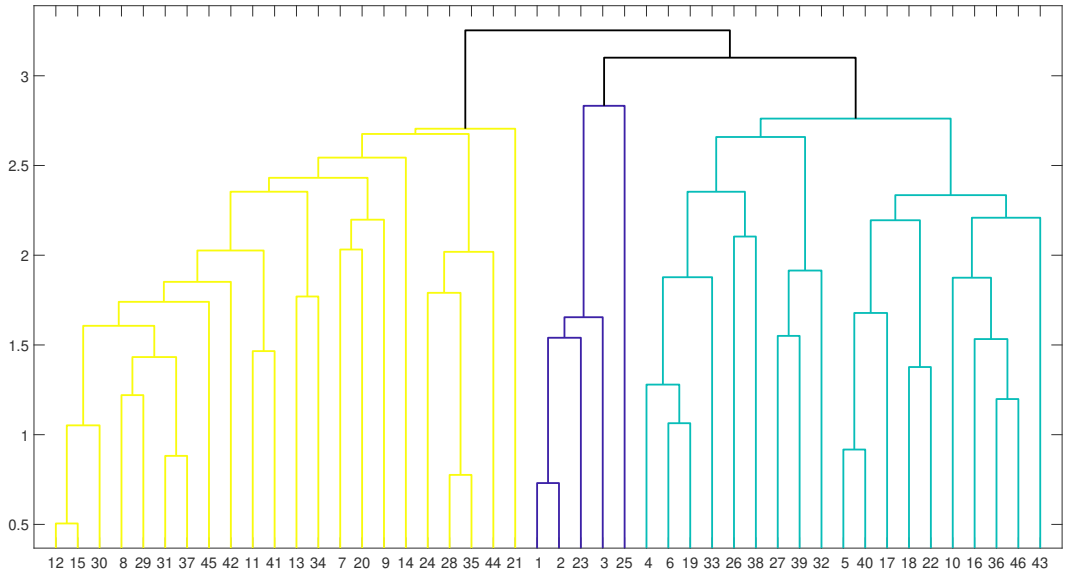
Figure 6.8: Clustering based on the rank of the annual floods in the reference station near the Heidelberg

the two subcatchments at the Oberensingen (station Nr. 2477) and 36056 at the Pforzheim (station Nr. 36056) are clustered as the class II (see Figure 6.8c). The similarity between these two maps is for the rest of the catchment downstream, where the main river is massively regulated. Also, three small subcatchments are always clustered differently. It means, these areas did not follow the floods at Heidelberg. These regions have belonged to the small size and headwater subcatchments on the upper Jurassic units.

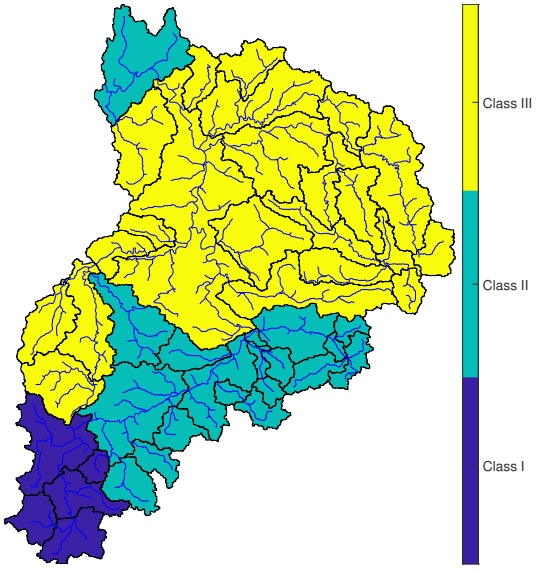
In addition, the two areas colored in green in Figure 6.8c are the areas that Figure 3.7 in the

Single linkage method showed. These are the regions that are distinguished from the others. Furthermore, it can be because of the shell limestone hydrogeological unite in this region.

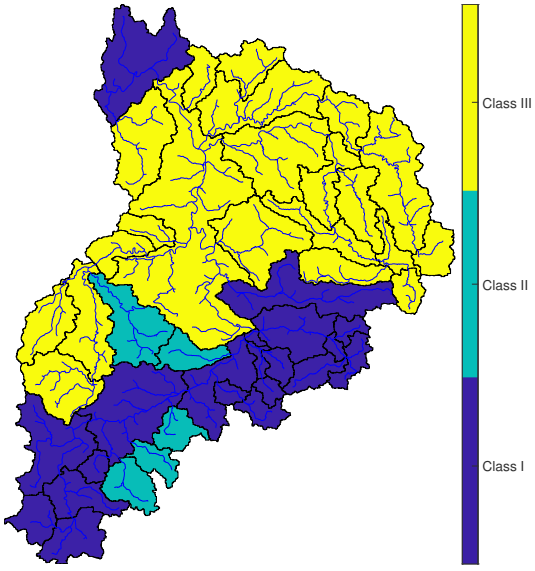
The difference between these maps is a sign that the low and high volume floods had different behavior in this study area, which has to be investigated.



(a) The raw dendrogram of hierarchical clustering for the first 45 biggest normalized floods' magnitudes



(b) Event bases clustering map of the only 45 biggest normalized floods' magnitudes



(c) Event bases clustering map of the normalized floods' magnitudes for the whole time series

Figure 6.9: Clustering based on the normalized magnitudes of floods in the reference station near the Heidelberg

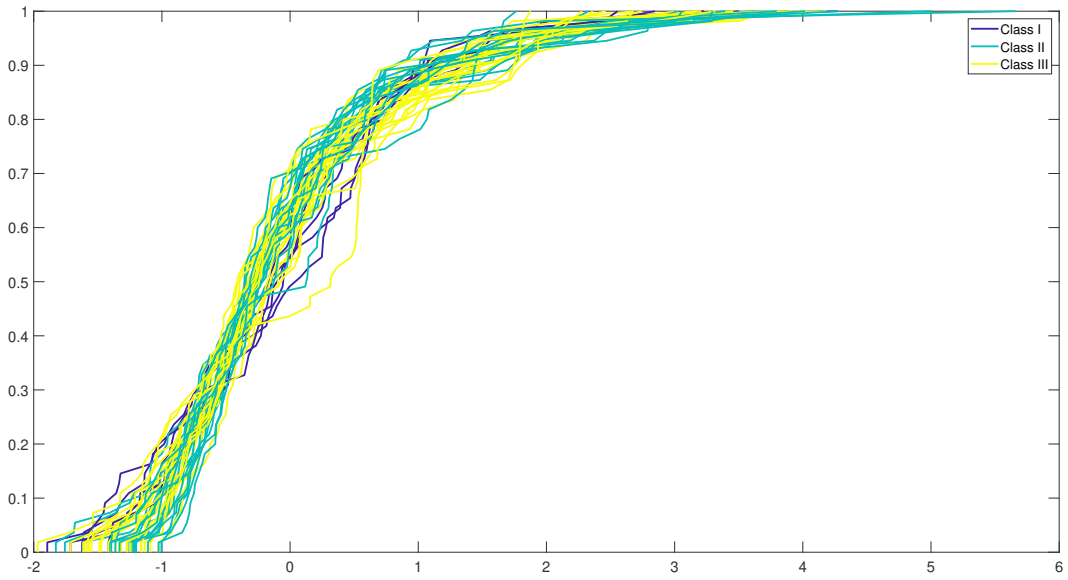
Figure 6.9 uses normalized flood magnitudes instead of the rank orders. Here, by using only 45 highest rows or having all the input matrix, the upper Neckar and the Würm subcatchment (Nr. 36056) showed different clusters and consequently different hydrological behavior.

Figure 6.9c using all the information illustrates that the upper Neckar except four small subcatchments on its border, plus the Rems river are in the first cluster. The second cluster includes the mentioned four sub-basins in upper Neckar and the Würm river.

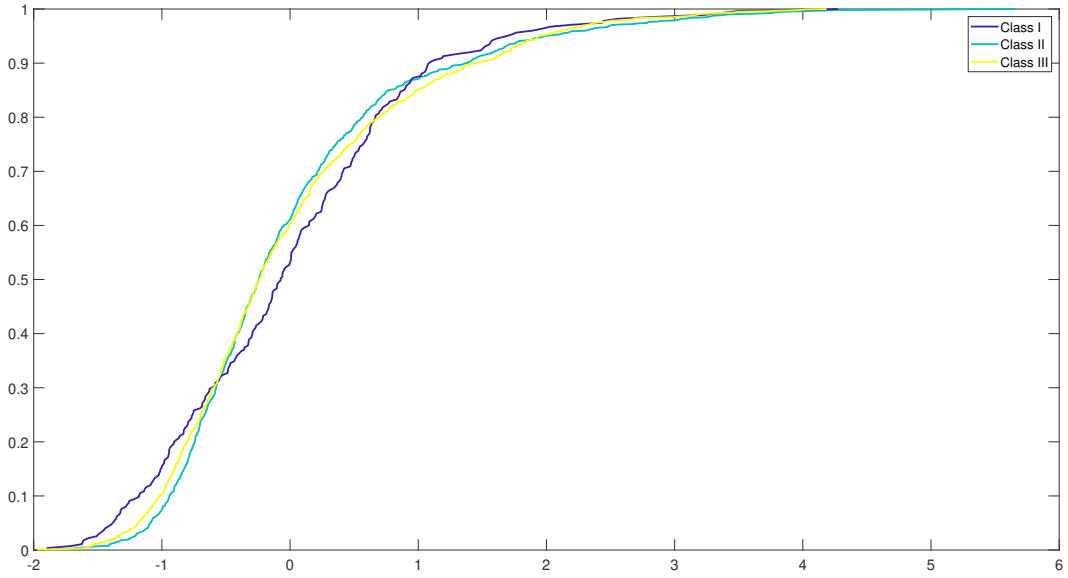
By concentrating on the most prominent 45 events, the first four sub-catchments on the Neckar river upstream show separation from the other parts. It means the enormous magnitudes of floods are not that much related to this area (dark blue in Figure 6.9b). Also, the Rems' tributary (Nr. 1470) are grouped as class III, the same as all the rest of the downstream regions.

In these figures, the outlet of Neckar is clusters similar to the upper Neckar, or at least a part of it. However, the rest of the downstream tributaries, such as Enz, Jagst, and Kocher, are in the same cluster until the Rockenau (one station before the outlet). It can be due to the existence of the Heidelberg city as a population center; some river management was performed there differently.

The obtained cluster maps express that there are differences between heavy and light floods in the Neckar basin. However, all the identified floods were the largest floods each year.



(a) The ecdfs of normalized floods in different clusters for individual measurement gauges



(b) Three ecdfs of the obtained clusters

Figure 6.10: Empirical CDFs of the extreme normalized flood series comparing the events at Heidelberg

Figure 6.10 illustrates the empirical CDFs of normalized flood's magnitudes as described in Equation 6.1. The clusters are the same as Figure 6.9c. The top figure shows no clear differences from the yellow and green clusters. But, the blue one has recognizable, especially for the low and medium size floods (in between -2 to 1 in the x-axis of Figure 6.10a). However, it is still not clear that how does each cluster works together.

Therefore, all values in the same clusters are gathered as a single vector, and then the ecdfs are drawn (Figure 6.10b). Here, the class I cluster is shown distinct behavior for low, medium, and high volume floods. The second and third clusters in green and yellow are mainly similar

to each other. However, they are twisted, and in some parts, the second and sometimes the third cluster has a higher F(x) value.

As a result, it is reasonable to say that the upstream regions except for the four small subcatchment in the Neckar catchments have a different form of distribution than the others.

6.4 Conclusion

The robust simulated annealing aims to cluster flood events with a distinct approach by implementing an optimization scheme. It is not a traditional test for clustering. The applied and designated method tries to neglect initial presumption to find the optimal answer. Clustering is considered as an optimization problem, where groups with small distances within a group and large distances between groups are to be found. SA finds the optimal global possibility of ordering stations in clusters.

The RSA clustering method is developed and needs to be investigated more in different regions, the number of clusters, and objective functions. Each of these parameters may have a significant effect on how data can be differently clustered. RSA solves the problem of the number of points in a cluster and the best number of clusters. It dynamically changes the size during the optimization procedure. Despite the good results in the inner rings, the objective function in the outer rings did not reach a suitable level. Including at least five members in each cluster may cause a low magnitude of the objective function. Furthermore, it seems that other objective functions may work better. It is possible to change the control factor to set a different number of stations inside each cluster to better compare among clustering methods. Also, the dissimilarity matrix is adopted from a non-parametric KS test, which is a maximum distance between two empirical CDFs. Changing this coefficient to another dissimilarity-based factor may change the resulted maps.

In total, due to some optimization assumptions and initial values, the resulting clusters are partly different from the resulted maps in the previous chapter. The mapping of clusters presents areas with distinct flood behavior and demands different flood protection action plans. Almost upstream and small-scale sub-catchments are grouped separately.

Stations in the second cluster in the west of the basin are followed a similar slope of change in time series as Figure 2.12. Also, yellow and green regions are primarily similar to the no trend sub-catchment in the yearly time interval in Figure 2.10d. Therefore, in the RSA method, the results do not agree to belong to a cluster in the west part of upper Neckar and they are divided into all three possible clusters, where the river network has a high hillslope gradient and high elevation (see figures 2.6c and 2.2b). However, the two green sub-catchments (Class II) are on the adjunct rivers and the flow independently leads to the mainstream. Therefore, it is a sign that these small and independent areas have to be in a separate cluster from the mainstream (station Nr. 2471 & 4408). However, station Nr. 40640 reacted differently. To conclude, it can be argued that the RSA optimization method is at the initial stage of its development and has to be further investigated. These investigations can be carried out by running the method with different objective functions and distance matrices.

Moreover, the event-based clustering is done to show to what extent the floods are reacted similarly to the outlet of catchment near the Heidelberg. The result shows that using the

rank orders and normalized flood magnitudes has relatively similar clustering results. Thus, the upper Neckar is grouped as a part that brought floods differently. Also, some small subcatchments in this area are separated from the other neighboring regions. Also, the Würm river is shown the same reaction as the mentioned small headwater sub-catchments. In this area, there are shell limestone and Jurassic hydrological units. Another point is that all the measurement gauges on the upstream of the main river, Neckar plus Fils tributary in the East of upper Neckar with the karstic feature have distinguished empirical CDF for normalized floods compared to the reference point in Heidelberg. Thus, almost three stages of low, medium, and high volume flood are separated from the downstream of the Neckar basin.

7 Conclusions

The thesis aims to address three main challenges in flood frequency analysis. One is a study on simultaneous floods' reaction areas. Another one is the analysis of general flood behavior and spatial analysis of similar reaction regions. The last one is clustering these excessive floods and investigating hybrid and innovative algorithms to reach this goal. The mentioned challenges are addressed as follows:

For the simultaneous occurrence of floods, three forms of analysis are employed to determine clusters where floods reacted similarly. The agglomerate hierarchical clustering, multidimensional scaling for visualization, and a fusion combination of principal component analysis and hierarchical clustering are implemented to illustrate to what extent concurrent floods can enhance risk together. Each of these applied methods has its mathematical concepts, which leads to different results. Both MDS and PCA-AHCT work with the principal components; however, PCA-AHCT operates with PCA residual series, which is an innovative aspect in this field. Consequently, the resulting clusters agree on the main clustering pattern with partly different areas, which are mainly small sub-catchments.

For the second challenge of investigating general flood behavior, different methods are considered from previous chapters. A hybrid application of distribution-based clustering and hierarchical clustering are merged to illustrate the regions where floods reacted similarly in the Neckar catchment. The idea behind this is to group extreme floods by comparing the maximum difference among CDFs of flood series. Further, an innovative and robust optimization scheme for clustering is performed to recognize different groups of sub-catchments that similarly floods occurred. Due to the distinct nature of applied clustering methods does not assume that the final clustered maps have to be the same or primarily similar. In the third chapter, clustering is based on traditional hierarchical clustering. The fourth chapter presented the hierarchical clusters based on residual PCA. Both aforementioned chapters concentrate on simultaneous occurrences of flood events. The fifth and sixth chapters have different objectives on general flood events, which may cover some of the simultaneous flood occurrences. These two chapters performed and introduced distribution-based clustering and optimization-based clustering. The initial estimation of applying different methods has different clustered maps. Therefore, five different clustering algorithms are applied on simultaneous and general flood terms.

The final challenge for clustering is embedded in two previously mentioned challenges. In a way, a total of five newly developed clustering methods have been employed or introduced in this thesis. The used clustering algorithms are compared and validated to determine the commonalities of their pattern/cluster areas concerning the mentioned challenges. Hydrological models usually struggle with extremes, and it is possible to enhance the efficiency of current models by running separately in different clusters.

Algorithms such as the RSA tries to optimize the clustering evaluation criterion for over-fitting and have some type of bias. Therefore, it has to be considered that the clustered map may have an over-fitting feature. So, implementing an internal evaluation method demands understanding the floods and their properties in the catchment. Using different cross-validation approaches, like splitting data into test and validation parts, may solve the mentioned problem. Probably, changing evaluation criteria can show different performances and maps. Moreover, many factors contribute to the cooling and controlling factors, such as other dissimilarity indices, altering the minimum number of points in a cluster, and the probability rate curve.

Likewise, a comprehensive assessment of the catchment properties and data series is investigated to answer and address geophysical issues and the relationship between clusters and basin areas in the second chapter. Almost half of the sub-catchments belong to the upper Neckar, which shows a high concentration of the sub-catchments there due to high elevation. Therefore, expecting to have different clusters in this part of the basin is not far-fetched. However, two different aspects of evaluations expressed the fact that the general extreme floods' clusters followed the mainstream connected sub-basin. The simultaneous occurrences of floods pursued more topographical and geophysics of the catchment. They indicated and separated explicitly the upper Neckar from the rest of the catchment. It is why the investigation of simultaneous occurrence of floods has to be considered in future research separately and substantially. Also, clustering after implementation of PCA is less sensitive to utilizing different distances metrics and linkage methods. Therefore, it can be concluded that it is an efficient approach for further clustering studies on flood analysis in hydrology.

In the second chapter, it is shown that there is an abrupt change point in the late 80's in the discharge time series (see fig. 2.13). It is possible to assess and cluster on two time periods before and after this breaking point, and consequently, different clustering results will be illustrated. Also, different slopes of changes in discharge time series in various sub-catchments in Figure 2.12, can cause instability in clustering for the long-term time series. Moreover, due to the different time distribution of flood events in a year (see fig. 2.9), dividing the whole time series into the wet and cold seasons and performing the same analysis may show different clusters.

All areas alongside the mainstream river show a significant trend in different time intervals (see fig. 2.10), but usually, the upper Neckar is separated in the clustering maps, which do not follow the trend significance's areas. Thus, trend analysis evaluation on the flood series may illustrate a similar pattern as the resultant clusters. Furthermore, there is no precise number of clusters, but there is an optimal one.

The clustering maps for simultaneous floods express that these floods depend on the distance to the outlet. It means the upstream in the upper Neckar is divided into two clusters, and the rest of the catchment is in another cluster. The upper Neckar and all the southern edge of the basin have high elevation topography and are on Jurassic hydrogeological units. In addition, here, the concentration of small size sub-basins is high.

The similar general floods regions are sensitive to the headwaters. Usually, these small subcatchments reacted differently from the others. Also, the Jagst and Kocher in the East and North of the Neckar basin showed different behavior like the three first sub-catchments

of the main Neckar river in Weighted and Ward maps. Also, employing the RSA technique shows that almost all the measurement areas along the main river are grouped as the same cluster. The same as distribution-based clustering, headwater sub-basins are the points that different clusters came up with.

Furthermore, by comparing the annual maxima at the outlet of the Neckar with the other gauges, it is recognized that the upper Neckar is an independent area. This region has its own behavior and does not impact the extremes very much on the outlet. Station on the main river, and Fils tributary, both in the upper Neckar, has distinct cumulative distribution function than the downstream (see Figure 6.10 and 6.9).

In terms of geological units, The Nagold and origin of the Enz river are on the red sandstone. The Würm, which always showed different reactions from neighboring regions, is on shell limestone. Also, in general, western subcatchments and partly the Würm geological formation are similar to the origin of the main river, the Neckar river. It can be a reason why these mountainous sub-basins have similar behaviors. However, floods do not follow the flow regimes and the mentioned factors mainly have a great impact on the high temporal resolution discharge time series.

Another point is, all the southern edges of the Neckar plus the origins of the Jagst and Kocher in the East have upper, middle, and lower Jurassic hydrological unites. These tributaries also have a significant trend in weekly, monthly, and yearly discharge temporal resolution in Figure 2.10. It can be due to possible different water management in these regions. However, as mentioned, the flood time series is utterly independent of the river flow time series. This parameter changes massively the behavior of these areas compared to the rest of the catchment.

The presence of more than four thousand hydraulic structures, massive river regulations, and management have greatly influenced the hydrological analysis. In such a way, one of the central catchment areas, headwater and not on the border of the basin, has the largest number of different structures (Murr sub-basin). Also, the lack of measurement gauges between the outlet of the upper Neckar at the Plochingen (station Nr. 427) and the Rockenau (Nr. 454) may cause some uncertainty in the clustered areas. In between, some of the main Neckar's tributaries connect to the main river, and consequently, it is impossible to investigate more detail.

Bibliography

Abdi, H. and Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459.

Ahmad, M. I., Sinclair, C., and Spurr, B. (1988). Assessment of flood frequency models using empirical distribution function statistics. *Water Resources Research*, 24(8):1323–1328.

Ahmad, Q. (2003). Regional cooperation in flood management in the Ganges-Brahmaputra-Meghna region: Bangladesh perspective. *Natural Hazards*, 28(1):191–198.

Alfieri, L., Salamon, P., Bianchi, A., Neal, J., Bates, P., and Feyen, L. (2014). Advances in pan-European flood hazard mapping. *Hydrological Processes*, 28(13):4067–4077.

Bahrami, S. (2019). *Global Ensemble Streamflow and Flood Modeling with Application of Large Data Analytics, Deep learning and GIS*. PhD thesis, University of Nevada, Reno.

Baldassarre, G. D., Viglione, A., Carr, G., Kuil, L., Salinas, J., and Blöschl, G. (2013). Socio-hydrology: conceptualising human-flood interactions. *Hydrology and Earth System Sciences*, 17(8):3295–3303.

Ballabio, C. (2009). Spatial prediction of soil properties in temperate mountain regions using support vector regression. *Geoderma*, 151(3-4):338–350.

Bandyopadhyay, S., Maulik, U., and Pakhira, M. K. (2001). Clustering using simulated annealing with probabilistic redistribution. *International Journal of Pattern Recognition and Artificial Intelligence*, 15(02):269–285.

Bar-Joseph, Z., Gifford, D. K., and Jaakkola, T. S. (2001). Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics*, 17(suppl_1):S22–S29.

Barata, J. C. A. and Hussein, M. S. (2012). The Moore–Penrose pseudoinverse: A tutorial review of the theory. *Brazilian Journal of Physics*, 42(1-2):146–165.

Bárdossy, A. and Pegram, G. (2014). Infilling missing precipitation records–A comparison of a new Copula-based method with other techniques. *Journal of Hydrology*, 519:1162–1170.

Batool, F. and Hennig, C. (2021). Clustering with the Average Silhouette Width. *Computational Statistics & Data Analysis*, 158:107190.

Bertle, F. A. (1973). Selecting spillway floods of existing structures (bureau of reclamation techniques). In *Inspection, Maintenance and Rehabilitation of Old Dams*, pages 328–336. ASCE.

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A., Merz, B., Arheimer, B., Aronica, G. T., Bilibashi, A., Bonacci, O., Borga, M., et al. (2017). Changing climate shifts timing of European floods. *Science*, 357(6351):588–590.

Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G. T., Bilibashi, A., et al. (2019). Changing climate both increases and decreases European river floods. *Nature*, pages 1–4.

Bormann, H. (2010). Runoff regime changes in German rivers due to climate change. *Erdkunde*, pages 257–279.

Bouguettaya, A. (1996). On-line clustering. *IEEE Transactions on Knowledge and Data Engineering*, 8(2):333–339.

Bouguettaya, A., Yu, Q., Liu, X., Zhou, X., and Song, A. (2015). Efficient agglomerative hierarchical clustering. *Expert Systems with Applications*, 42(5):2785 – 2797.

Brázdil, R., Kundzewicz, Z. W., and Benito, G. (2006). Historical hydrology for studying flood risk in Europe. *Hydrological Sciences Journal*, 51(5):739–764.

Brown, D. E. and Huntley, C. L. (1992). A practical application of simulated annealing to clustering. *Pattern recognition*, 25(4):401–412.

Bui, Q.-T., Nguyen, Q.-H., Nguyen, X. L., Pham, V. D., Nguyen, H. D., and Pham, V.-M. (2020). Verification of novel integrations of swarm intelligence algorithms into deep learning neural network for flood susceptibility mapping. *Journal of Hydrology*, 581:124379.

Bürger, K., Dostal, P., Seidel, J., Imbery, F., Barriendos, M., Mayer, H., and Glaser, R. (2006). Hydrometeorological reconstruction of the 1824 flood event in the Neckar river basin (southwest Germany). *Hydrological Sciences Journal*, 51(5):864–877.

Caliński, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27.

Chaimontree, S., Atkinson, K., and Coenen, F. (2010). Best clustering configuration metrics: towards multiagent based clustering. In *International Conference on Advanced Data Mining and Applications*, pages 48–59. Springer.

Chen, L., Singh, V. P., Shenglian, G., Hao, Z., and Li, T. (2012). Flood coincidence risk analysis using multivariate Copula functions. *Journal of Hydrologic Engineering*, 17(6):742–755.

Corporal-Lodangco, I. L. and Leslie, L. M. (2017). Defining Philippine climate zones using surface and high-resolution satellite data. *Procedia Computer Science*, 114:324–332.

Cox, T. F. and Cox, M. A. (2000). *Multidimensional scaling*. Chapman and hall/CRC Press.

Cunnane, C. (1988). Methods and merits of regional flood frequency analysis. *Journal of Hydrology*, 100(1-3):269–290.

Dalrymple, T. (1960). Flood-frequency analyses, manual of hydrology: Part 3. Technical report, USGPO.

Danziger, Z. (2020). Theil-Sen robust linear regression. *https://www.mathworks.com/matlabcentral/fileexchange/48294-theil-sen-robust-linear-regression*. MATLAB Central File Exchange.

Davidson, I. and Ravi, S. S. (2005). Agglomerative hierarchical clustering with constraints: Theoretical and empirical results. In Jorge, A. M., Torgo, L., Brazdil, P., Camacho, R., and Gama, J., editors, *Knowledge Discovery in Databases: PKDD 2005*, pages 59–70, Berlin, Heidelberg. Springer Berlin Heidelberg.

Davies, D. L. and Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227.

Day, W. H. E. and Edelsbrunner, H. (1984). Efficient algorithms for agglomerative hierarchical clustering methods. *Journal of Classification*, 1(1):7–24.

de Moel, H., van Alphen, J., and Aerts, J. (2009). Flood maps in Europe - methods, availability and use. *Natural Hazards and Earth System Sciences*, 9(2):289–301.

Deng, J. and Pandey, M. (2009). Using partial probability weighted moments and partial maximum entropy to estimate quantiles from censored samples. *Probabilistic Engineering Mechanics*, 24(3):407–417.

Dewan, A., Hu, K., Kamruzzaman, M., and Uddin, M. R. (2019). Chapter eight - Evaluating the spatiotemporal pattern of concentration, aggressiveness and seasonality of precipitation over Bangladesh with time–series Tropical Rainfall Measuring Mission data. In Maggioni, V. and Massari, C., editors, *Extreme Hydroclimatic Events and Multivariate Hazards in a Changing Environment*, pages 191 – 219. Elsevier.

Dey, A., Bhattacharyya, S., Dey, S., Snasel, V., and Hassanien, A. E. (2019). *7. Quantum inspired simulated annealing technique for automatic clustering*, pages 145–166. De Gruyter.

Diana, G. and Tommasi, C. (2002). Cross-validation methods in principal component analysis: a comparison. *Statistical Methods and Applications*, 11(1):71–82.

Diederen, D., Liu, Y., Gouldby, B., Diermanse, F., and Vorogushyn, S. (2019). Stochastic generation of spatially coherent river discharge peaks for continental event-based flood risk assessment. *Natural Hazards and Earth System Sciences*, 19(5):1041–1053.

Dudek, A. (2019). Silhouette index as clustering evaluation tool. In *Conference of the Section on Classification and Data Analysis of the Polish Statistical Association*, pages 19–33. Springer.

Dunn, J. C. (1973). A Fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3):32–57.

Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218.

Ehmele, F. and Kunz, M. (2019). Flood-related extreme precipitation in southwestern Ger-

many: development of a two-dimensional stochastic precipitation model. *Hydrology and Earth System Sciences*, 23(2):1083–1102.

Emblemsvåg, J. (2012). *Risk Management for the Future: Theory and Cases*. IntechOpen.

England Jr, J. F., Cohn, T. A., Faber, B. A., Stedinger, J. R., Thomas Jr, W. O., Veilleux, A. G., Kiang, J. E., and Mason Jr, R. R. (2019). Guidelines for determining flood flow frequency–bulletin 17c. Technical report, Reston, VA.

European Environment Agency (2019). Economic losses from climate-related extremes in Europe (temporal coverage 1980- 2017).

Falter, D., Schröter, K., Dung, N. V., Vorogushyn, S., Kreibich, H., Hundecha, Y., Apel, H., and Merz, B. (2015). Spatially coherent flood risk assessment based on long-term continuous simulation with a coupled model chain. *Journal of Hydrology*, 524:182 – 193.

Farris, J. S. (1969). On the cophenetic correlation coefficient. *Systematic Zoology*, 18(3):279–285.

Favre, A.-C., El Adlouni, S., Perreault, L., Thiémonge, N., and Bobée, B. (2004). Multivariate hydrological frequency analysis using Copulas. *Water Resources Research*, 40(1).

Fowlkes, E. B. and Mallows, C. L. (1983). A method for comparing two hierarchical clusterings. *Journal of the American statistical association*, 78(383):553–569.

Geertsema, T. J., Teuling, A. J., Uijlenhoet, R., Torfs, P. J., and Hoitink, A. J. (2018). Anatomy of simultaneous flood peaks at a lowland confluence. *Hydrology and Earth System Sciences*, 22(10):5599–5613.

Glaser, R., Riemann, D., Schönbein, J., Barriendos, M., Brázdil, R., Bertolin, C., Camuffo, D., Deutsch, M., Dobrovolnỳ, P., van Engelen, A., et al. (2010). The variability of European floods since ad 1500. *Climatic Change*, 101(1-2):235–256.

Golub, G. and Kahan, W. (1965). Calculating the singular values and pseudo-inverse of a matrix. *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis*, 2(2):205–224.

Greene, C. A., Thirumalai, K., Kearney, K. A., Delgado, J. M., Schwanghart, W., Wolfenbarger, N. S., Thyng, K. M., Gwyther, D. E., Gardner, A. S., and Blankenship, D. D. (2019). The climate data toolbox for MATLAB. *Geochemistry, Geophysics, Geosystems*.

Greis, N. P. and Wood, E. F. (1981). Regional flood frequency estimation and network design. *Water Resources Research*, 17(4):1167–1177.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. (2012). A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773.

Götzinger, J., Barthel, R., Jagelke, J., and Bárdossy, A. (2008). The role of groundwater recharge and baseflow in integrated models. In *Proceedings of Symposium HS1002 at IUGG2007, Perugia, July 2007*, volume 321, pages 103–109. IAHS Publication.

Haigh, I. D., Wadey, M. P., Wahl, T., Ozsoy, O., Nicholls, R. J., Brown, J. M., Horsburgh, K., and Gouldby, B. (2016). Spatial and temporal analysis of extreme sea level and storm surge events around the coastline of the UK. *Scientific data*, 3:160107.

Hall, J. and Blöschl, G. (2018). Spatial patterns and characteristics of flood seasonality in Europe. *Hydrology and Earth System Sciences*, 22(7):3883–3901.

Hamzadayi, A. and Yildiz, G. (2013). A simulated annealing algorithm based approach for balancing and sequencing of mixed-model u-lines. *Computers & Industrial Engineering*, 66(4):1070–1084.

Hao, Z., Singh, V. P., and Hao, F. (2018). Compound extremes in hydroclimatology: a review. *Water*, 10(6):718.

Hattermann, F. F., Kundzewicz, Z. W., Huang, S., Vetter, T., Kron, W., Burghoff, O., Merz, B., Bronstert, A., Krysanova, V., and Gerstengarbe, F.-W. (2012). *Flood Risk from a Holistic Perspective–Observed Changes in Germany*, chapter 11. CRC Press.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):497–546.

Hintze, J. (2019). *Multidimensional Scaling, Help Documentation, Chapter 435. NCSS 2019 Statistical Software*. NCSS, LLC., Kaysville, Utah, USA.

Holgersson, M. (1978). The limited value of cophenetic correlation as a clustering criterion. *Pattern Recognition*, 10(4):287–295.

Hosseini, F. S., Choubin, B., Mosavi, A., Nabipour, N., Shamshirband, S., Darabi, H., and Haghighi, A. T. (2020). Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: Application of the simulated annealing feature selection method. *Science of The Total Environment*, 711:135161.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of classification*, 2(1):193–218.

Hudson, A. R., Alfaro-Sanchez, R., Babst, F., Belmecheri, S., Moore, D. J., and Trouet, V. (2019). Seasonal and synoptic climatic drivers of tree growth in the Bighorn mountains, WY, USA (1654–1983 CE). *Dendrochronologia*, 58:125633.

Hudson, P. F. and Colditz, R. R. (2003). Flood delineation in a large and complex alluvial valley, lower Panuco basin, Mexico. *Journal of Hydrology*, 280(1-4):229–245.

Jackson, D. A. (1993). Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology*, 74(8):2204–2214.

Jagelke, J. and Barthel, R. (2005). Conceptualization and implementation of a regional groundwater model for the Neckar catchment in the framework of an integrated regional model. *Advances in Geosciences*, 5:105–111.

Jain, A. K. and Dubes, R. C. (1988). Algorithms for clustering data. *Englewood Cliffs: Prentice Hall, 1988*.

Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323.

Jalón-Rojas, I., Schmidt, S., and Sottolichio, A. (2016). Evaluation of spectral methods for high-frequency multiannual time series in coastal transitional waters: advantages of combined analyses. *Limnology and Oceanography: Methods*, 14(6):381–396.

Johnson, R. M. (1963). On a theorem stated by Eckart and Young. *Psychometrika*, 28(3):259–263.

Jolliffe, I. T. (2002). *Principal component analysis*. Springer, 2 edition.

Josse, J. and Husson, F. (2012). Selecting the number of components in principal component analysis using cross-validation approximations. *Computational Statistics & Data Analysis*, 56(6):1869–1879.

Kajambeu, R., Sigauke, C., Bere, A., Chikobvu, D., Maposa, D., and Maurel, M. (2020). Probabilistic flood height estimation of the Limpopo River at the Beitbridge using r-largest order statistics. *Appl. Math*, 14(2):191–204.

Kalantar, B., Ueda, N., Saeidi, V., Janizadeh, S., Shabani, F., Ahmadi, K., and Shabani, F. (2021). Deep neural network utilizing remote sensing datasets for flood hazard susceptibility mapping in Brisbane, Australia. *Remote Sensing*, 13(13):2638.

Kalweit, H., Buck, W., and auf das Abflussregime, A. A. E. (1993). *Der Rhein unter der Einwirkung des Menschen: Ausbau, Schiffahrt, Wasserwirtschaft*. Secretariaat CHR.

Kang, L., Jiang, S., Hu, X., and Li, C. (2019). Evaluation of return period and risk in bivariate non-stationary flood frequency analysis. *Water*, 11(1):79.

Karamizadeh, S., Abdullah, S. M., Manaf, A. A., Zamani, M., and Hooman, A. (2013). An overview of principal component analysis. *Journal of Signal and Information Processing*, 4(3B):173.

Katzenbeisser, W. and Hackl, P. (1986). An alternative to the kolmogrov-smirnov two-sample test. *Communications in Statistics-Theory and Methods*, 15(4):1163–1177.

Kaufman, L. and Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons.

Kendall, M. G. (1948). *Rank correlation methods*. Charles Griffin & Co. Ltd., London.

Killick, R., Fearnhead, P., and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598.

Kim, J. H. et al. (2014). Meta-heuristic algorithms as tools for hydrological science. *Geoscience Letters*, 1(1):1–7.

King, J. R. and Jackson, D. A. (1999). Variable selection in large environmental data sets using principal components analysis. *Environmetrics: The official journal of the International Environmetrics Society*, 10(1):67–77.

Kingrani, S. K., Levene, M., and Zhang, D. (2018). Estimating the number of clusters using diversity. *Artificial Intelligence Research*, 7(1):15–22.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *science*, 220(4598):671–680.

Klein, R. W. and Dubes, R. C. (1989). Experiments in projection and clustering by simulated annealing. *Pattern Recognition*, 22(2):213–220.

Koradia, A., Bhalala, A., and Tiwari, M. (2019). Rainfall-runoff simulation modelling using artificial neural networks in semi-arid middle Gujarat region. *Indian Journal of Soil Conservation*, 47(3):231–238.

Kourgialas, N. N., Dokou, Z., and Karatzas, G. P. (2015). Statistical analysis and ann modeling for predicting hydrological extremes under climate change scenarios: The example of a small Mediterranean agro-watershed. *Journal of environmental management*, 154:86–101.

Kruskal, J. B. and Wish, M. (1978). *Multidimensional scaling (quantitative applications in the social sciences)*. Sage Publications, Inc.

Krzanowski, W. (1987). Cross-validation in principal component analysis. *Biometrics*, pages 575–584.

Kuiper, N. H. (1960). Tests concerning random points on a circle. In *Nederl. Akad. Wetensch. Proc. Ser. A*, volume 63, pages 38–47.

Kundzewicz, Z. W. (2019). *Changes in flood risk in Europe*. CRC Press.

Lavielle, M. (2005). Using penalized contrasts for the change-point problem. *Signal processing*, 85(8):1501–1510.

Lehmann, E. L. (1951). Consistency and unbiasedness of certain nonparametric tests. *The annals of mathematical statistics*, pages 165–179.

Lengyel, A. and Botta-Dukát, Z. (2019). Silhouette width using generalized mean—A flexible method for assessing clustering efficiency. *Ecology and evolution*, 9(23):13231–13243.

Leonard, M., Westra, S., Phatak, A., Lambert, M., van den Hurk, B., McInnes, K., Risbey, J., Schuster, S., Jakob, D., and Stafford-Smith, M. (2014). A compound event framework for understanding extreme impacts. *Wiley Interdisciplinary Reviews: Climate Change*, 5(1):113–128.

Leščešen, I. and Dolinaj, D. (2019). Regional flood frequency analysis of the Pannonian basin. *Water*, 11(2).

Liu, Y., Li, Z., Xiong, H., Gao, X., and Wu, J. (2010). Understanding of internal clustering validation measures. In *2010 IEEE international conference on data mining*, pages 911–916. IEEE.

Lodhi, P., Mishra, O., and Rajpoot, D. S. (2019). Sorted outlier detection approach based

on Silhouette coefficient. In *Advances in Signal Processing and Communication*, pages 187–197. Springer.

Lohre, M., Sibbertsen, P., and Könning, T. (2003). Modeling water flow of the Rhine River using seasonal long memory. *Water Resources Research*, 39(5).

LUBW (2021). Geobasisdaten Landesamt für Geoinformation und Landentwicklung Baden-Württemberg. www.lhl-bw.de, Az:2851.9-1/19.

Łukasik, S., Kowalski, P. A., Charytanowicz, M., and Kulczycki, P. (2016). Clustering using flower pollination algorithm and Calinski-Harabasz index. In *2016 IEEE Congress on Evolutionary Computation (CEC)*, pages 2724–2728. IEEE.

Lyra, G. B., Oliveira-Júnior, J. F., and Zeri, M. (2014). Cluster analysis applied to the spatial and temporal variability of monthly rainfall in Alagoas state, northeast of Brazil. *International Journal of Climatology*, 34(13):3546–3558.

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica: Journal of the econometric society*, pages 245–259.

Martínez, C., Cienfuegos, R., Inzunza, S., Urrutia, A., and Guerrero, N. (2020). Worst-case tsunami scenario in Cartagena Bay, central Chile: Challenges for coastal risk management. *Ocean & Coastal Management*, 185:105060.

Massey Jr, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78.

Maulik, U. and Bandyopadhyay, S. (2002). Performance evaluation of some clustering algorithms and validity indices. *IEEE Transactions on pattern analysis and machine intelligence*, 24(12):1650–1654.

McPhillips, L. E., Chang, H., Chester, M. V., Depietri, Y., Friedman, E., Grimm, N. B., Kominoski, J. S., McPhearson, T., Méndez-Lázaro, P., Rosi, E. J., et al. (2018). Defining extreme events: A cross-disciplinary review. *Earth's Future*, 6(3):441–455.

Meade, R. H., Rayol, J. M., Da Conceicão, S. C., and Natividade, J. R. (1991). Backwater effects in the Amazon river basin of Brazil. *Environmental Geology and Water Sciences*, 18(2):105–114.

Meilă, M. (2003). Comparing clusterings by the variation of information. In Schölkopf, B. and Warmuth, M. K., editors, *Learning Theory and Kernel Machines*, pages 173–187, Berlin, Heidelberg. Springer Berlin Heidelberg.

Milliner, C., Materna, K., Bürgmann, R., Fu, Y., Moore, A. W., Bekaert, D., Adhikari, S., and Argus, D. F. (2018). Tracking the weight of hurricane Harvey's stormwater using GPS data. *Science advances*, 4(9):eaau2477.

Ming, F., Yang, Y., Zeng, A., and Zhao, B. (2017). Spatiotemporal filtering for regional GPS network in China using independent component analysis. *Journal of Geodesy*, 91(4):419–440.

Mirza, M. M. Q. (2002). Global warming and changes in the probability of occurrence of floods in Bangladesh and implications. *Global environmental change*, 12(2):127–138.

Mirza, M. M. Q. (2003). Three recent extreme floods in Bangladesh: a hydro-meteorological analysis. In *Flood problem and management in South Asia*, pages 35–64. Springer.

Modaresi Rad, A. and Khalili, D. (2015). Appropriateness of clustered raingauge stations for spatio-temporal meteorological drought applications. *Water Resources Management*, 29(11):4157–4171.

Modiri, E. (2021). Towards new applications of spaceborne technology on flood protection. United Nations Office for Outer Space Affairs | Space4water.

Modiri, E. and Bárdossy, A. (2018). Clustering the extreme flood magnitude based on dependency structure in upper Neckar catchment. In *EGU General Assembly Conference Abstracts*, volume 20, page 8841.

Modiri, E. and Bárdossy, A. (2019a). Identifying joint peaks behavior in neighboring catchments. In *27th IUGG General Assembly*, Montreal, Canada. International Union of Geodesy and Geophysics.

Modiri, E. and Bárdossy, A. (2019b). Identifying the simultaneous extreme flood behavior in the Neckar catchment. In *EGU General Assembly Conference Abstracts*, volume 21, page 1251.

Modiri, E. and Bárdossy, A. (2020). A robust simulated annealing technique for flood clustering. In *AGU Fall Meeting Abstracts*, volume 2020, pages H203–05.

Modiri, E. and Bárdossy, A. (2020). Spatio-temporal determination of the similarity of extreme floods in the Neckar catchment. In *EGU General Assembly Conference Abstracts*, page 10008.

Modiri, E. and Bárdossy, A. (2021). Clustering simultaneous occurrences of the extreme floods in the Neckar catchment. *Water*, 13(4):399.

Moeeni, H., Bonakdari, H., and Fatemi, S. E. (2017). Stochastic model stationarization by eliminating the periodic term and its effect on time series prediction. *Journal of Hydrology*, 547:348–364.

Mora-López, L. and Mora, J. (2015). An adaptive algorithm for clustering cumulative probability distribution functions using the Kolmogorov–Smirnov two-sample test. *Expert Systems with Applications*, 42(8):4016–4021.

MunichRe (2019). Flood / flash flood events in Germany 1980 – 2018. Technical report, NatCatSERVICE, Münchener Rückversicherungs-Gesellschaft “Munich Re”.

Murtagh, F. and Contreras, P. (2011). Methods of hierarchical clustering. *arXiv preprint arXiv:1105.0121*.

Murtagh, F. and Contreras, P. (2012). Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1):86–97.

Murtagh, F. and Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? *Journal of Classification*, 31(3):274–295.

Nemec, A. and Brinkhurst, R. (1988). The Fowlkes–Mallows statistic and the comparison of two independently determined dendrograms. *Canadian Journal of Fisheries and Aquatic Sciences*, 45(6):971–975.

Newton, D. W. (1983). Realistic assessment of maximum flood potentials. *Journal of Hydraulic Engineering*, 109(6):905–918.

Nikitina, E., Ostrovskaya, E., and Fomenko, M. (2010). Towards better water governance in river basins: some lessons learned from the Volga. *Regional Environmental Change*, 10(4):285–297.

Palop, J. J., Mucke, L., and Roberson, E. D. (2010). Quantifying biomarkers of cognitive dysfunction and neuronal network hyperexcitability in mouse models of Alzheimer's disease: depletion of calcium-dependent proteins and inhibitory hippocampal remodeling. In *Alzheimer's Disease and Frontotemporal Dementia*, pages 245–262. Springer.

Pauwels, E. J. and Frederix, G. (2000). Image segmentation by nonparametric clustering based on the Kolmogorov-Smirnov distance. In *European Conference on Computer Vision*, pages 85–99. Springer.

Pedrozo-Acuña, A., Breña-Naranjo, J. A., and Domínguez-Mora, R. (2014). The hydrological setting of the 2013 floods in Mexico. *Weather*, 69(11):295–302.

Peng, H., Wang, S., and Wang, X. (2008). Consistency and asymptotic distribution of the Theil–Sen estimator. *Journal of Statistical Planning and Inference*, 138(6):1836–1850.

Penrose, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(3):406–413.

Petrovic, S. (2006). A comparison between the Silhouette index and the Davies-Bouldin index in labelling IDS clusters. In *Proceedings of the 11th Nordic Workshop of Secure IT Systems*, volume 2006, pages 53–64. Citeseer.

Petrow, T., Thieken, A. H., Kreibich, H., Merz, B., and Bahlburg, C. H. (2006). Improvements on flood alleviation in Germany: Lessons learned from the Elbe flood in August 2002. *Environmental management*, 38(5):717–732.

Petts, G. E. (1999). *River regulation*, pages 521–528. Springer Netherlands, Dordrecht.

Prohaska, S., Ilic, A., and Majkic, B. (2008). Multiple-coincidence of flood waves on the Main river and its tributaries. In *IOP conference series: Earth and environmental science*, volume 4, page 012013. IOP Publishing.

Radziejewski, M., Bardossy, A., and Kundzewicz, Z. W. (2000). Detection of change in river flow using phase randomization. *Hydrological Sciences Journal*, 45(4):547–558.

Rahman, A. S. and Rahman, A. (2020). Application of principal component analysis and cluster analysis in regional flood frequency analysis: A case study in New South Wales, Australia. *Water*, 12(3):781.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850.

Reis Jr, D. S. and Stedinger, J. R. (2005). Bayesian MCMC flood frequency analysis with historical information. *Journal of hydrology*, 313(1-2):97–116.

Rendón, E., Abundez, I., Arizmendi, A., and Quiroz, E. M. (2011a). Internal versus external cluster validation indexes. *International Journal of computers and communications*, 5(1):27–34.

Rendón, E., Abundez, I. M., Gutierrez, C., Zagal, S. D., Arizmendi, A., Quiroz, E. M., and Arzate, H. E. (2011b). A comparison of internal and external cluster validation indexes. In *Proceedings of the 2011 American Conference, San Francisco, CA, USA*, volume 29, pages 1–10.

Rosbjerg, D. and Madsen, H. (1995). Uncertainty measures of regional flood frequency estimators. *Journal of Hydrology*, 167(1-4):209–224.

Rosenblatt, M. (1952). Limit theorems associated with variants of the von Mises statistic. *The Annals of Mathematical Statistics*, pages 617–623.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

Saf, B. (2009). Regional flood frequency analysis using L-moments for the West Mediterranean region of Turkey. *Water Resources Management*, 23(3):531–551.

Santos, C. A. G., Moura, R., da Silva, R. M., and Costa, S. G. F. (2019). Cluster analysis applied to spatiotemporal variability of monthly precipitation over Paraíba state using tropical rainfall measuring mission (TRMM) data. *Remote Sensing*, 11(6):637.

Schmidt, R., Petrovic, S., Güntner, A., Barthelmes, F., Wünsch, J., and Kusche, J. (2008). Periodic components of water storage changes from GRACE and global hydrology models. *Journal of Geophysical Research: Solid Earth*, 113(B8).

Schröter, K., Kunz, M., Elmer, F., Mühr, B., and Merz, B. (2015). What made the June 2013 flood in Germany an exceptional event? A hydro-meteorological evaluation. *Hydrology and Earth System Sciences*, 19(1):309–327.

Schwanghart, W. and Kuhn, N. J. (2010). Topotoolbox: A set of Matlab functions for topographic analysis. *Environmental Modelling & Software*, 25(6):770–781.

Schwanghart, W. and Scherler, D. (2014). Topotoolbox 2–MATLAB-based software for topographic analysis and modeling in Earth surface sciences. *Earth Surface Dynamics*, 2(1):1–7.

Schwanghart, W. and Scherler, D. (2017). Bumps in river profiles: uncertainty assessment and smoothing using quantile regression techniques. *Earth Surface Dynamics*, 5(4):821–839.

Searcy, J. K. (1959). *Flow-duration curves*. Number 1542. US Government Printing Office.

Seber, G. (1984). *Multivariate observations*. John Wiley & Sons, New York.

Seidel, J., Dostal, P., and Imbery, F. (2012). Analysis of historical river floods-A contribution towards modern flood risk management. In Emblemsvåg, J., editor, *Risk Management for the Future-Theory and Cases*, chapter 12, pages 275–294. IntechOpen.

Seifollahi, S., Bagirov, A., Zare Borzeshi, E., and Piccardi, M. (2019). A simulated annealing-based maximum-margin clustering algorithm. *Computational Intelligence*, 35(1):23–41.

Selaman, O. S., Said, S., and Putuhena, F. (2007). Flood frequency analysis for Sarawak using Weibull, Gringorten and L-moments formula. *J. Inst. Eng*, 68:43–52.

Selim, S. Z. and Alsultan, K. (1991). A simulated annealing algorithm for the clustering problem. *Pattern recognition*, 24(10):1003–1008.

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

Shutaywi, M. and Kachouie, N. N. (2021). Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy*, 23(6):759.

Sokal, R. R. and Rohlf, F. J. (1962). The comparison of dendrograms by objective methods. *Taxon*, 11(2):33–40.

Stedinger, J. R. (1983). Estimating a regional flood frequency distribution. *Water Resources Research*, 19(2):503–510.

Stojković, M., Plavšić, J., and Prohaska, S. (2017). Annual and seasonal discharge prediction in the middle Danube River basin based on a modified TIPS (Tendency, Intermittency, Periodicity, Stochasticity) methodology. *Journal of Hydrology and Hydromechanics*, 65(2):165–174.

Tan, W., Chen, J., Dong, D., Qu, W., and Xu, X. (2020). Analysis of the potential contributors to common mode error in Chuandian region of China. *Remote Sensing*, 12(5):751.

Theil, H. (1992). A rank-invariant method of linear and polynomial regression analysis. In *Henri Theil's contributions to economics and econometrics*, pages 345–381. Springer.

Thieken, A. H. (2009). Floods, flood losses and flood risk management in Germany.

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, 18(4):267–276.

Thornton, P. K., Jones, P. G., Owiyo, T., Kruska, R., Herrero, M. T., Kristjanson, P. M., Notenbaert, A. M. O., Bekele, N., and Omolo, A. (2006). Mapping climate vulnerability and poverty in Africa. Technical report, International Livestock Research Institute (ILRI).

Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423.

Torgerson, W. S. (1952). Multidimensional scaling: I. theory and method. *Psychometrika*, 17(4):401–419.

Tsivtsivadze, N., Bregvadze, G., Laghidze, L., Trapaidze, V., and Motsonelidze, N. I. (2019).

Measures of flood risk mitigation in downsream of river Rioni. In *19th International Multidisciplinary Scientific GeoConference SGEM 2019*, volume 19, pages 355–362.

Uhlemann, S., Thieken, A. H., and Merz, B. (2010). A consistent set of trans-basin floods in Germany between 1952-2002. *Hydrology and Earth System Sciences*, 14(7):1277–1295.

Unal, Y., Kindap, T., and Karaca, M. (2003). Redefining the climate zones of Turkey using cluster analysis. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 23(9):1045–1055.

United Nations (1964). Manual of standards and criteria for planning water resource projects. *Water Resources Series*, (26).

Ünlü, R. and Xanthopoulos, P. (2019). Estimating the number of clusters in a dataset via consensus clustering. *Expert Systems with Applications*, 125:33–39.

van Oldenborgh, G. J., Philip, S., Aalbers, E., Vautard, R., Otto, F., Haustein, K., Habets, F., Singh, R., and Cullen, H. (2016). Rapid attribution of the May/June 2016 flood-inducing precipitation in France and Germany to climate change. *Hydrol Earth Syst Sci Discuss*, 10.

Vogel, R. M. and Fennessey, N. M. (1994). Flow-duration curves. I: New interpretation and confidence intervals. *Journal of Water Resources Planning and Management*, 120(4):485–504.

Vogel, R. M. and Fennessey, N. M. (1995). Flow duration curves II: A review of applications in water resources planning 1. *JAWRA Journal of the American Water Resources Association*, 31(6):1029–1039.

Wagner, S. and Wagner, D. (2007). *Comparing clusterings: an overview*. Universität Karlsruhe, Fakultät für Informatik Karlsruhe.

Wang, T., Li, Q., Bucci, D. J., Liang, Y., Chen, B., and Varshney, P. K. (2019). K-medoids clustering of data sequences with composite distributions. *IEEE Transactions on Signal Processing*, 67(8):2093–2106.

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244.

Williams, A. (2016). Everything you did and didn't know about PCA.

Wyncoll, D. and Gouldby, B. (2015). Integrating a multivariate extreme value method within a system flood risk analysis model. *Journal of Flood Risk Management*, 8(2):145–160.

Yinkang, Z. (1996). A preliminary study on the characteristics of floods in Huaihe river basin. *Geographical Research*, 1.

Yokoo, Y. and Sivapalan, M. (2011). Towards reconstruction of the flow duration curve: development of a conceptual framework with a physical basis. *Hydrology and Earth System Sciences*, 15(9):2805–2819.

Yuan, P., Li, Z., Jiang, W., Ma, Y., Chen, W., and Sneeuw, N. (2018). Influences of environmental loading corrections on the nonlinear variations and velocity uncertainties for the reprocessed global positioning system height time series of the crustal movement observation network of China. *Remote Sensing*, 10(6):958.

Zahn, C. T. (1971). Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Transactions on Computers*, 100(1):68–86.

Zhang, B. and Chen, R. (2018). Nonlinear time series clustering based on Kolmogorov-Smirnov 2d statistic. *Journal of Classification*, 35(3):394–421.

Zhang, Q., Gu, X., Singh, V. P., Xiao, M., and Chen, X. (2015). Evaluation of flood frequency under non-stationarity resulting from climate indices and reservoir indices in the East River basin, China. *Journal of Hydrology*, 527:565–575.

Zhou, T., Liu, Z., Jin, J., and Hu, H. (2019). Assessing the impacts of univariate and bivariate flood frequency approaches to flood risk accounting for reservoir operation. *Water*, 11(3):475.

Zhu, Y., Deng, Q., Huang, D., Jing, B., and Zhang, B. (2021). Clustering based on Kolmogorov–Smirnov statistic with application to bank card transaction data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*.

Institut für Wasser- und Umweltsystemmodellierung
Universität Stuttgart

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

Verzeichnis der Mitteilungshefte

5 Plate, Erich: *Beitrag zur Bestimmung der Windgeschwindigkeitsverteilung in der durch eine Wand gestörten bodennahen Luftschicht*, und
Röhnisch, Arthur; Marotz, Günter: *Neue Baustoffe und Bauausführungen für den Schutz der Böschungen und der Sohle von Kanälen, Flüssen und Häfen; Gestehungskosten und jeweilige Vorteile*, sowie
Unny, T.E.: *Schwingungsuntersuchungen am Kegelstrahlschieber*, 1967

6 Seiler, Erich: *Die Ermittlung des Anlagenwertes der bundeseigenen Binnenschiffahrtsstraßen und Talsperren und des Anteils der Binnenschiffahrt an diesem Wert*, 1967

7 *Sonderheft anläßlich des 65. Geburtstages von Prof. Arthur Röhnisch mit Beiträgen von* Benk, Dieter; Breitling, J.; Gurr, Siegfried; Haberhauer, Robert; Honekamp, Hermann; Kuz, Klaus Dieter; Marotz, Günter; Mayer-Vorfelder, Hans-Jörg; Miller, Rudolf; Plate, Erich J.; Radomski, Helge; Schwarz, Helmut; Vollmer, Ernst; Wildenhahn, Eberhard; 1967

8 Jumikis, Alfred: *Beitrag zur experimentellen Untersuchung des Wassernachschubs in einem gefrierenden Boden und die Beurteilung der Ergebnisse*, 1968

9 Marotz, Günter: *Technische Grundlagen einer Wasserspeicherung im natürlichen Untergrund*, 1968

10 Radomski, Helge: *Untersuchungen über den Einfluß der Querschnittsform wellenförmiger Spundwände auf die statischen und rammtechnischen Eigenschaften*, 1968

11 Schwarz, Helmut: *Die Grenztragfähigkeit des Baugrundes bei Einwirkung vertikal gezogener Ankerplatten als zweidimensionales Bruchproblem*, 1969

12 Erbel, Klaus: *Ein Beitrag zur Untersuchung der Metamorphose von Mittelgebirgsschneedecken unter besonderer Berücksichtigung eines Verfahrens zur Bestimmung der thermischen Schneequalität*, 1969

13 Westhaus, Karl-Heinz: *Der Strukturwandel in der Binnenschiffahrt und sein Einfluß auf den Ausbau der Binnenschiffskanäle*, 1969

14 Mayer-Vorfelder, Hans-Jörg: *Ein Beitrag zur Berechnung des Erdwiderstandes unter Ansatz der logarithmischen Spirale als Gleitflächenfunktion*, 1970

15 Schulz, Manfred: *Berechnung des räumlichen Erddruckes auf die Wandung kreiszylindrischer Körper*, 1970

16 Mobasseri, Manoutschehr: *Die Rippenstützmauer. Konstruktion und Grenzen ihrer Standsicherheit*, 1970

17 Benk, Dieter: *Ein Beitrag zum Betrieb und zur Bemessung von Hochwasserrückhaltebecken*, 1970

18 Gàl, Attila: *Bestimmung der mitschwingenden Wassermasse bei überströmten Fischbauchklappen mit kreiszylindrischem Staublech*, 1971, vergriffen

19 Kuz, Klaus Dieter: *Ein Beitrag zur Frage des Einsetzens von Kavitationserscheinungen in einer Düsenströmung bei Berücksichtigung der im Wasser gelösten Gase*, 1971, vergriffen

20 Schaak, Hartmut: *Verteilleitungen von Wasserkraftanlagen*, 1971

21 *Sonderheft zur Eröffnung der neuen Versuchsanstalt des Instituts für Wasserbau der Universität Stuttgart mit Beiträgen von* Brombach, Hansjörg; Dirksen, Wolfram; Gàl, Attila; Gerlach, Reinhard; Giesecke, Jürgen; Holthoff, Franz-Josef; Kuz, Klaus Dieter; Marotz, Günter; Minor, Hans-Erwin; Petrikat, Kurt; Röhnisch, Arthur; Rueff, Helge; Schwarz, Helmut; Vollmer, Ernst; Wildenhahn, Eberhard; 1972

22 Wang, Chung-su: *Ein Beitrag zur Berechnung der Schwingungen an Kegelstrahlschiebern*, 1972

23 Mayer-Vorfelder, Hans-Jörg: *Erdwiderstandsbeiwerte nach dem Ohde-Variationsverfahren*, 1972

24 Minor, Hans-Erwin: *Beitrag zur Bestimmung der Schwingungsanfachungsfunktionen überströmter Stauklappen*, 1972, vergriffen

25 Brombach, Hansjörg: *Untersuchung strömungsmechanischer Elemente (Fluidik) und die Möglichkeit der Anwendung von Wirbelkammerelementen im Wasserbau*, 1972, vergriffen

26 Wildenhahn, Eberhard: *Beitrag zur Berechnung von Horizontalfilterbrunnen*, 1972

27 Steinlein, Helmut: *Die Eliminierung der Schwebstoffe aus Flußwasser zum Zweck der unterirdischen Wasserspeicherung, gezeigt am Beispiel der Iller*, 1972

28 Holthoff, Franz Josef: *Die Überwindung großer Hubhöhen in der Binnenschiffahrt durch Schwimmerhebewerke*, 1973

29 Röder, Karl: *Einwirkungen aus Baugrundbewegungen auf trog- und kastenförmige Konstruktionen des Wasser- und Tunnelbaues*, 1973

30 Kretschmer, Heinz: *Die Bemessung von Bogenstaumauern in Abhängigkeit von der Talform*, 1973

31 Honekamp, Hermann: *Beitrag zur Berechnung der Montage von Unterwasserpipelines*, 1973

32 Giesecke, Jürgen: *Die Wirbelkammertriode als neuartiges Steuerorgan im Wasserbau*, und Brombach, Hansjörg: *Entwicklung, Bauformen, Wirkungsweise und Steuereigenschaften von Wirbelkammerverstärkern*, 1974

33 Rueff, Helge: *Untersuchung der schwingungserregenden Kräfte an zwei hintereinander angeordneten Tiefschützen unter besonderer Berücksichtigung von Kavitation*, 1974

34 Röhnisch, Arthur: *Einpreßversuche mit Zementmörtel für Spannbeton - Vergleich der Ergebnisse von Modellversuchen mit Ausführungen in Hüllwellrohren*, 1975

35 *Sonderheft anläßlich des 65. Geburtstages von Prof. Dr.-Ing. Kurt Petrikat mit Beiträgen von*: Brombach, Hansjörg; Erbel, Klaus; Flinspach, Dieter; Fischer jr., Richard; Gàl, Attila; Gerlach, Reinhard; Giesecke, Jürgen; Haberhauer, Robert; Hafner Edzard; Hausenblas, Bernhard; Horlacher, Hans-Burkhard; Hutarew, Andreas; Knoll, Manfred; Krummet, Ralph; Marotz, Günter; Merkle, Theodor; Miller, Christoph; Minor, Hans-Erwin; Neumayer, Hans; Rao, Syamala; Rath, Paul; Rueff, Helge; Ruppert, Jürgen; Schwarz, Wolfgang; Topal-Gökceli, Mehmet; Vollmer, Ernst; Wang, Chung-su; Weber, Hans-Georg; 1975

36 Berger, Jochum: *Beitrag zur Berechnung des Spannungszustandes in rotationssymmetrisch belasteten Kugelschalen veränderlicher Wandstärke unter Gas- und Flüssigkeitsdruck durch Integration schwach singulärer Differentialgleichungen*, 1975

37 Dirksen, Wolfram: *Berechnung instationärer Abflußvorgänge in gestauten Gerinnen mittels Differenzenverfahren und die Anwendung auf Hochwasserrückhaltebecken*, 1976

38 Horlacher, Hans-Burkhard: *Berechnung instationärer Temperatur- und Wärmespannungsfelder in langen mehrschichtigen Hohlzylindern*, 1976

39 Hafner, Edzard: *Untersuchung der hydrodynamischen Kräfte auf Baukörper im Tiefwasserbereich des Meeres*, 1977, ISBN 3-921694-39-6

40 Ruppert, Jürgen: *Über den Axialwirbelkammerverstärker für den Einsatz im Wasserbau*, 1977, ISBN 3-921694-40-X

41 Hutarew, Andreas: *Beitrag zur Beeinflußbarkeit des Sauerstoffgehalts in Fließgewässern an Abstürzen und Wehren*, 1977, ISBN 3-921694-41-8, *vergriffen*

42 Miller, Christoph: *Ein Beitrag zur Bestimmung der schwingungserregenden Kräfte an unterströmten Wehren*, 1977, ISBN 3-921694-42-6

43 Schwarz, Wolfgang: *Druckstoßberechnung unter Berücksichtigung der Radial- und Längsverschiebungen der Rohrwandung*, 1978, ISBN 3-921694-43-4

44 Kinzelbach, Wolfgang: *Numerische Untersuchungen über den optimalen Einsatz variabler Kühlsysteme einer Kraftwerkskette am Beispiel Oberrhein*, 1978, ISBN 3-921694-44-2

45 Barczewski, Baldur: *Neue Meßmethoden für Wasser-Luftgemische und deren Anwendung auf zweiphasige Auftriebsstrahlen*, 1979, ISBN 3-921694-45-0

46 Neumayer, Hans: *Untersuchung der Strömungsvorgänge in radialen Wirbelkammerverstärkern*, 1979, ISBN 3-921694-46-9

47 Elalfy, Youssef-Elhassan: *Untersuchung der Strömungsvorgänge in Wirbelkammerdioden und -drosseln*, 1979, ISBN 3-921694-47-7

48 Brombach, Hansjörg: *Automatisierung der Bewirtschaftung von Wasserspeichern*, 1981, ISBN 3-921694-48-5

49 Geldner, Peter: *Deterministische und stochastische Methoden zur Bestimmung der Selbstdichtung von Gewässern*, 1981, ISBN 3-921694-49-3, *vergriffen*

50 Mehlhorn, Hans: *Temperaturveränderungen im Grundwasser durch Brauchwassereinleitungen*, 1982, ISBN 3-921694-50-7, *vergriffen*

51 Hafner, Edzard: *Rohrleitungen und Behälter im Meer*, 1983, ISBN 3-921694-51-5

52 Rinnert, Bernd: *Hydrodynamische Dispersion in porösen Medien: Einfluß von Dichteunterschieden auf die Vertikalvermischung in horizontaler Strömung*, 1983, ISBN 3-921694-52-3, *vergriffen*

53 Lindner, Wulf: *Steuerung von Grundwasserentnahmen unter Einhaltung ökologischer Kriterien*, 1983, ISBN 3-921694-53-1, *vergriffen*

54 Herr, Michael; Herzer, Jörg; Kinzelbach, Wolfgang; Kobus, Helmut; Rinnert, Bernd: *Methoden zur rechnerischen Erfassung und hydraulischen Sanierung von Grundwasserkontaminationen*, 1983, ISBN 3-921694-54-X

55 Schmitt, Paul: *Wege zur Automatisierung der Niederschlagsermittlung*, 1984, ISBN 3-921694-55-8, *vergriffen*

56 Müller, Peter: *Transport und selektive Sedimentation von Schwebstoffen bei gestautem Abfluß*, 1985, ISBN 3-921694-56-6

57 El-Qawasmeh, Fuad: *Möglichkeiten und Grenzen der Tropfbewässerung unter besonderer Berücksichtigung der Verstopfungsanfälligkeit der Tropfelemente*, 1985, ISBN 3-921694-57-4, *vergriffen*

58 Kirchenbaur, Klaus: *Mikroprozessorgesteuerte Erfassung instationärer Druckfelder am Beispiel seegangsbelasteter Baukörper*, 1985, ISBN 3-921694-58-2

59 Kobus, Helmut (Hrsg.): *Modellierung des großräumigen Wärme- und Schadstofftransports im Grundwasser*, Tätigkeitsbericht 1984/85 (DFG-Forschergruppe an den Universitäten Hohenheim, Karlsruhe und Stuttgart), 1985, ISBN 3-921694-59-0, *vergriffen*

60 Spitz, Karlheinz: *Dispersion in porösen Medien: Einfluß von Inhomogenitäten und Dichteunterschieden*, 1985, ISBN 3-921694-60-4, *vergriffen*

61 Kobus, Helmut: *An Introduction to Air-Water Flows in Hydraulics*, 1985, ISBN 3-921694-61-2

62 Kaleris, Vassilios: *Erfassung des Austausches von Oberflächen- und Grundwasser in horizontalebenen Grundwassermodellen*, 1986, ISBN 3-921694-62-0

63 Herr, Michael: *Grundlagen der hydraulischen Sanierung verunreinigter Porengrundwasserleiter*, 1987, ISBN 3-921694-63-9

64 Marx, Walter: *Berechnung von Temperatur und Spannung in Massenbeton infolge Hydratation*, 1987, ISBN 3-921694-64-7

65 Koschitzky, Hans-Peter: *Dimensionierungskonzept für Sohlbelüfter in Schußrinnen zur Vermeidung von Kavitationsschäden*, 1987, ISBN 3-921694-65-5

66 Kobus, Helmut (Hrsg.): *Modellierung des großräumigen Wärme- und Schadstofftransports im Grundwasser*, Tätigkeitsbericht 1986/87 (DFG-Forschergruppe an den Universitäten Hohenheim, Karlsruhe und Stuttgart) 1987, ISBN 3-921694-66-3

67 Söll, Thomas: *Berechnungsverfahren zur Abschätzung anthropogener Temperaturanomalien im Grundwasser*, 1988, ISBN 3-921694-67-1

68 Dittrich, Andreas; Westrich, Bernd: *Bodenseeufererosion, Bestandsaufnahme und Bewertung*, 1988, ISBN 3-921694-68-X, *vergriffen*

69 Huwe, Bernd; van der Ploeg, Rienk R.: *Modelle zur Simulation des Stickstoffhaushaltes von Standorten mit unterschiedlicher landwirtschaftlicher Nutzung*, 1988, ISBN 3-921694-69-8, *vergriffen*

70 Stephan, Karl: *Integration elliptischer Funktionen*, 1988, ISBN 3-921694-70-1

71 Kobus, Helmut; Zilliox, Lothaire (Hrsg.): *Nitratbelastung des Grundwassers, Auswirkungen der Landwirtschaft auf die Grundwasser- und Rohwasserbeschaffenheit und Maßnahmen zum Schutz des Grundwassers*. Vorträge des deutsch-französischen Kolloquiums am 6. Oktober 1988, Universitäten Stuttgart und Louis Pasteur Strasbourg (Vorträge in deutsch oder französisch, Kurzfassungen zweisprachig), 1988, ISBN 3-921694-71-X

72 Soyeaux, Renald: *Unterströmung von Stauanlagen auf klüftigem Untergrund unter Berücksichtigung laminarer und turbulenter Fließzustände*,1991, ISBN 3-921694-72-8

73 Kohane, Roberto: *Berechnungsmethoden für Hochwasserabfluß in Fließgewässern mit überströmten Vorländern*, 1991, ISBN 3-921694-73-6

74 Hassinger, Reinhard: *Beitrag zur Hydraulik und Bemessung von Blocksteinrampen in flexibler Bauweise*, 1991, ISBN 3-921694-74-4, vergriffen

75 Schäfer, Gerhard: *Einfluß von Schichtenstrukturen und lokalen Einlagerungen auf die Längsdispersion in Porengrundwasserleitern*, 1991, ISBN 3-921694-75-2

76 Giesecke, Jürgen: *Vorträge, Wasserwirtschaft in stark besiedelten Regionen; Umweltforschung mit Schwerpunkt Wasserwirtschaft*, 1991, ISBN 3-921694-76-0

77 Huwe, Bernd: *Deterministische und stochastische Ansätze zur Modellierung des Stickstoffhaushalts landwirtschaftlich genutzter Flächen auf unterschiedlichem Skalenniveau*, 1992, ISBN 3-921694-77-9, vergriffen

78 Rommel, Michael: *Verwendung von Kluftdaten zur realitätsnahen Generierung von Kluftnetzen mit anschließender laminar-turbulenter Strömungsberechnung*, 1993, ISBN 3-92 1694-78-7

79 Marschall, Paul: *Die Ermittlung lokaler Stofffrachten im Grundwasser mit Hilfe von Einbohrloch-Meßverfahren*, 1993, ISBN 3-921694-79-5, vergriffen

80 Ptak, Thomas: *Stofftransport in heterogenen Porenaquiferen: Felduntersuchungen und stochastische Modellierung*, 1993, ISBN 3-921694-80-9, vergriffen

81 Haakh, Frieder: *Transientes Strömungsverhalten in Wirbelkammern,* 1993, ISBN 3-921694-81-7

82 Kobus, Helmut; Cirpka, Olaf; Barczewski, Baldur; Koschitzky, Hans-Peter: *Versuchseinrichtung zur Grundwasser- und Altlastensanierung VEGAS, Konzeption und Programmrahmen*, 1993, ISBN 3-921694-82-5

83 Zang, Weidong: *Optimaler Echtzeit-Betrieb eines Speichers mit aktueller Abflußregenerierung*, 1994, ISBN 3-921694-83-3, vergriffen

84 Franke, Hans-Jörg: *Stochastische Modellierung eines flächenhaften Stoffeintrages und Transports in Grundwasser am Beispiel der Pflanzenschutzmittelproblematik*, 1995, ISBN 3-921694-84-1

85 Lang, Ulrich: *Simulation regionaler Strömungs- und Transportvorgänge in Karstaquiferen mit Hilfe des Doppelkontinuum-Ansatzes: Methodenentwicklung und Parameteridentifikation*, 1995, ISBN 3-921694-85-X, vergriffen

86 Helmig, Rainer: *Einführung in die Numerischen Methoden der Hydromechanik*, 1996, ISBN 3-921694-86-8, vergriffen

87 Cirpka, Olaf: *CONTRACT: A Numerical Tool for Contaminant Transport and Chemical Transformations - Theory and Program Documentation -*, 1996, ISBN 3-921694-87-6

88 Haberlandt, Uwe: *Stochastische Synthese und Regionalisierung des Niederschlages für Schmutzfrachtberechnungen*, 1996, ISBN 3-921694-88-4

89 Croisé, Jean: *Extraktion von flüchtigen Chemikalien aus natürlichen Lockergesteinen mittels erzwungener Luftströmung*, 1996, ISBN 3-921694-89-2, vergriffen

90 Jorde, Klaus: *Ökologisch begründete, dynamische Mindestwasserregelungen bei Ausleitungskraftwerken*, 1997, ISBN 3-921694-90-6, vergriffen

91 Helmig, Rainer: *Gekoppelte Strömungs- und Transportprozesse im Untergrund - Ein Beitrag zur Hydrosystemmodellierung-*, 1998, ISBN 3-921694-91-4, vergriffen

92 Emmert, Martin: *Numerische Modellierung nichtisothermer Gas-Wasser Systeme in porösen Medien,* 1997, ISBN 3-921694-92-2

93 Kern, Ulrich: *Transport von Schweb- und Schadstoffen in staugeregelten Fließgewässern am Beispiel des Neckars*, 1997, ISBN 3-921694-93-0, vergriffen

94 Förster, Georg: *Druckstoßdämpfung durch große Luftblasen in Hochpunkten von Rohrleitungen 1997*, ISBN 3-921694-94-9

95 Cirpka, Olaf: *Numerische Methoden zur Simulation des reaktiven Mehrkomponenten-transports im Grundwasser*, 1997, ISBN 3-921694-95-7, *vergriffen*

96 Färber, Arne: *Wärmetransport in der ungesättigten Bodenzone: Entwicklung einer thermischen In-situ-Sanierungstechnologie*, 1997, ISBN 3-921694-96-5

97 Betz, Christoph: *Wasserdampfdestillation von Schadstoffen im porösen Medium: Entwicklung einer thermischen In-situ-Sanierungstechnologie*, 1998, SBN 3-921694-97-3

98 Xu, Yichun: *Numerical Modeling of Suspended Sediment Transport in Rivers*, 1998, ISBN 3-921694-98-1, *vergriffen*

99 Wüst, Wolfgang: *Geochemische Untersuchungen zur Sanierung CKW-kontaminierter Aquifere mit Fe(0)-Reaktionswänden,* 2000, ISBN 3-933761-02-2

100 Sheta, Hussam: *Simulation von Mehrphasenvorgängen in porösen Medien unter Einbeziehung von Hysterese-Effekten*, 2000, ISBN 3-933761-03-4

101 Ayros, Edwin: *Regionalisierung extremer Abflüsse auf der Grundlage statistischer Verfahren*, 2000, ISBN 3-933761-04-2, *vergriffen*

102 Huber, Ralf: *Compositional Multiphase Flow and Transport in Heterogeneous Porous Media*, 2000, ISBN 3-933761-05-0

103 Braun, Christopherus: *Ein Upscaling-Verfahren für Mehrphasenströmungen in porösen Medien*, 2000, ISBN 3-933761-06-9

104 Hofmann, Bernd: *Entwicklung eines rechnergestützten Managementsystems zur Beurteilung von Grundwasserschadensfällen*, 2000, ISBN 3-933761-07-7

105 Class, Holger: *Theorie und numerische Modellierung nichtisothermer Mehrphasenprozesse in NAPL-kontaminierten porösen Medien,* 2001, ISBN 3-933761-08-5

106 Schmidt, Reinhard: *Wasserdampf- und Heißluftinjektion zur thermischen Sanierung kontaminierter Standorte*, 2001, ISBN 3-933761-09-3

107 Josef, Reinhold: *Schadstoffextraktion mit hydraulischen Sanierungsverfahren unter Anwendung von grenzflächenaktiven Stoffen*, 2001, ISBN 3-933761-10-7

108 Schneider, Matthias: *Habitat- und Abflussmodellierung für Fließgewässer mit unscharfen Berechnungsansätzen*, 2001, ISBN 3-933761-11-5

109 Rathgeb, Andreas: *Hydrodynamische Bemessungsgrundlagen für Lockerdeckwerke an überströmbaren Erddämmen*, 2001, ISBN 3-933761-12-3

110 Lang, Stefan: *Parallele numerische Simulation instätionärer Probleme mit adaptiven Methoden auf unstrukturierten Gittern*, 2001, ISBN 3-933761-13-1

111 Appt, Jochen; Stumpp Simone: *Die Bodensee-Messkampagne 2001, IWS/CWR Lake Constance Measurement Program 2001*, 2002, ISBN 3-933761-14-X

112 Heimerl, Stephan: *Systematische Beurteilung von Wasserkraftprojekten,* 2002, ISBN 3-933761-15-8, *vergriffen*

113 Iqbal, Amin: *On the Management and Salinity Control of Drip Irrigation,* 2002, ISBN 3-933761-16-6

114 Silberhorn-Hemminger, Annette: *Modellierung von Kluftaquifersystemen: Geostatistische Analyse und deterministisch-stochastische Kluftgenerierung,* 2002, ISBN 3-933761-17-4

115 Winkler, Angela: *Prozesse des Wärme- und Stofftransports bei der In-situ-Sanierung mit festen Wärmequellen,* 2003, ISBN 3-933761-18-2

116 Marx, Walter: *Wasserkraft, Bewässerung, Umwelt - Planungs- und Bewertungsschwerpunkte der Wasserbewirtschaftung,* 2003, ISBN 3-933761-19-0

117 Hinkelmann, Reinhard: *Efficient Numerical Methods and Information-Processing Techniques in Environment Water,* 2003, ISBN 3-933761-20-4

118 Samaniego-Eguiguren, Luis Eduardo: *Hydrological Consequences of Land Use / Land Cover and Climatic Changes in Mesoscale Catchments*, 2003, ISBN 3-933761-21-2

119 Neunhäuserer, Lina: *Diskretisierungsansätze zur Modellierung von Strömungs- und Transportprozessen in geklüftet-porösen Medien*, 2003, ISBN 3-933761-22-0

120 Paul, Maren: *Simulation of Two-Phase Flow in Heterogeneous Poros Media with Adaptive Methods*, 2003, ISBN 3-933761-23-9

121 Ehret, Uwe: *Rainfall and Flood Nowcasting in Small Catchments using Weather Radar*, 2003, ISBN 3-933761-24-7

122 Haag, Ingo: *Der Sauerstoffhaushalt staugeregelter Flüsse am Beispiel des Neckars - Analysen, Experimente, Simulationen -,* 2003, ISBN 3-933761-25-5

123 Appt, Jochen: *Analysis of Basin-Scale Internal Waves in Upper Lake Constance,* 2003, ISBN 3-933761-26-3

124 Hrsg.: Schrenk, Volker; Batereau, Katrin; Barczewski, Baldur; Weber, Karolin und Koschitzky, Hans-Peter: *Symposium Ressource Fläche und VEGAS - Statuskolloquium 2003, 30. September und 1. Oktober 2003,* 2003, ISBN 3-933761-27-1

125 Omar Khalil Ouda: *Optimisation of Agricultural Water Use: A Decision Support System for the Gaza Strip,* 2003, ISBN 3-933761-28-0

126 Batereau, Katrin: *Sensorbasierte Bodenluftmessung zur Vor-Ort-Erkundung von Schadensherden im Untergrund*, 2004, ISBN 3-933761-29-8

127 Witt, Oliver: *Erosionsstabilität von Gewässersedimenten mit Auswirkung auf den Stofftransport bei Hochwasser am Beispiel ausgewählter Stauhaltungen des Oberrheins*, 2004, ISBN 3-933761-30-1

128 Jakobs, Hartmut: *Simulation nicht-isothermer Gas-Wasser-Prozesse in komplexen Kluft-Matrix-Systemen*, 2004, ISBN 3-933761-31-X

129 Li, Chen-Chien: *Deterministisch-stochastisches Berechnungskonzept zur Beurteilung der Auswirkungen erosiver Hochwasserereignisse in Flussstauhaltungen,* 2004, ISBN 3-933761-32-8

130 Reichenberger, Volker; Helmig, Rainer; Jakobs, Hartmut; Bastian, Peter; Niessner, Jennifer: *Complex Gas-Water Processes in Discrete Fracture-Matrix Systems: Up-scaling, Mass-Conservative Discretization and Efficient Multilevel Solution,* 2004, ISBN 3-933761-33-6

131 Hrsg.: Barczewski, Baldur; Koschitzky, Hans-Peter; Weber, Karolin; Wege, Ralf: *VEGAS - Statuskolloquium 2004,* Tagungsband zur Veranstaltung am 05. Oktober 2004 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2004, ISBN 3-933761-34-4

132 Asie, Kemal Jabir: *Finite Volume Models for Multiphase Multicomponent Flow through Porous Media*. 2005, ISBN 3-933761-35-2

133 Jacoub, George: *Development of a 2-D Numerical Module for Particulate Contaminant Transport in Flood Retention Reservoirs and Impounded Rivers,* 2004, ISBN 3-933761-36-0

134 Nowak, Wolfgang: *Geostatistical Methods for the Identification of Flow and Transport Parameters in the Subsurface,* 2005, ISBN 3-933761-37-9

135 Süß, Mia: *Analysis of the influence of structures and boundaries on flow and transport processes in fractured porous media*, 2005, ISBN 3-933761-38-7

136 Jose, Surabhin Chackiath: *Experimental Investigations on Longitudinal Dispersive Mixing in Heterogeneous Aquifers,* 2005, ISBN: 3-933761-39-5

137 Filiz, Fulya: *Linking Large-Scale Meteorological Conditions to Floods in Mesoscale Catchments,* 2005, ISBN 3-933761-40-9

138 Qin, Minghao: *Wirklichkeitsnahe und recheneffiziente Ermittlung von Temperatur und Spannungen bei großen RCC-Staumauern*, 2005, ISBN 3-933761-41-7

139 Kobayashi, Kenichiro: *Optimization Methods for Multiphase Systems in the Subsurface - Application to Methane Migration in Coal Mining Areas*, 2005, ISBN 3-933761-42-5

140 Rahman, Md. Arifur: *Experimental Investigations on Transverse Dispersive Mixing in Heterogeneous Porous Media,* 2005, ISBN 3-933761-43-3

141 Schrenk, Volker: *Ökobilanzen zur Bewertung von Altlastensanierungsmaßnahmen,* 2005, ISBN 3-933761-44-1

142 Hundecha, Hirpa Yeshewatesfa: *Regionalization of Parameters of a Conceptual Rainfall-Runoff Model*, 2005, ISBN: 3-933761-45-X

143 Wege, Ralf: *Untersuchungs- und Überwachungsmethoden für die Beurteilung natürlicher Selbstreinigungsprozesse im Grundwasser,* 2005, ISBN 3-933761-46-8

144 Breiting, Thomas: *Techniken und Methoden der Hydroinformatik - Modellierung von komplexen Hydrosystemen im Untergrund*, 2006, ISBN 3-933761-47-6

145 Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Müller, Martin: *Ressource Untergrund: 10 Jahre VEGAS: Forschung und Technologieentwicklung zum Schutz von Grundwasser und Boden,* Tagungsband zur Veranstaltung am 28. und 29. September 2005 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2005, ISBN 3-933761-48-4

146 Rojanschi, Vlad: *Abflusskonzentration in mesoskaligen Einzugsgebieten unter Berücksichtigung des Sickerraumes,* 2006, ISBN 3-933761-49-2

147 Winkler, Nina Simone: *Optimierung der Steuerung von Hochwasserrückhaltebeckensystemen,* 2006, ISBN 3-933761-50-6

148 Wolf, Jens: *Räumlich differenzierte Modellierung der Grundwasserströmung alluvialer Aquifere für mesoskalige Einzugsgebiete*, 2006, ISBN: 3-933761-51-4

149 Kohler, Beate: *Externe Effekte der Laufwasserkraftnutzung*, 2006, ISBN 3-933761-52-2

150 Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Stuhrmann, Matthias: *VEGAS-Statuskolloquium 2006,* Tagungsband zur Veranstaltung am 28. September 2006 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2006, ISBN 3-933761-53-0

151 Niessner, Jennifer: *Multi-Scale Modeling of Multi-Phase - Multi-Component Processes in Heterogeneous Porous Media*, 2006, ISBN 3-933761-54-9

152 Fischer, Markus: *Beanspruchung eingeerdeter Rohrleitungen infolge Austrocknung bindiger Böden,* 2006, ISBN 3-933761-55-7

153 Schneck, Alexander: *Optimierung der Grundwasserbewirtschaftung unter Berücksichtigung der Belange der Wasserversorgung, der Landwirtschaft und des Naturschutzes*, 2006, ISBN 3-933761-56-5

154 Das, Tapash: *The Impact of Spatial Variability of Precipitation on the Predictive Uncertainty of Hydrological Models,* 2006, ISBN 3-33761-57-3

155 Bielinski, Andreas: *Numerical Simulation of CO_2 sequestration in geological formations*, 2007, ISBN 3-933761-58-1

156 Mödinger, Jens: *Entwicklung eines Bewertungs- und Entscheidungsunterstützungssystems für eine nachhaltige regionale Grundwasserbewirtschaftung*, 2006, ISBN 3-933761-60-3

157 Manthey, Sabine: *Two-phase flow processes with dynamic effects in porous media - parameter estimation and simulation,* 2007, ISBN 3-933761-61-1

158 Pozos Estrada, Oscar: *Investigation on the Effects of Entrained Air in Pipelines,* 2007, ISBN 3-933761-62-X

159 Ochs, Steffen Oliver: *Steam injection into saturated porous media – process analysis including experimental and numerical investigations*, 2007, ISBN 3-933761-63-8

160 Marx, Andreas: *Einsatz gekoppelter Modelle und Wetterradar zur Abschätzung von Niederschlagsintensitäten und zur Abflussvorhersage, 2007,* ISBN 3-933761-64-6

161 Hartmann, Gabriele Maria: *Investigation of Evapotranspiration Concepts in Hydrological Modelling for Climate Change Impact Assessment*, 2007, ISBN 3-933761-65-4

162 Kebede Gurmessa, Tesfaye: *Numerical Investigation on Flow and Transport Characteristics to Improve Long-Term Simulation of Reservoir Sedimentation*, 2007, ISBN 3-933761-66-2

163 Trifković, Aleksandar: *Multi-objective and Risk-based Modelling Methodology for Planning, Design and Operation of Water Supply Systems*, 2007, ISBN 3-933761-67-0

164 Götzinger, Jens: *Distributed Conceptual Hydrological Modelling - Simulation of Climate, Land Use Change Impact and Uncertainty Analysis*, 2007, ISBN 3-933761-68-9

165 Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Stuhrmann, Matthias: *VEGAS – Kolloquium 2007,* Tagungsband zur Veranstaltung am 26. September 2007 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2007, ISBN 3-933761-69-7

166 Freeman, Beau: *Modernization Criteria Assessment for Water Resources Planning; Klamath Irrigation Project, U.S.*, 2008, ISBN 3-933761-70-0

167 Dreher, Thomas: *Selektive Sedimentation von Feinstschwebstoffen in Wechselwirkung mit wandnahen turbulenten Strömungsbedingungen*, 2008, ISBN 3-933761-71-9

168 Yang, Wei: *Discrete-Continuous Downscaling Model for Generating Daily Precipitation Time Series*, 2008, ISBN 3-933761-72-7

169 Kopecki, Ianina: *Calculational Approach to FST-Hemispheres for Multiparametrical Benthos Habitat Modelling,* 2008*,* ISBN 3-933761-73-5

170 Brommundt, Jürgen: *Stochastische Generierung räumlich zusammenhängender Niederschlagszeitreihen*, 2008, ISBN 3-933761-74-3

171 Papafotiou, *Alexandros: Numerical Investigations of the Role of Hysteresis in Heterogeneous Two-Phase Flow Systems*, 2008, ISBN 3-933761-75-1

172 He, Yi: *Application of a Non-Parametric Classification Scheme to Catchment Hydrology*, 2008, ISBN 978-3-933761-76-7

173 Wagner, Sven: *Water Balance in a Poorly Gauged Basin in West Africa Using Atmospheric Modelling and Remote Sensing Information*, 2008, ISBN 978-3-933761-77-4

174 Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Stuhrmann, Matthias; Schrenk, Volker: *VEGAS-Kolloquium 2008 Ressource Fläche III*, Tagungsband zur Veranstaltung am 01. Oktober 2008 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2008, ISBN 978-3-933761-78-1

175 Patil, Sachin: *Regionalization of an Event Based Nash Cascade Model for Flood Predictions in Ungauged Basins*, 2008, ISBN 978-3-933761-79-8

176 Assteerawatt, Anongnart: *Flow and Transport Modelling of Fractured Aquifers based on a Geostatistical Approach,* 2008, ISBN 978-3-933761-80-4

177 Karnahl, Joachim Alexander: *2D numerische Modellierung von multifraktionalem Schwebstoff- und Schadstofftransport in Flüssen*, 2008, ISBN 978-3-933761-81-1

178 Hiester, Uwe: *Technologieentwicklung zur In-situ-Sanierung der ungesättigten Bodenzone mit festen Wärmequellen,* 2009, ISBN 978-3-933761-82-8

179 Laux, Patrick: *Statistical Modeling of Precipitation for Agricultural Planning in the Volta Basin of West Africa,* 2009, ISBN 978-3-933761-83-5

180 Ehsan, Saqib: *Evaluation of Life Safety Risks Related to Severe Flooding,* 2009, ISBN 978-3-933761-84-2

181 Prohaska, Sandra: *Development and Application of a 1D Multi-Strip Fine Sediment Transport Model for Regulated Rivers,* 2009, ISBN 978-3-933761-85-9

182 Kopp, Andreas: *Evaluation of CO_2 Injection Processes in Geological Formations for Site Screening,* 2009, ISBN 978-3-933761-86-6

183 Ebigbo, Anozie: *Modelling of biofilm growth and its influence on CO_2 and water (two-phase) flow in porous media,* 2009, ISBN 978-3-933761-87-3

184 Freiboth, Sandra: *A phenomenological model for the numerical simulation of multiphase multicomponent processes considering structural alterations of porous media*, 2009, ISBN 978-3-933761-88-0

185 Zöllner, Frank: *Implementierung und Anwendung netzfreier Methoden im Konstruktiven Wasserbau und in der Hydromechanik*, 2009, ISBN 978-3-933761-89-7

186 Vasin, Milos: *Influence of the soil structure and property contrast on flow and transport in the unsaturated zone*, 2010, ISBN 978-3-933761-90-3

187 Li, Jing: *Application of Copulas as a New Geostatistical Tool*, 2010, ISBN 978-3-933761-91-0

188 AghaKouchak, Amir: *Simulation of Remotely Sensed Rainfall Fields Using Copulas*, 2010, ISBN 978-3-933761-92-7

189 Thapa, Pawan Kumar: *Physically-based spatially distributed rainfall runoff modelling for soil erosion estimation*, 2010, ISBN 978-3-933761-93-4

190 Wurms, Sven: *Numerische Modellierung der Sedimentationsprozesse in Retentionsanlagen zur Steuerung von Stoffströmen bei extremen Hochwasserabflussereignissen*, 2011, ISBN 978-3-933761-94-1

191 Merkel, Uwe: *Unsicherheitsanalyse hydraulischer Einwirkungen auf Hochwasserschutzdeiche und Steigerung der Leistungsfähigkeit durch adaptive Strömungsmodellierung*, 2011, ISBN 978-3-933761-95-8

192 Fritz, Jochen: *A Decoupled Model for Compositional Non-Isothermal Multiphase Flow in Porous Media and Multiphysics Approaches for Two-Phase Flow*, 2010, ISBN 978-3-933761-96-5

193 Weber, Karolin (Hrsg.): *12. Treffen junger WissenschaftlerInnen an Wasserbauinstituten*, 2010, ISBN 978-3-933761-97-2

194 Bliefernicht, Jan-Geert: *Probability Forecasts of Daily Areal Precipitation for Small River Basins,* 2011, ISBN 978-3-933761-98-9

195 Hrsg.: Koschitzky, Hans-Peter; Braun, Jürgen: *VEGAS-Kolloquium 2010 In-situ-Sanierung - Stand und Entwicklung Nano und ISCO -,* Tagungsband zur Veranstaltung am 07. Oktober 2010 an der Universität Stuttgart, Campus Stuttgart-Vaihingen*,* 2010*,* ISBN 978-3-933761-99-6

196 Gafurov, Abror: *Water Balance Modeling Using Remote Sensing Information - Focus on Central Asia,* 2010, ISBN 978-3-942036-00-9

197 Mackenberg, Sylvia: *Die Quellstärke in der Sickerwasserprognose: Möglichkeiten und Grenzen von Labor- und Freilanduntersuchungen,* 2010, ISBN 978-3-942036-01-6

198 Singh, Shailesh Kumar: *Robust Parameter Estimation in Gauged and Ungauged Basins,* 2010, ISBN 978-3-942036-02-3

199 Doğan, Mehmet Onur: *Coupling of porous media flow with pipe flow,* 2011, ISBN 978-3-942036-03-0

200 Liu, Min: *Study of Topographic Effects on Hydrological Patterns and the Implication on Hydrological Modeling and Data Interpolation,* 2011, ISBN 978-3-942036-04-7

201 Geleta, Habtamu Itefa: *Watershed Sediment Yield Modeling for Data Scarce Areas,* 2011, ISBN 978-3-942036-05-4

202 Franke, Jörg: *Einfluss der Überwachung auf die Versagenswahrscheinlichkeit von Staustufen,* 2011, ISBN 978-3-942036-06-1

203 Bakimchandra, Oinam: *Integrated Fuzzy-GIS approach for assessing regional soil erosion risks,* 2011, ISBN 978-3-942036-07-8

204 Alam, Muhammad Mahboob: *Statistical Downscaling of Extremes of Precipitation in Mesoscale Catchments from Different RCMs and Their Effects on Local Hydrology,* 2011, ISBN 978-3-942036-08-5

205 Hrsg.: Koschitzky, Hans-Peter; Braun, Jürgen: *VEGAS-Kolloquium 2011 Flache Geothermie - Perspektiven und Risiken,* Tagungsband zur Veranstaltung am 06. Oktober 2011 an der Universität Stuttgart, Campus Stuttgart-Vaihingen*,* 2011*,* ISBN 978-3-933761-09-2

206 Haslauer, Claus: *Analysis of Real-World Spatial Dependence of Subsurface Hydraulic Properties Using Copulas with a Focus on Solute Transport Behaviour,* 2011, ISBN 978-3-942036-10-8

207 Dung, Nguyen Viet: *Multi-objective automatic calibration of hydrodynamic models – development of the concept and an application in the Mekong Delta,* 2011, ISBN 978-3-942036-11-5

208 Hung, Nguyen Nghia: *Sediment dynamics in the floodplain of the Mekong Delta, Vietnam,* 2011, ISBN 978-3-942036-12-2

209 Kuhlmann, Anna: *Influence of soil structure and root water uptake on flow in the unsaturated zone,* 2012, ISBN 978-3-942036-13-9

210 Tuhtan, Jeffrey Andrew: *Including the Second Law Inequality in Aquatic Ecodynamics: A Modeling Approach for Alpine Rivers Impacted by Hydropeaking,* 2012, ISBN 978-3-942036-14-6

211 Tolossa, Habtamu: *Sediment Transport Computation Using a Data-Driven Adaptive Neuro-Fuzzy Modelling Approach,* 2012, ISBN 978-3-942036-15-3

212 Tatomir, Alexandru-Bodgan: *From Discrete to Continuum Concepts of Flow in Fractured*

Porous Media, 2012, ISBN 978-3-942036-16-0

213 Erbertseder, Karin: *A Multi-Scale Model for Describing Cancer-Therapeutic Transport in the Human Lung,* 2012, ISBN 978-3-942036-17-7

214 Noack, Markus: *Modelling Approach for Interstitial Sediment Dynamics and Reproduction of Gravel Spawning Fish,* 2012, ISBN 978-3-942036-18-4

215 De Boer, Cjestmir Volkert: *Transport of Nano Sized Zero Valent Iron Colloids during Injection into the Subsurface*, 2012, ISBN 978-3-942036-19-1

216 Pfaff, Thomas: *Processing and Analysis of Weather Radar Data for Use in Hydrology*, 2013, ISBN 978-3-942036-20-7

217 Lebrenz, Hans-Henning: *Addressing the Input Uncertainty for Hydrological Modeling by a New Geostatistical Method*, 2013, ISBN 978-3-942036-21-4

218 Darcis, Melanie Yvonne: *Coupling Models of Different Complexity for the Simulation of CO_2 Storage in Deep Saline Aquifers*, 2013, ISBN 978-3-942036-22-1

219 Beck, Ferdinand: *Generation of Spatially Correlated Synthetic Rainfall Time Series in High Temporal Resolution - A Data Driven Approach*, 2013, ISBN 978-3-942036-23-8

220 Guthke, Philipp: *Non-multi-Gaussian spatial structures: Process-driven natural genesis, manifestation, modeling approaches, and influences on dependent processes*, 2013, ISBN 978-3-942036-24-5

221 Walter, Lena: *Uncertainty studies and risk assessment for CO_2 storage in geological formations*, 2013, ISBN 978-3-942036-25-2

222 Wolff, Markus: *Multi-scale modeling of two-phase flow in porous media including capillary pressure effects,* 2013, ISBN 978-3-942036-26-9

223 Mosthaf, Klaus Roland: *Modeling and analysis of coupled porous-medium and free flow with application to evaporation processes,* 2014, ISBN 978-3-942036-27-6

224 Leube, Philipp Christoph: *Methods for Physically-Based Model Reduction in Time: Analysis, Comparison of Methods and Application*, 2013, ISBN 978-3-942036-28-3

225 Rodríguez Fernández, Jhan Ignacio: *High Order Interactions among environmental variables: Diagnostics and initial steps towards modeling*, 2013, ISBN 978-3-942036-29-0

226 Eder, Maria Magdalena: *Climate Sensitivity of a Large Lake*, 2013, ISBN 978-3-942036-30-6

227 Greiner, Philipp: *Alkoholinjektion zur In-situ-Sanierung von CKW Schadensherden in Grundwasserleitern: Charakterisierung der relevanten Prozesse auf unterschiedlichen Skalen*, 2014, ISBN 978-3-942036-31-3

228 Lauser, Andreas: *Theory and Numerical Applications of Compositional Multi-Phase Flow in Porous Media*, 2014, ISBN 978-3-942036-32-0

229 Enzenhöfer, Rainer: *Risk Quantification and Management in Water Production and Supply Systems*, 2014, ISBN 978-3-942036-33-7

230 Faigle, Benjamin: *Adaptive modelling of compositional multi-phase flow with capillary pressure*, 2014, ISBN 978-3-942036-34-4

231 Oladyshkin, Sergey: *Efficient modeling of environmental systems in the face of complexity and uncertainty*, 2014, ISBN 978-3-942036-35-1

232 Sugimoto, Takayuki: *Copula based Stochastic Analysis of Discharge Time Series*, 2014, ISBN 978-3-942036-36-8

233 Koch, Jonas: *Simulation, Identification and Characterization of Contaminant Source Architectures in the Subsurface*, 2014, ISBN 978-3-942036-37-5

234 Zhang, Jin: *Investigations on Urban River Regulation and Ecological Rehabilitation Measures, Case of Shenzhen in China*, 2014, ISBN 978-3-942036-38-2

235 Siebel, Rüdiger: *Experimentelle Untersuchungen zur hydrodynamischen Belastung und Standsicherheit von Deckwerken an überströmbaren Erddämmen,* 2014, ISBN 978-3-942036-39-9

236 Baber, Katherina: *Coupling free flow and flow in porous media in biological and technical applications: From a simple to a complex interface description,* 2014,

ISBN 978-3-942036-40-5

237 Nuske, Klaus Philipp: *Beyond Local Equilibrium — Relaxing local equilibrium assumptions in multiphase flow in porous media,* 2014, ISBN 978-3-942036-41-2

238 Geiges, Andreas: *Efficient concepts for optimal experimental design in nonlinear environmental systems*, 2014, ISBN 978-3-942036-42-9

239 Schwenck, Nicolas: *An XFEM-Based Model for Fluid Flow in Fractured Porous Media*, 2014, ISBN 978-3-942036-43-6

240 Chamorro Chávez, Alejandro: *Stochastic and hydrological modelling for climate change prediction in the Lima region, Peru*, 2015, ISBN 978-3-942036-44-3

241 Yulizar: *Investigation of Changes in Hydro-Meteorological Time Series Using a Depth-Based Approach*, 2015, ISBN 978-3-942036-45-0

242 Kretschmer, Nicole: *Impacts of the existing water allocation scheme on the Limarí watershed – Chile, an integrative approach*, 2015, ISBN 978-3-942036-46-7

243 Kramer, Matthias: *Luftbedarf von Freistrahlturbinen im Gegendruckbetrieb*, 2015, ISBN 978-3-942036-47-4

244 Hommel, Johannes: *Modeling biogeochemical and mass transport processes in the subsurface: Investigation of microbially induced calcite precipitation*, 2016, ISBN 978-3-942036-48-1

245 Germer, Kai: *Wasserinfiltration in die ungesättigte Zone eines makroporösen Hanges und deren Einfluss auf die Hangstabilität*, 2016, ISBN 978-3-942036-49-8

246 Hörning, Sebastian: *Process-oriented modeling of spatial random fields using copulas*, 2016, ISBN 978-3-942036-50-4

247 Jambhekar, Vishal: *Numerical modeling and analysis of evaporative salinization in a coupled free-flow porous-media system,* 2016, ISBN 978-3-942036-51-1

248 Huang, Yingchun: *Study on the spatial and temporal transferability of conceptual hydrological models*, 2016, ISBN 978-3-942036-52-8

249 Kleinknecht, Simon Matthias: *Migration and retention of a heavy NAPL vapor and remediation of the unsaturated zone*, 2016, ISBN 978-3-942036-53-5

250 Kwakye, Stephen Oppong: *Study on the effects of climate change on the hydrology of the West African sub-region*, 2016, ISBN 978-3-942036-54-2

251 Kissinger, Alexander: *Basin-Scale Site Screening and Investigation of Possible Impacts of CO_2 Storage on Subsurface Hydrosystems*, 2016, ISBN 978-3-942036-55-9

252 Müller, Thomas: *Generation of a Realistic Temporal Structure of Synthetic Precipitation Time Series for Sewer Applications*, 2017, ISBN 978-3-942036-56-6

253 Grüninger, Christoph: *Numerical Coupling of Navier-Stokes and Darcy Flow for Soil-Water Evaporation*, 2017, ISBN 978-3-942036-57-3

254 Suroso: *Asymmetric Dependence Based Spatial Copula Models: Empirical Investigations and Consequences on Precipitation Fields*, 2017, ISBN 978-3-942036-58-0

255 Müller, Thomas; Mosthaf, Tobias; Gunzenhauser, Sarah; Seidel, Jochen; Bárdossy, András: *Grundlagenbericht Niederschlags-Simulator (NiedSim3)*, 2017, ISBN 978-3-942036-59-7

256 Mosthaf, Tobias: *New Concepts for Regionalizing Temporal Distributions of Precipitation and for its Application in Spatial Rainfall Simulation,* 2017, ISBN 978-3-942036-60-3

257 Fenrich, Eva Katrin: *Entwicklung eines ökologisch-ökonomischen Vernetzungsmodells für Wasserkraftanlagen und Mehrzweckspeicher,* 2018, ISBN 978-3-942036-61-0

258 Schmidt, Holger: *Microbial stabilization of lotic fine sediments*, 2018, ISBN 978-3-942036-62-7

259 Fetzer, Thomas: *Coupled Free and Porous-Medium Flow Processes Affected by Turbulence and Roughness – Models, Concepts and Analysis*, 2018, ISBN 978-3-942036-63-4

260 Schröder, Hans Christoph: *Large-scale High Head Pico Hydropower Potential Assessment*, 2018, ISBN 978-3-942036-64-1

261 Bode, Felix: *Early-Warning Monitoring Systems for Improved Drinking Water Resource*

Protection, 2018, ISBN 978-3-942036-65-8

262 Gebler, Tobias: *Statistische Auswertung von simulierten Talsperrenüberwachungsdaten zur Identifikation von Schadensprozessen an Gewichtsstaumauern*, 2018, ISBN 978-3-942036-66-5

263 Harten, Matthias von: *Analyse des Zuppinger-Wasserrades – Hydraulische Optimierungen unter Berücksichtigung ökologischer Aspekte*, 2018, ISBN 978-3-942036-67-2

264 Yan, Jieru: *Nonlinear estimation of short time precipitation using weather radar and surface observations*, 2018, ISBN 978-3-942036-68-9

265 Beck, Martin: *Conceptual approaches for the analysis of coupled hydraulic and geomechanical processes*, 2019, ISBN 978-3-942036-69-6

266 Haas, Jannik: *Optimal planning of hydropower and energy storage technologies for fully renewable power systems*, 2019, ISBN 978-3-942036-70-2

267 Schneider, Martin: *Nonlinear Finite Volume Schemes for Complex Flow Processes and Challenging Grids*, 2019, ISBN 978-3-942036-71-9

268 Most, Sebastian Christopher: *Analysis and Simulation of Anomalous Transport in Porous Media*, 2019, ISBN 978-3-942036-72-6

269 Buchta, Rocco: *Entwicklung eines Ziel- und Bewertungssystems zur Schaffung nachhaltiger naturnaher Strukturen in großen sandgeprägten Flüssen des norddeutschen Tieflandes*, 2019, ISBN 978-3-942036-73-3

270 Thom, Moritz: *Towards a Better Understanding of the Biostabilization Mechanisms of Sediment Beds*, 2019, ISBN 978-3-942036-74-0

271 Stolz, Daniel: *Die Nullspannungstemperatur in Gewichtsstaumauern unter Berücksichtigung der Festigkeitsentwicklung des Betons*, 2019, ISBN 978-3-942036-75-7

272 Rodriguez Pretelin, Abelardo: *Integrating transient flow conditions into groundwater well protection*, 2020, ISBN: 978-3-942036-76-4

273 Weishaupt, Kilian: *Model Concepts for Coupling Free Flow with Porous Medium Flow at the Pore-Network Scale: From Single-Phase Flow to Compositional Non-Isothermal Two-Phase Flow*, 2020, ISBN: 978-3-942036-77-1

274 Koch, Timo: *Mixed-dimension models for flow and transport processes in porous media with embedded tubular network systems*, 2020, ISBN: 978-3-942036-78-8

275 Gläser, Dennis: *Discrete fracture modeling of multi-phase flow and deformation in fractured poroelastic media*, 2020, ISBN: 978-3-942036-79-5

276 Seitz, Lydia: *Development of new methods to apply a multi-parameter approach – A first step towards the determination of colmation*, 2020, ISBN: 978-3-942036-80-1

277 Ebrahim Bakhshipour, Amin: *Optimizing hybrid decentralized systems for sustainable urban drainage infrastructures planning*, 2021, ISBN: 978-3-942036-81-8

278 Seitz, Gabriele: *Modeling Fixed-Bed Reactors for Thermochemical Heat Storage with the Reaction System $CaO/Ca(OH)_2$*, 2021, ISBN: 978-3-942036-82-5

279 Emmert, Simon: *Developing and Calibrating a Numerical Model for Microbially Enhanced Coal-Bed Methane Production*, 2021, ISBN: 978-3-942036-83-2

280 Heck, Katharina Klara: *Modelling and analysis of multicomponent transport at the interface between free- and porous-medium flow - influenced by radiation and roughness*, 2021, ISBN: 978-3-942036-84-9

281 Ackermann, Sina: *A multi-scale approach for drop/porous-medium interaction*, 2021, ISBN: 978-3-942036-85-6

282 Beckers, Felix: *Investigations on Functional Relationships between Cohesive Sediment Erosion and Sediment Characteristics*, 2021, ISBN: 978-3-942036-86-3

283 Schlabing, Dirk: *Generating Weather for Climate Impact Assessment on Lakes*, 2021, ISBN: 978-3-942036-87-0

284 Becker, Beatrix: *Efficient multiscale multiphysics models accounting for reversible flow at various subsurface energy storage sites*, 2021, ISBN: 978-3-942036-88-7

285 Reuschen, Sebastian: *Bayesian Inversion and Model Selection of Heterogeneities in Geo-*

statistical Subsurface Modeling, 2021, ISBN: 978-3-942036-89-4

286 Michalkowski, Cynthia: *Modeling water transport at the interface between porous GDL and gas distributor of a PEM fuel cell cathode*, 2022, ISBN: 978-3-942036-90-0

287 Koca, Kaan: *Advanced experimental methods for investigating flow-biofilm-sediment interactions*, 2022, ISBN: 978-3-942036-91-7

288 Modiri, Ehsan: *Clustering simultaneous occurrences of extreme floods in the Neckar catchment*, 2022, ISBN: 978-3-942036-92-4

Die Mitteilungshefte ab der Nr. 134 (Jg. 2005) stehen als pdf-Datei über die Homepage des Instituts: www.iws.uni-stuttgart.de zur Verfügung.