



Ingenieur fakultät Bau Geo Umwelt  
Lehrstuhl für Hydrologie  
und Flussgebietsmanagement

Department of Civil Environmental  
and Mechanical Engineering

Diese Promotion wurde von der Universität Trient und von der Technischen Universität München gemeinsam betreut.

# Homogenization and analysis of hydrological time series

Giorgia Marcolini

Vollständiger Abdruck der von der Ingenieur fakultät Bau Geo Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Ralf Ludwig  
Prüfer der Dissertation:

1. Prof. Dr.-Ing. Markus Disse
2. Prof. Dr. Alberto Bellin
3. Prof. Dr. Salvatore Grimaldi
4. Prof. Dr. Ralf Ludwig

Die Dissertation wurde am 11.07.2017 bei der Technischen Universität München eingereicht und durch die Ingenieur fakultät Bau Geo Umwelt am 11.10.2017 angenommen.





Department of Civil, Geo  
and Environmental Engineering

Department of Civil Environmental  
and Mechanical Engineering

Chair of Hydrology  
and River Basin Management

This doctoral thesis was cosupervised by the University of Trento and by the Technical University of Munich.

# Homogenization and analysis of hydrological time series

Giorgia Marcolini

Complete copy of the dissertation approved by the degree-awarding institution of the Department of Civil Environmental and Mechanical Engineering of the University of Trento for the obtainment of the academic degree of

Dottore di Ricerca in Ingegneria Ambientale

Chair: Prof. Dr. Ralf Ludwig

Dissertation examiners:

1. Prof. Dr.-Ing. Markus Disse
2. Prof. Dr. Alberto Bellin
3. Prof. Dr. Salvatore Grimaldi
4. Prof. Dr. Ralf Ludwig

The dissertation was submitted to the Technical University of Munich on the 11.07.2017 and accepted by the degree-awarding institution of Department of Civil, Geo and Environmental Engineering on the 11.10.2017.



*To my family*



# Contents

<b>Abstract</b>	<b>i</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Description of the datasets</b>	<b>5</b>
2.1 Trentino snow depth dataset . . . . .	5
2.2 Austrian snow depth dataset . . . . .	8
2.3 Trentino - A.A. snow depth and temperature dataset	10
2.3.1 Study area . . . . .	10
2.3.2 Description of the dataset . . . . .	10
2.4 Adige and upper Inn river basins discharge dataset	15
2.5 Lower Inn river discharge dataset . . . . .	18
<b>3 Homogenization</b>	<b>21</b>
3.1 Introduction . . . . .	21
3.2 SNHT for the homogenization of snow depth data	23
3.2.1 The algorithm . . . . .	23
3.2.2 Testing procedure of the algorithm . . . . .	30
3.3 Homogenization of Trentino snow depth data . . .	34
3.3.1 Detected breakpoints . . . . .	34
3.3.2 Validation using artificially created break- points . . . . .	40
3.4 Intercomparison Experiment . . . . .	43
3.4.1 HOMOP . . . . .	44
3.4.2 Breakpoint detection . . . . .	45
3.4.3 Correction of the inhomogeneities . . . . .	47
3.4.4 Implication for time series analysis . . . . .	48
3.5 Conclusions . . . . .	51

---

<b>4</b>	<b>Wavelet analysis</b>	<b>53</b>
4.1	Introduction . . . . .	53
4.2	Continuous wavelet transform . . . . .	53
4.2.1	Wavelet coherence . . . . .	55
4.3	Application to snow data . . . . .	56
4.3.1	Analysis of the Trentino - Alto Adige snow depth dataset . . . . .	56
4.3.2	Wavelet transform . . . . .	64
4.3.3	Wavelet coherence with the NAOI and the MOI . . . . .	68
4.4	Application to discharge data . . . . .	73
4.4.1	Continuous wavelet transform . . . . .	74
4.4.2	Wavelet coherence analysis . . . . .	77
4.5	Conclusions . . . . .	83
<b>5</b>	<b>Copula</b>	<b>87</b>
5.1	Introduction . . . . .	87
5.1.1	Measures of dependence . . . . .	89
5.1.2	Tail dependence . . . . .	89
5.1.3	Bivariate copula families . . . . .	90
5.1.4	Multivariate copula families . . . . .	92
5.1.5	Empirical copula . . . . .	93
5.1.6	Software . . . . .	96
5.2	Analyzing snow dynamics with copula . . . . .	96
5.2.1	Mean seasonal snow depth and snow cover duration . . . . .	97
5.2.2	Mean seasonal snow depth and temperature	99
5.2.3	Snow cover duration and temperature . . .	100
5.2.4	Mean seasonal snow depth and MOI . . . .	102
5.3	Modeling flood events with copula . . . . .	104
5.3.1	Bivariate application: direct peak discharge - direct volume modeling . . . . .	105
5.3.2	Multivariate application: $Q_D-V_D-T_R-Q_B$ . .	110
5.3.3	Multivariate application: $Q_D-V_D-T_R-Th$ . .	117
5.4	Conclusions . . . . .	125
<b>6</b>	<b>Conclusions</b>	<b>127</b>
<b>A</b>	<b>4-dimensional vine copulas</b>	<b>133</b>



# List of Figures

2.1	Map of the stations of the Province of Trento . . . . .	6
2.2	Number of data of the time series. . . . .	7
2.3	Number of available time series per year. . . . .	7
2.4	Map of the stations of the intercomparison experiment	8
2.5	Location of the analyzed time series. . . . .	11
2.6	Location of the examined gauging stations in the Adige river basin. . . . .	15
2.7	Location of the examined gauging stations in the Adige river basin . . . . .	16
2.8	Map of the Inn River Basin with the location of the analyzed gauging stations . . . . .	19
3.1	Scheme of the homogeneity test for the detection of the breakpoints. . . . .	24
3.2	Scheme of the semihierarchical splitting and merg- ing algorithm . . . . .	27
3.3	General scheme of the testing procedure . . . . .	31
3.4	Example of candidate reference time series for the second phase of the homogenization analysis . . . . .	33
3.5	Distribution of the self-homogeneity breakpoints . . . . .	38
3.6	Example of breakpoint correction: Val Noana time series . . . . .	39
3.7	Example of breakpoint correction: Madonna di Campiglio - Pancugolo time series . . . . .	40
3.8	Percentage of detected artificial breakpoints . . . . .	41
3.9	Influence of artificial breakpoints on the time series of Passo Valles . . . . .	42
3.10	Example of the time series used for the intercom- parison experiment . . . . .	48

---

3.11	Anomalies of Galtür before and after the correction of the breakpoint . . . . .	49
3.12	Anomalies of Bad Gastein before and after the correction of the breakpoint . . . . .	50
3.13	Anomalies of St.Leonhard i.P. before and after the correction of the breakpoint . . . . .	51
4.1	Hovmöller-type diagramm of the mean seasonal snow depth, 5-year moving average of $HS_m$ and 5-year moving average of the snow depth . . . . .	57
4.2	Hovmöller-type diagramm of the snow cover duration, 5-year moving average of $SCD_m$ and 5-year moving average of the snow cover . . . . .	58
4.3	Moving average of the mean seasonal maximum temperature at different altitude classes. . . . .	61
4.4	Anomalies of the mean seasonal maximum temperature . . . . .	61
4.5	Correlation between the mean seasonal maximum temperature and the mean seasonal snow depth . . . . .	63
4.6	Correlation between the mean seasonal maximum temperature and the snow cover duration. . . . .	64
4.7	Wavelet spectrum of the average mean seasonal snow depth for different altitude classes. . . . .	65
4.8	Wavelet spectrum of the average snow cover for different altitude classes. . . . .	66
4.9	Global wavelet spectrum for different altitude classes. . . . .	67
4.10	Scale average time series between the periods of 2 and 8 years for different altitude classes. . . . .	67
4.11	Wavelet coherence analysis between the average mean seasonal snow depth of the stations below 1350 m a.s.l. and the NAOI (plot above) and the MOI (plot below). . . . .	69
4.12	Wavelet coherence analysis between the average mean seasonal snow depth of the stations between 1350 m and 1650 m a.s.l. and the NAOI (plot above) and the MOI (plot below). . . . .	70
4.13	Wavelet coherence analysis between the average mean seasonal snow depth of the stations between 1650 m and 2000 m a.s.l. and the NAOI (plot above) and the MOI (plot below). . . . .	71

4.14	Wavelet coherence analysis between the average mean seasonal snow depth of the stations below 2000 m a.s.l. and the NAOI (plot above) and the MOI (plot below). . . . .	72
4.15	Comparison of the global wavelet spectrum for the gauging stations of the Adige (left column) and for the Inn (right column) catchment. . . . .	74
4.16	Comparison of the 2-8 years scale-average for the stations of the Adige (left column) and for the Inn (right column) catchment. . . . .	75
4.17	Wavelet coherence analysis with the NAOI for the gauging stations of the Adige river catchment. . . .	78
4.18	Wavelet coherence analysis with the NAOI for the gauging stations of the Inn river catchment. . . . .	79
4.19	Wavelet coherence analysis with the MOI for the gauging stations of the Adige river catchment. . . .	81
4.20	Wavelet coherence analysis with the MOI for the gauging stations of the Inn river catchment. . . . .	82
5.1	Example of density distribution of a normal copula	90
5.2	Example of density distribution of a t-copula . . . .	91
5.3	Structure of a 3-dimensional vine copula . . . . .	93
5.4	Example of possible structures of a D-vine copula.	94
5.5	Examples of possible structures of a C-vine copula.	94
5.6	Scheme of the procedure applied for the application of the copula procedure and the generation of simulated data. . . . .	95
5.7	Comparison of the pseudoobservations of mean seasonal snow depth and snow cover duration with the isolines of the fitted copula. . . . .	97
5.8	Comparison of the pseudoobservations of mean seasonal snow depth and temperature with the isolines of the fitted copula. . . . .	100
5.9	Comparison of the pseudoobservations of snow cover duration and temperature with the isolines of the fitted copula. . . . .	101
5.10	Comparison of the pseudoobservations of mean seasonal snow depth and MOI with the isolines of the fitted copula. . . . .	102
5.11	Example of the determination of $Q_{MHQ}$ and $V_{MHQ}$ .	105

---

5.12	Autocorrelation analysis for the Q-V time series of Wasserburg. . . . .	106
5.13	Autocorrelation analysis for the Q-V time series of Passau. . . . .	107
5.14	Comparison between the observed pseudo-observations and the simulated ones of the gauging stations of Wasserburg and Passau. . . . .	108
5.15	Comparison between the isolines of the cdf of the empirical copula of the observed and of the simulated pseudo-observations of the gauging stations of Wasserburg and Passau. . . . .	108
5.16	Comparison between the observed and simulated $Q_MHQ-V_MHQ$ data for the gauging station of Wasserburg. . . . .	109
5.17	Comparison between the observed and simulated $Q_MHQ-V_MHQ$ data for the gauging station of Passau Ingling. . . . .	109
5.18	Example of the determination of the direct flood peak $Q_D$ , the direct volume $V_D$ , the rising time $T_R$ and the base flow $Q_B$ . . . . .	110
5.19	Coupled scatter plots for the variables $Q_D$ , $V_D$ , $T_R$ and $Q_B$ for Wasserburg. . . . .	112
5.20	Coupled scatter plots for the variables $Q_D$ , $V_D$ , $T_R$ and $Q_B$ for Passau. . . . .	112
5.21	Coupled scatter plots for the pseudo-observations of the variables $Q_D$ , $V_D$ , $T_R$ and $Q_B$ for Wasserburg. . . . .	113
5.22	Coupled scatter plots for the pseudo-observations of the variables $Q_D$ , $V_D$ , $T_R$ and $Q_B$ for Passau. . . . .	113
5.23	Isolines of the cdf of the bivariate empirical copula between each couple of the variables $Q_D$ , $V_D$ , $T_R$ and $Q_B$ of Wasserburg. . . . .	114
5.24	Isolines of the cdf of the bivariate empirical copula between each couple of the variables $Q_D$ , $V_D$ , $T_R$ and $Q_B$ of Passau. . . . .	115
5.25	Coupled scatter plots of the simulated and of the observed values of the variables $Q_D$ , $V_D$ , $T_R$ and $Q_B$ for Wasserburg . . . . .	116
5.26	Coupled scatter plots of the simulated and of the observed values of the variables $Q_D$ , $V_D$ , $T_R$ and $Q_B$ for Passau . . . . .	116

5.27	Example of the determination of the direct flood peak $Q_D$ , the direct volume $V_D$ , the rising time $T_R$ and the threshold $Th$ . . . . .	117
5.28	Coupled scatter plots for the variables $Q_D$ , $V_D$ , $T_R$ and $Th$ for Wasserburg. . . . .	119
5.29	Coupled scatter plots for the variables $Q_D$ , $V_D$ , $T_R$ and $Th$ for Passau. . . . .	119
5.30	Coupled scatter plots for the pseudo-observations of the variables $Q_D$ , $V_D$ , $T_R$ and $Th$ for Wasserburg. . . . .	120
5.31	Coupled scatter plots for the pseudo-observations of the variables $Q_D$ , $V_D$ , $T_R$ and $Th$ for Passau. . . . .	120
5.32	Isolines of the cdf of the bivariate empirical copula between each couple of the variables $Q_D$ , $V_D$ , $T_R$ and $Th$ of Wasserburg. . . . .	122
5.33	Isolines of the cdf of the bivariate empirical copula between each couple of the variables $Q_D$ , $V_D$ , $T_R$ and $Th$ of Passau. . . . .	123
5.34	Coupled scatter plots of the simulated values of the variables $Q_D$ , $V_D$ , $T_R$ and $Th$ for Wasserburg. . . . .	123
5.35	Coupled scatter plots of the simulated values of the variables $Q_D$ , $V_D$ , $T_R$ and $Th$ for Passau. . . . .	124



# List of Tables

2.1	List of the stations analyzed for homogeneity in the intercomparison experiment. . . . .	9
2.2	Location of the stations below 1350 m a.s.l. . . . .	12
2.3	Location of the stations between 1350 m and 1650 m a.s.l. . . . .	12
2.4	Location of the stations between 1650 m and 2000 m a.s.l. . . . .	13
2.5	Location of the stations above 2000 m a.s.l. . . . .	13
2.6	Location of the temperature stations. . . . .	14
2.7	Location of the gauging stations. . . . .	17
2.8	Characteristics of the gauging stations. . . . .	18
2.9	Location and dimension of the drainage basin of the gauging stations. . . . .	18
3.1	Scheme of the procedure for the application of the algorithm for the detection of the breakpoints. . .	32
3.2	List of the inhomogeneous time series. . . . .	35
3.3	Results of the homogenization of the snow depth dataset of the Province of Trento . . . . .	36
3.4	Inconsistent results of the homogenization of the snow depth dataset of the Province of Trento. . . .	37
3.5	Stations forming the VAL NOANA site. . . . .	39
3.6	Results of the intercomparison experiment . . . . .	46
3.7	Correction factors of the intercomparison experiment	47
4.1	P-values of the Kolmogorov-Smirnoff applied to the average mean seasonal snow depth . . . . .	59
4.2	P-values of the Mann-Whitney Tests applied to the average mean seasonal snow depth . . . . .	60

---

4.3	Correlation coefficients of the mean seasonal snow depth and of the snow cover duration with the mean maximum temperature . . . . .	63
5.1	Examples of Archimedean copulas [Embrechts et al., 2001]. . . . .	92
5.2	Kendall's tau $\tau_k$ and the Spearman's rho $\rho_S$ correlation coefficients of the observed and of the simulated $Q_{MHQ}$ and $V_{MHQ}$ . . . . .	110
5.3	Kendall's tau $\tau_k$ and the Spearman's rho $\rho_S$ correlation coefficients of the variables $Q_D$ , $V_D$ , $T_R$ and $Q_B$ for Wasserburg and Passau. . . . .	111
5.4	Kendall's tau $\tau_k$ and the Spearman's rho $\rho_S$ correlation coefficients of the simulated variables $Q_D$ , $V_D$ , $T_R$ and $Q_B$ for Wasserburg and Passau . . . .	114
5.5	Kendall's tau $\tau_k$ and the Spearman's rho $\rho_S$ correlation coefficients of the variables $Q_D$ , $V_D$ , $T_R$ and $Th$ for Wasserburg and Passau. . . . .	118
5.6	Kendall's tau $\tau_k$ and the Spearman's rho $\rho_S$ correlation coefficients of the simulated variables $Q_D$ , $V_D$ , $T_R$ and $Th$ for Wasserburg and Passau. . . . .	121
A.1	Vine Copula for the modeling of the variables $Q_D$ , $V_D$ , $T_R$ and $Q_B$ of Wasserburg. . . . .	133
A.2	Vine Copula for the modeling of the variables $Q_D$ , $V_D$ , $T_R$ and $Q_B$ of Passau. . . . .	134
A.3	Vine Copula for the modeling of the variables $Q_D$ , $V_D$ , $T_R$ and $Th$ of Wasserburg. . . . .	134
A.4	Vine Copula for the modeling of the variables $Q_D$ , $V_D$ , $T_R$ and $Th$ of Passau. . . . .	134



# Abstract

In hydrological studies, it is very important to properly analyze the relationship among the different components of the water cycle, due to the complex feedback mechanisms typical of this system. The analysis of available time series is hence a fundamental step, which has to be performed before any modeling activity. Moreover, time series analysis can shed light over the spatial and temporal dynamics of correlated hydrological and climatological processes. In this work, we focus on three tools applied for time series analysis: homogeneity tests, wavelet analysis and copula analysis.

Homogeneity tests allow to identify a first important kind of variability in the time series, which is not due to climate nor seasonal variability. Testing for inhomogeneities is therefore an important step that should be always performed on a time series before using it for any application. The homogenization of snow depth data, in particular, is a challenging task. Up to now, it has been performed analyzing available metadata, which often present contradictions and are rarely complete. In this work, we present a procedure to test the homogeneity of snow depth time series based on the Standard Normal Homogeneity Test (SNHT). The performance of the SNHT for the detection of inhomogeneities in snow depth data is further investigated with a comparison experiment, in which a dataset of snow depth time series relative to Austrian stations has been analyzed with both the SNHT and the HOMOP algorithm. The intercomparison study indicates that the two algorithms show comparable performance.

The wavelet transform analysis allows to obtain a different kind of information about the variability of a time series. In fact, it determines the different frequency content of a signal in different time intervals. Moreover, the wavelet coherence analysis allows to identify periods where two time series are correlated and their phase

shift. We apply the wavelet transform to a dataset of snow depth time series of stations distributed in the Adige catchment and on a dataset of 16 discharge time series located in the Adige and in the Inn catchments. The same datasets are used to perform a wavelet coherence analysis considering the Mediterranean Oscillation Index (MOI) and the North Atlantic Oscillation Index (NAOI). This analysis highlights a difference in the behavior of the snow time series collected below and above 1650 m a.s.l.. We also observe a difference between low and high elevation sites in the amount of mean seasonal snow depth and snow cover duration. More interestingly, snow time series collected at different elevations respond differently to temperature and more in general to climate changes. The wavelet analysis allows us also to distinguish between gauging stations belonging to different catchments, while the wavelet coherence analysis revealed non-stationary correlations with the MOI and NAOI, indicating a very complex relation between the measured quantities and climatic indexes.

Finally the application of copulas allows modeling the marginal of each variable and their dependence structure independently. We apply this technique to two relevant cases. First we study snow related variables in relation with temperature, the NAOI and the MOI, which we already investigated with the wavelet coherence analysis. Then we model flood events registered at two stations of the Inn river: Wasserburg and Passau. This last analysis is performed with the goal of predicting future flood events and derive construction parameters for retention basins. We test three different combinations of variables (direct peak discharge-direct volume, direct peak discharge-direct volume-rising time-base flow, direct peak discharge-direct volume-rising time-moving threshold) describing the flood events and compare the results. The consistency in the results indicates that the proposed methodology is robust and reliable.

This study shows the importance of approaching the analysis to hydrological time series from several points of view: quality of the data, variability of the time series and relation between different variables. Moreover, it shows that integrating the use of various time series analysis methods can greatly improve our understanding of the system behavior.

# Abstract

Negli studi idrologici é molto importante analizzare adeguatamente la relazione tra le diverse componenti coinvolte nel ciclo dell'acqua, per via del complesso meccanismo di risposta tipico di questo sistema. L'analisi delle serie temporali a disposizione rappresenta perciò un passaggio fondamentale, che deve essere effettuato prima di qualsiasi attività di modellazione. Inoltre, l'analisi delle serie temporali può chiarire la correlazione esistente tra dinamiche spaziali e temporali di processi idrologici e climatologici. In questo lavoro, ci concentriamo su tre metodologie applicate nell'analisi di serie temporali: test di omogeneità, analisi wavelet e analisi per mezzo delle copule.

I test di omogeneità permettono di identificare un primo tipo importante di variabilità nelle serie temporali, che non é dovuto né ad una variabilità climatica né ad una stagionale. Il controllo della presenza di inomogeneità é perciò un passaggio importante che dovrebbe essere sempre effettuato prima del loro utilizzo per qualsiasi applicazione. L'omogenizzazione di dati di neve al suolo, in particolare, é un compito impegnativo. Fino ad oggi é stato effettuato analizzando i metadati disponibili, i quali però presentano spesso contraddizioni e sono raramente completi. In questo lavoro, presentiamo una procedura per verificare l'omogeneità delle serie di neve al suolo basata sullo Standard Normal Homogeneity Test (SNHT). Le prestazioni dell'SNHT per il rilevamento di inomogeneità nei dati di neve al suolo sono investigate inoltre attraverso un esperimento di confronto, nel quale un set di dati di serie temporali di neve al suolo relativi a stazioni austriache é stato analizzato con i due algoritmi SNHT e HOMOP. Questo studio indica che i due algoritmi hanno prestazioni comparabili.

L'analisi wavelet permette di ottenere un diverso tipo di informazione riguardo alla variabilità delle serie temporali. Infatti, determina il diverso contenuto di frequenze di un segnale in di-

versi intervalli temporali. Inoltre, la cosiddetta *wavelet coherence* permette di identificare sia periodi dove due serie temporali sono correlate che la loro differenza di fase. In questo lavoro, applichiamo la trasformata wavelet ad un set di dati di serie temporali di neve al suolo relative a stazioni distribuite nel bacino dell'Adige e su un set di dati di 16 serie temporali di portata relative a stazioni situate nei bacini dell'Adige e dell'Inn. Gli stessi set di dati sono usati per effettuare un'analisi di coerenza wavelet considerando il *Mediterranean Oscillation Index* (MOI) ed il *North Atlantic Oscillation Index* (NAOI). Questa analisi evidenzia una differenza nel comportamento delle serie temporali di neve registrate al di sotto e al di sopra dei 1650 m s.l.m.. Si può anche osservare una differenza tra i siti a quote basse ed alte in termini di media stagionale di neve al suolo e di durata di copertura nevosa. Inoltre, le serie temporali di neve al suolo registrate a diverse quote rispondono diversamente a cambiamenti di temperatura e, più in generale, ai cambiamenti climatici. L'analisi wavelet permette anche di distinguere tra idrometri appartenenti a diversi bacini, mentre la *wavelet coherence* rivela correlazioni non stazionarie con il MOI ed il NAOI, indicando una relazione molto complessa tra le quantità misurate e gli indici climatici.

Infine, l'applicazione delle copule permette di modellare indipendentemente le distribuzioni marginali di ogni variabile e la loro struttura di dipendenza. Appliciamo questa tecnica a due casi scientificamente rilevanti. Prima studiamo le variabili legate alla neve in relazione alla temperatura, al NAOI e al MOI, come abbiamo già effettuato utilizzando la *wavelet coherence*. Quindi modelliamo eventi di piena registrati a due stazioni sul fiume Inn: Wasserburg e Passau. Quest'ultima analisi è effettuata con lo scopo di predire eventi di piena futuri e di derivare parametri di costruzione per bacini di ritenzione. Testiamo tre diverse combinazioni di variabili (portata massima diretta-volume diretto, portata massima diretta-volume diretto-tempo di crescita-portata di base, portata massima diretta-volume diretto-tempo di crescita-soglia mobile) per descrivere gli eventi di piena e confrontare i risultati. La consistenza nei risultati indica che la metodologia proposta è robusta ed affidabile.

Questo studio dimostra l'importanza di affrontare l'analisi di serie temporali idrologiche da diversi punti di vista: la qualità dei dati, la variabilità delle serie temporali e la relazione tra diverse variabili. Inoltre, mostra come integrando l'uso di diversi metodi

di analisi di serie temporali si possa migliorare sensibilmente la nostra comprensione del comportamento del sistema.



# Abstract

In hydrologischen Studien ist es auf Grund der für dieses System typischen komplexen Rückkopplungseffekte sehr wichtig, die Zusammenhänge zwischen verschiedenen Komponenten des Wasserkreislaufes genau zu analysieren. Die statistische Analyse der verfügbaren Zeitreihen ist also ein fundamentaler Schritt, der vor jeder Modellierung durchgeführt werden sollte. Insbesondere kann mit Hilfe der Zeitreihenanalyse die räumliche und zeitliche Dynamik von korrelierten hydrologischen und klimatologischen Prozessen aufgezeigt werden. Diese Arbeit konzentriert sich auf drei Verfahren, die in der Zeitreihenanalyse Anwendung finden: Homogenitätstests, Wavelet-Analyse und Copula-Analyse.

Mit Hilfe von Homogenitätstests kann ein erster wichtiger Variabilitätstyp in den Zeitreihen festgelegt werden, der nicht auf klimatologische oder saisonale Variabilität zurückzuführen ist. Das Testen auf Inhomogenität ist deshalb ein notwendiger Schritt, der immer durchgeführt werden sollte, bevor eine Zeitreihe für eine Anwendung benutzt wird. Homogenitätstests und insbesondere die Homogenisierung von Schneehöhendaten sind eine anspruchsvolle Aufgabe. Bisher wurde Letzteres mit verfügbaren Metadaten durchgeführt, wobei diese aber häufig Widersprüche zeigen und selten vollständig sind. In dieser Arbeit wird eine Methode zur Überprüfung der Homogenität von Schneehöhendaten vorgestellt, die auf dem Standard Normal Homogeneity Test (SNHT) basiert. Die Leistungsfähigkeiten des SNHT für die Detektion von Inhomogenitäten in Schneehöhendaten wird zusätzlich mit einem Vergleichsexperiment getestet. Dabei wird einen Datensatz von Schneehöhen österreichischer Stationen mit den SNHT- und HOMOP- Algorithmen analysiert. Die Vergleichsstudie zeigt, dass die Leistungsfähigkeit der beiden Algorithmen vergleichbar ist.

Die Wavelet-Transformations-Analyse ermöglicht es, eine andere Art von Informationen über die Variabilität der Zeitreihen

zu erhalten, denn sie bestimmt die inhärenten Frequenzen eines Signals in verschiedenen Zeitintervallen. Die Wavelet-Kohärenz-Analyse erlaubt es, Perioden von zwei korrelierten Zeitreihen und deren Phasenverschiebung zu bestimmen. Es werden Wavelet-Transformationen an Schneehöhen-Zeitreihen des Adige-Einzugsgebiets und an 16 Abflusszeitreihen der Adige- und Inn-Einzugsgebiete angewendet. Dieselben Datensätze werden verwendet, um eine Wavelet-Kohärenzanalyse mit dem Mediterranen Oszillations Index (MOI) und mit dem Nordatlantischen Oszillations Index (NAOI) durchzuführen. Diese Analyse ergibt ein unterschiedliches Verhalten von Schneezeitreihen, die unterhalb und oberhalb von 1650 m u.N.N. gemessen wurden. Weiterhin wurden Unterschiede zwischen höher und niedriger gelegenen Stationen bei der saisonalen mittleren Schneehöhe und der Schneedeckendauer festgestellt. Zusätzlich reagieren Schneezeitreihen verschiedener Höhen unterschiedlich auf Veränderungen der Temperatur und auf Klimaveränderungen. Die Wavelet-Analyse erlaubt auch die Unterscheidung von Pegeln verschiedener Einzugsgebiete. Die Wavelet-Kohärenzanalyse weist nicht-stationäre Korrelationen mit dem MOI und mit dem NAOI nach, wobei diese einen sehr komplexen Zusammenhang zwischen den gemessenen Größen und den klimatischen Indizes zeigen.

Schließlich erlaubt die Anwendung von Copulas die Randverteilungen jeder Variable und ihre gemeinsame Abhängigkeitsstruktur unabhängig zu modellieren. Diese Methode wird für zwei Fälle angewendet. Zuerst werden Schneevariablen bezüglich der Temperatur, des NAOI und des MOI untersucht, vergleichbar zu den Wavelet-Kohärenzanalysen. Anschließend werden Hochwasserwellen modelliert, die an zwei Pegeln am Inn (Wasserburg und Passau) gemessen wurden. Diese Analyse wird mit dem Ziel durchgeführt, Bemessungswerte bzw. ereignisse für den Bau von Poldern abzuleiten. Drei verschiedene Kombinationen von Variablen (Direktabflussscheitel - Direktabflussvolumen, Direktabflussscheitel - Direktabflussvolumen Anstiegszeit - Basisabfluss, Direktabflussscheitel - Direktabflussvolumen - Anstiegszeit - variabler Schwellenwert) werden für die Beschreibung von Hochwasserereignissen getestet. Die Konsistenz in den Ergebnissen zeigt, dass die vorgeschlagene Methodik robust und zuverlässig ist.

Diese Studie zeigt, wie wichtig es ist, bei der Analyse von hydrologischen Zeitreihen verschiedene Aspekte zu untersuchen: Datenqualität, Variabilität der Zeitreihen und Zusammenhänge



verschiedener Variablen. Zudem zeigt sie, dass die Integration verschiedener Methoden der Zeitreihenanalyse das Verständnis des Systemverhaltens nachhaltig verbessern kann.



# Chapter 1

## Introduction

The analysis of time series is a practice common to many fields of science and several techniques have been developed in the past years for this goal [see e.g. Anderson, 2011, Chatfield, 2016]. A time series shows how a certain variable has changed as function of time. Time series can be a valuable source of information, in particular in our case they can bring information about the functioning of the hydrological cycle at different elevations, as well as about climate changes. These information are encoded in the time series as different kinds of variability. When analyzing a time series, our goal is to identify these different variations and address them to their probable causes. In this work we deal with hydrological time series, and in particular with snow depth and river discharge time series relative to alpine catchments. Hydrological time series relative to alpine hydrology are conditioned by the fact, that most of the territory we consider is mountainous. This does not affect only the variables themselves, but, as shown in this work, also their interaction.

Another aspect specific of the alpine regions is the difficulty in collecting reliable and representative data. Some locations are in fact difficult to reach and monitor. The exposition of measuring stations has an important role and relocation of the stations can strongly affect the time series, especially if they are located on steep slopes. Therefore, the first kind of variability, which is important to identify, is the one that is not depending on climatic or seasonal variability. These variations may have an anthropogenic cause, such as the relocation of a station or the change of the measuring equipment [see e.g., Hanssen-Bauer and

Førland, 1994, Brunetti et al., 2006, Aguilar et al., 2003]. Since our goal is to investigate time series to better understand their variability driven by physical forcings, we need to identify these spurious changes. The homogeneity analysis allows to identify those time series, whose variability has been strongly affected by the anthropogenic causes mentioned before. Several algorithms based on statistical tests have been developed for the homogenization of variables such as temperature and precipitation [see e.g. Easterling and Peterson, 1995, Vincent, 1998, Aguilar et al., 2003, Caussinus and Mestre, 2004, Mestre et al., 2013], but, to the best of our knowledge, in most cases the homogeneity of the mean seasonal snow depth has been mainly performed with the analysis of the metadata (i.e. all documents regarding the history of the tested station), which are unfortunately often not complete and not fully reliable [Marcolini et al., 2017]. Snow is a very important factor in mountainous regions, since it affects many aspects of the human life, such as water availability [Beniston et al., 2003, Barnett et al., 2005, Beniston, 2006], the ecosystem functioning [Theurillat and Guisan, 2001] and economical activities, such as the winter tourism [Koenig and Abegg, 1997] and hydropower production [Beniston, 2012a, Majone et al., 2015]. Moreover, snow-pack is extremely sensitive to small variations in temperature and precipitation [Barnett et al., 2005, Bartlett et al., 2004, Beniston et al., 2003, Beniston, 2006] and exerts an important control on the hydrological cycle of Alpine catchments [Chiogna et al., 2014, Penna et al., 2014]. For these reasons it is important to develop a robust homogenization procedure for snow depth time series. We present an algorithm based on the Standard Normal Homogeneity Test [Alexandersson, 1986, Alexandersson and Moberg, 1997] for the homogenization of mean seasonal snow depth data and apply it for the dataset available in the Province of Trento. We also show an intercomparison experiment with the Central Institute of Meteorology and Geodynamics of Vienna, which applied for the same goal the algorithm HOMOP [Vincent et al., 2002, Caussinus and Mestre, 2004, Nemec et al., 2013].

Once we have established that there is no evidence, that the variability of the examined time series derives from anthropogenic factors, we can further analyze it in order to evaluate its correlation with other components of the hydrological cycle or climate indexes. The goal here is to analyze its response to the climate changes recorded in the last years. The first analysis of this kind

---

shown in this work is the wavelet analysis.[see e.g., Lau and Weng, 1995, Torrence and Compo, 1998, Coulibaly and Burn, 2004, Grinsted et al., 2004, Guan et al., 2011, Carey et al., 2013]. Similarly to the Fourier transform analysis, the wavelet transform analysis decomposes the signal of the analyzed time series in different signals with different periods. The advantage of the wavelet analysis respect to the Fourier transform analysis is that it allows to analyze the components of a non-stationary signal, while Fourier transform does not contain any local information [see e.g. Sifuz-zaman et al., 2009]. Another possibility offered by this kind of analysis is the wavelet coherence analysis [Torrence and Compo, 1998, Grinsted et al., 2004]. This method investigates the coherence between two signals in time frequency space. In this work we show the application of the wavelet analysis for the study of mean seasonal snow depth in the Adige catchment and discharge data of the Adige and Inn river basins.

The complex feedback mechanisms typical of the water cycle makes it a challenging task to investigate the dependence structure between different time series. Nonetheless, this is a very important aspect to be analyzed, since it can shed light over the spatial and temporal dynamics of correlated hydrological and climatological processes. In this case a valuable tool is the copula analysis [Joe, 1997, Nelsen, 1999, Genest and Favre, 2007, Yan et al., 2007]. Copulas have been initially applied in finance [Genest and Favre, 2007], but in the last years have found a wider and wider application also in the field of hydrology [see e.g. Favre et al., 2004, Salvadori and De Michele, 2004, 2007, Bárdossy and Li, 2008, Grimaldi et al., 2016]. Copulas are multivariate distributions with uniform marginals. Their most valuable property is that they allow to model the dependence structure between two or more variables and their marginals separately and are as such a pure expression of the dependence structure [Joe, 1997, Nelsen, 1999, Genest and Favre, 2007, Yan et al., 2007]. In this work, they are used for two applications. We first apply them in order to further study the correlation between snow related variables and temperature or climate indexes. We show how copula analysis can reveal more information about the relation between different variables than standard time series statistical analysis. We also show the performance of the copula for modeling flood events with three different variables definition for two gauging stations of the Inn River. In the first case we only analyze the direct peak

discharge and direct volume of the event respect to a threshold, fixed as the mean yearly maximum discharge of the gauging station. In the two last cases also the rising time and a threshold are modeled. The threshold is first defined as the baseflow, and then as an optimized threshold for the calculation of the retention potential with the Kozeny function. It is shown how the relationship between the variables changes depending on their definition and the difficulties and advantages of fitting copulas with a number of dimensions larger than 2.

This work is structured as it follows. In chapter 2 we introduce the datasets that are analyzed in this study. The algorithm based on the Standard Normal Homogeneity Test for the homogenization of mean seasonal snow depth and the intercomparison experiment with the Central Institute of Meteorology and Geodynamics of Vienna are presented in chapter 3. Chapter 4 presents the continuous wavelet transform and wavelet coherence analysis of the mean seasonal snow depth of the Adige catchment and discharge datasets of the Adige and Inn river basins. The results of the wavelet analysis of the mean seasonal snow depth are compared with the findings of the statistical analysis of the snow related variables in the Adige River basin, which is also presented in chapter 4. In chapter 5 we show the copula analysis of the mean seasonal snow depth and snow cover duration in the Adige river basin and we present the performance of three schemes for the modeling of flood events in the Inn river using copulas.

## Chapter 2

# Description of the datasets

In this work we analyze five different datasets by means of homogenization algorithms, wavelet transform and copulas. These datasets contain snow depth, temperature and discharge data. In sections 2.1 and 2.2 two snow depth datasets regarding the Province of Trento (Italy) and Austria are presented, whose homogenization is then shown in chapter 3. Section 2.3 describes a snow depth and temperature dataset. This will be used in chapter 4 for statistical and wavelet analysis to investigate the climatological changes in this area in the last decades. The snow and temperature dataset of section 2.3 are also analyzed using copulas in chapter 5. A statistical study is also conducted in chapter 4 on the discharge dataset of the Adige and Upper Inn rivers catchments presented in section 2.4. Further discharge data, but relative to the lower part of the Inn river catchment, are presented in section 2.5. These will be used in chapter 5 for the modeling of flood events with copulas.

### 2.1 Trentino snow depth dataset

The Trentino snow depth dataset [Marcolini et al., 2017] comprehends 106 mean seasonal snow depth (HS) time series collected in the Province of Trento (Italy). The time series of this dataset will be tested for homogeneity in chapter 3. This dataset is particularly valuable since it is composed by well correlated snow depth time series recorded at a high density network of snow stations

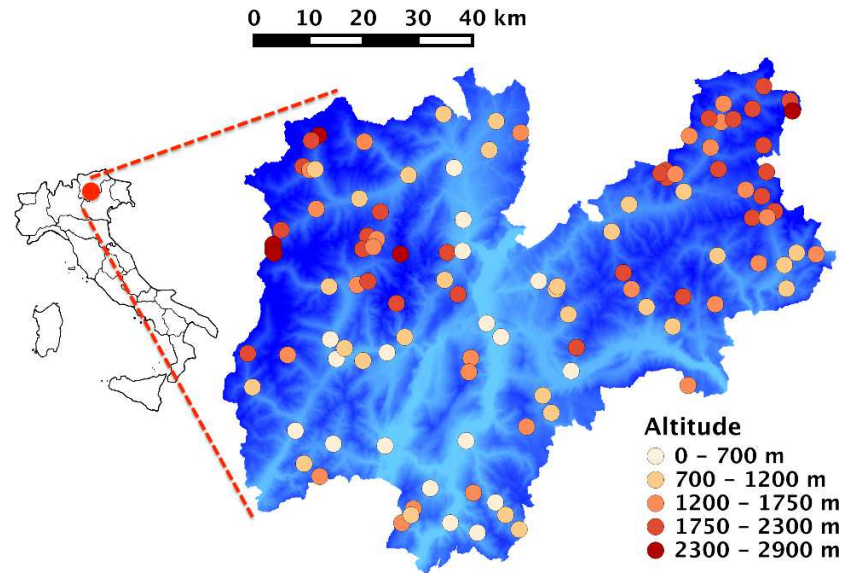


Figure 2.1: Location of the studied sites in the Province of Trento (north-east of Italy). The color of the dots indicates the site altitude.

(on average one station each  $200 \text{ Km}^2$ , see Figure 2.1), which allows to overcome the problems observed for example by Begert et al. [2008] in the homogenization of snow depth time series due to poor spatial and temporal correlation. The spatial density of the stations is comparable to that used in other studies [see e.g., Laternser and Schneebeli, 2003a, Hantel et al., 2000]. The mean seasonal values were computed by averaging the daily snow depth between November 1 and April 30. The time series were formed merging quality checked data from different sources, according to quality indices criteria, and missing data have not been interpolated. Manual data (measured by operators of the Province of Trento directly in the field) had the highest quality index, followed by historical data (collected from different sources, such as the Central Institute for Meteorology and Geodynamics of Vienna and the Hydrographic office of the Province of Trento) and automatic data (measured from automatic instruments). Historical data are generally considered less reliable than the manual ones because the procedure used for their collection is not always known [Marcolini et al., 2017].

Figure 2.1 shows the spatial distribution of the sites in the



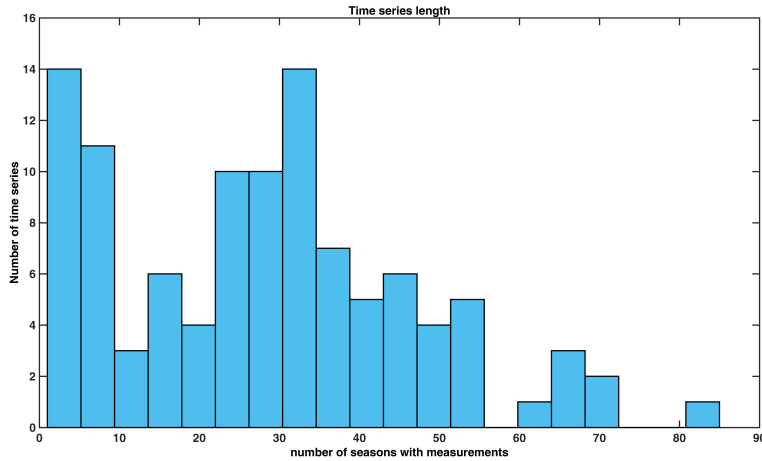


Figure 2.2: Number of data of the time series.

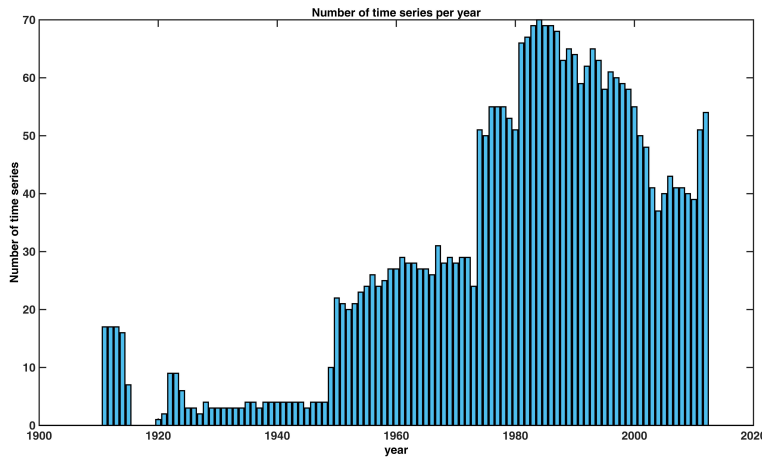


Figure 2.3: Number of available time series per year.

Province of Trento and their altitude. Most of the stations are at an altitude above 1000 m a.s.l., and, in particular, 50% of the sites are located between 1300 and 2000 m a.s.l.. 25 time series are shorter than 10 years, 55% of the time series contain between 10 and 40 years of data (Figure 2.2).

As shown in Figures 2.2 and 2.3, some of the time series start at the beginning of the 20th century and many of them extend up to the winter season 2012/13. Nevertheless, they often cover different time intervals and sometimes contain gaps, in some cases lasting for several years. The number of time series increases sharply in

the early fifties and in the early eighties (Figure 2.3). All these factors hamper the homogenization of the snow depth time series.

## 2.2 Austrian snow depth dataset

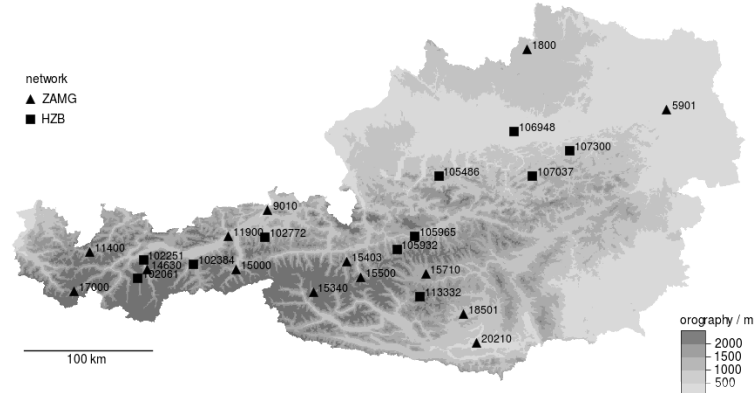


Figure 2.4: Map of the station used for the intercomparison experiment. The black triangle (square) represents stations in the ZAMG (HZB) observational network.

The Austrian snow depth dataset analyzed in this work comprises stations distributed in the whole Austria. The distribution is not uniform, due to the presence of a complex mountainous terrain. Figure 2.4 shows the 25 stations listed in Table 2.1, which will be analyzed for homogeneity in section 3.4. The stations belong to two major networks: the Central Institute for Meteorology and Geodynamics (ZAMG, triangles in Figure 2.4) and the Hydrographical Central Bureau (HZB, squares in Figure 2.4). These stations are located in different representative climatic regions of Austria and the mean length of snow depth time series is 73 years. The lowest station is Wien Hohe Warte at 198 m, while the highest one is Galtür at 1577 m a.s.l.

The time series of the ZAMG have been proofed with an extensive quality control and were tested for inner and outer consistency, respectively [Koch et al., 2014]. Metadata regarding for example

Table 2.1: List of the stations analyzed for homogeneity in the intercomparison experiment.

<b>Code</b>	<b>Station</b>	<b>Altitude</b>
5901	Wien Hohe Warte	198
106948	Oed	400
20210	Klagenfurt	450
107300	Frankenfels	465
9010	Kufstein	490
11900	Jenbach	530
107037	Göstling an der Ybbs	530
105486	Almsee (Forsthaus)	590
15000	Mayrhofen	643
18501	Weitensfeld	704
105965	Schladming	730
102251	Ötz	760
102772	Kelchsau	815
15403	Rauris	934
105932	Untertauern	1000
102384	Schönberg im Stubaital	1005
1800	Weitra	1024
15710	Tamsweg	1026
14630	Umhausen	1041
15500	Bad Gastein	1092
11400	Holzgau	1100
102061	St.Leonhard i.P.	1335
15340	Kals	1352
113332	Innerkrems	1520
17000	Galtür	1577

station relocation and changes in the observing system for these time series are quite detailed.

Plausibility checks were done for the digitalized HZB raw data before 1970 in order to reject major errors in the time series. Moreover, most of the climate data records provided by the HZB network are quality proofed backwards to the beginning of the 1970s. Unfortunately, for these time series only information about station relocations are available.

## 2.3 Trentino - Alto Adige snow depth and temperature dataset

The Trentino - Alto Adige snow depth and temperature dataset will be analyzed in chapter 4 by mean of the wavelet transform, coherence wavelet and other statistical analysis. Goal of this analysis is the individuation of climate patterns as well as the comprehension of the correlation between snow related variables and other forcings such as temperature and climate indexes. Moreover, this dataset will be analyzed in chapter 5 in order to investigate the dependence structure between mean seasonal snow depth, snow cover duration, temperature and climatological indices.

### 2.3.1 Study area

The Adige catchment is one of the most important river basins in Italy, not only due to its large catchment area (12.100 km<sup>2</sup>) and length (409 km), but also for the presence of more than 60 hydropower plants with an unit effective power of at least 500 kW. A recent review of Chiogna et al. [2016] describes in details the hydrological conditions in the catchment and its chemical and ecological status. The most important tributaries of the Adige catchment are located in the Alpine part of the basin, and hence they are strongly influenced by snow dynamics. An increasing concern is rising due to the effects of climate change in the area [see e.g. Lutz et al., 2016, Gampe et al., 2016], since this has already shown important implications for water resources management, and above all, for hydropower production and for winter tourism. In terms of atmospheric circulation patterns, the Adige catchment is mainly affected by southwest weather patterns and lee cyclones [Xoplaki et al., 2004, Buzzi and Tibaldi, 1978]. The works of Brunetti et al. [2006] and Brunetti et al. [2009] show that the parts of the Alps where the Adige catchment is located does not display a statistically significant (larger than 90%) change in precipitation in the period 1865-2003, while the increase in temperature has been positive and statistically significant.

### 2.3.2 Description of the dataset

Snow depth time series of meteorological stations located in the Adige catchment are a relevant source of information to study

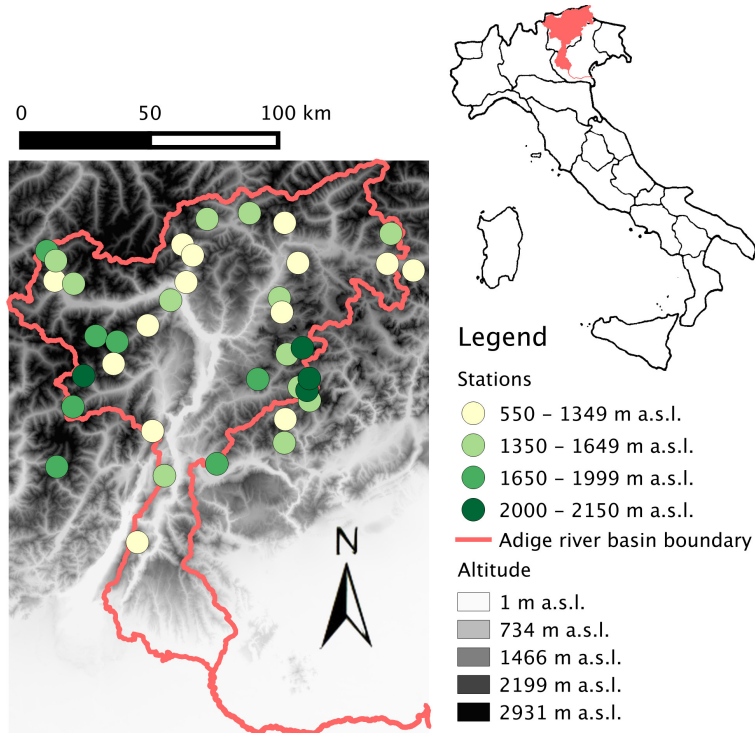


Figure 2.5: Location of the analyzed time series.

snow dynamics in the Alpine region (Figure 2.5) because of their spatial distribution over a wide elevation range and time spanning. In order to have a larger dataset we also considered five stations (Malga Bissina, Caoria, Brocon - Marande, San Martino di Castrozza and Sexten) which are very close to (yet not within) the catchment boundaries. This choice is justified by the high correlation ( $r > 0.9$ ) displayed by the time series at similar elevation in this region. A total of 37 stations are hence considered and we report their name, coordinates and elevation in Tables 2.2 - 2.5. The stations have been grouped into four elevation classes: below 1350 m a.s.l. (14 sites), between 1350 m and 1650 m a.s.l. (12 sites), between 1650 m and 2000 m a.s.l. (7 sites) and above 2000 m a.s.l. (4 sites). The reason for choosing these thresholds is the following. From a first analysis of the data we noticed a different behavior of the stations below and above 1650 m a.s.l. In order to better describe the variations depending on the different ele-

vations we then divided the two main classes into two subclasses. The two thresholds for the subclasses aimed at having a similar number of stations and a similar elevation variation between the four classes.

Table 2.2: Location of the stations below 1350 m a.s.l..

<b>Station</b>	<b>Altitude</b>	<b>Lat</b>	<b>Long</b>
St. Martin	588	46.783	11.228
Schenna	680	46.69	11.191
Caoria	875	46.199	11.672
Luesen	981	46.745	11.764
Andalo	1047	46.171	11.001
Zogglar Stausee	1142	46.542	10.99
Platt	1147	46.822	11.178
Pfunders	1159	46.887	11.703
St. Ulrich	1180	46.574	11.673
Toblach	1219	46.73	12.219
Marienberg	1310	46.706	10.521
Sexten	1310	46.703	12.35
Passo S. Valentino	1320	45.783	10.911
Rabbi	1323	46.410	10.812

Table 2.3: Location of the stations between 1350 m and 1650 m a.s.l..

<b>Station</b>	<b>Altitude</b>	<b>Lat</b>	<b>Long</b>
Flitzhof	1350	46.624	11.663
Ridnaun	1350	46.909	11.307
Wehr in Pfitsch	1365	46.925	11.525
Pozza di Fassa	1385	46.426	11.692
St. Magdalena	1398	46.835	12.243
Pawigl	1400	46.628	11.109
S. Martino di Castrozza	1462	46.262	11.798
Bondone Viote	1495	46.014	11.055
St. Valentin auf der Haide	1499	46.776	10.529
Paneveggio	1538	46.310	11.748
Matsch	1570	46.694	10.618
Brocon - Marande	1609	46.117	11.663

Each time series of the dataset refers to a site for the period

Table 2.4: Location of the stations between 1650 m and 2000 m a.s.l..

<b>Station</b>	<b>Altitude</b>	<b>Lat</b>	<b>Long</b>
Pampeago	1760	46.342	11.540
Panarotta	1775	46.050	11.320
Malga Bissina	1786	46.054	10.514
Ausserrojen	1833	46.81	10.484
Passo Tonale	1850	46.263	10.602
Zufrittsee	1851	46.509	10.725
Weissbrunnsee	1900	46.487	10.832

Table 2.5: Location of the stations above 2000 m a.s.l..

<b>Station</b>	<b>Altitude</b>	<b>Lat</b>	<b>Long</b>
Passo Rolle	2006	46.298	11.787
Pejo Tarlenta	2010	46.370	10.659
Passo Valles	2036	46.339	11.800
Ciampac	2145	46.447	11.771

going from 1st November to 30th April. When we refer to the season 1990, it means the data from 1st November 1990 to 30th April 1991 are considered. The time series regarding the Province of Trento are part also of the dataset described in section 2.1.

Verifying the homogeneity of the time series is an important prerequisite for detecting trends and changes in the time series [Auer et al., 2007, Brunetti et al., 2006]. To check for homogeneity of available snow depth time series, we applied the Standard Normal Homogeneity Test (SNHT) [Alexandersson and Moberg, 1997, Alexandersson, 1986, Marcolini et al., 2017] with the procedure described in section 3.2. This dataset, which is then used for the analysis in chapter 4, contains sites where the result of the homogeneity test showed that the mean seasonal snow depth time series, computed by averaging the daily snow depth between November 1 and April 30, are homogeneous in the timeframe 1980-2010.

Short gaps in the time series (i.e., shorter than 14 days), were filled by support vector machine regression [Smola and Schölkopf, 2004] applying the Matlab toolbox Spider (<http://www.kyb.tuebingen.mpg.de/bs/people/spider/> - Max Planck Institute for Biological Cybernetics), which uses the snow depth of the two

best correlated stations and, if available, the snowfall, temperature and precipitation data of the examined stations, as input variables for the regression.

Temperature data were also collected from the available databases of the meteorological survey of the Autonomous Province of Trento ([www.meteotrentino.it](http://www.meteotrentino.it)) and the Autonomous Province of Bolzano ([www.provincia.bz.it/meteo](http://www.provincia.bz.it/meteo)). The stations we considered in this work are summarized in Table 2.6. If possible, temperature stations located in the same sites as the snow stations were used. We then selected additional stations with the goal of creating a temperature dataset that could be associated with the elevation bands used to subdivide the snow depth time series.

Table 2.6: Location of the temperature stations.

<b>Station</b>	<b>Altitude</b>	<b>Lat</b>	<b>Long</b>
Bressanone Varna	590	46.730	11.644
Vallarsa (Diga di Speccheri)	875	45.768	11.135
Zogger Stausee	1142	46.542	10.990
Cogolo Pont (Centrale)	1190	46.365	10.689
Marienberg	1310	46.706	10.521
Passo Mendola	1360	46.419	11.197
Pawigl	1400	46.628	11.109
Predoi	1450	47.037	12.099
S.Valentin auf der Haide	1499	46.776	10.529
Redagno	1562	46.347	11.397
Vernago	1700	46.736	10.849
Pian Palu' (Diga)	1800	46.337	10.614
Zufrittsee	1851	46.509	10.725
Weissbrunnsee	1900	46.487	10.832
Passo Rolle	2012	46.298	11.787
Passo Valles	2032	46.338	11.800
Pian Fedaia (Diga)	2063	46.459	11.863
Cima Paganella	2125	46.143	11.037
Careser (Diga)	2600	46.423	10.699



## 2.4 Adige and upper Inn river basins discharge dataset

The discharge dataset of the Adige and upper Inn river basins will be analyzed in chapter 4 by means of the wavelet transform and wavelet coherence analysis.

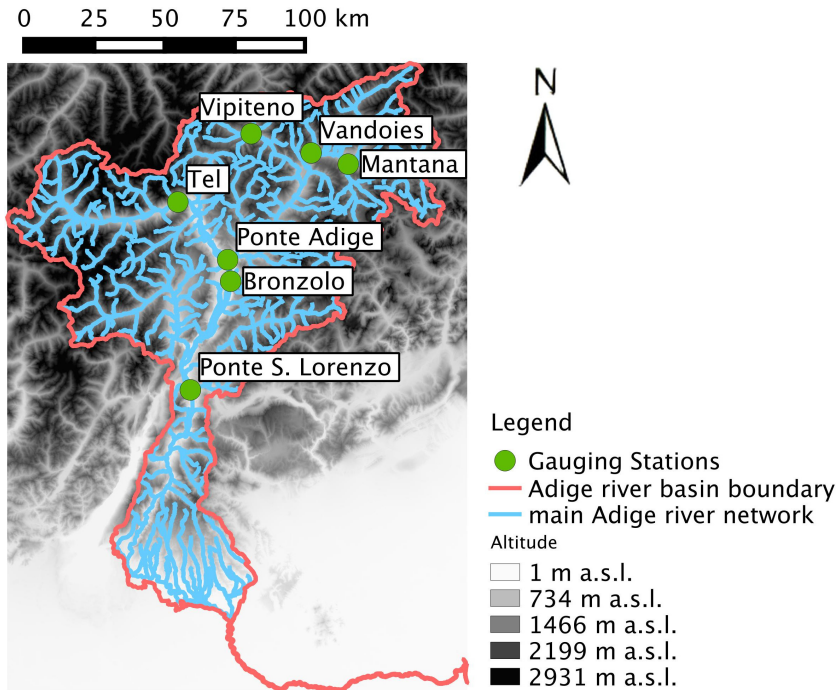


Figure 2.6: Location of the examined gauging stations in the Adige river basin.

The upper Inn and the Adige river basins have similar characteristics. The Adige river basin was already introduced in section 2.3.1. The area of the upper Inn catchment is similar to that of the Adige catchment, having a surface of 11960 km<sup>2</sup> closed in Wasserburg, which is the most downstream station we considered in this dataset. It crosses mainly three countries: Switzerland, Austria and Germany [Korck et al., 2012]. Both rivers have a typically alpine regime and are characterized by relatively humid and warm summers and falls, winter drought and late spring snow melt [see e.g. Zolezzi et al., 2009, Korck et al., 2012, Chiogna et al., 2016]. The two rivers present two peaks in discharge during the year:

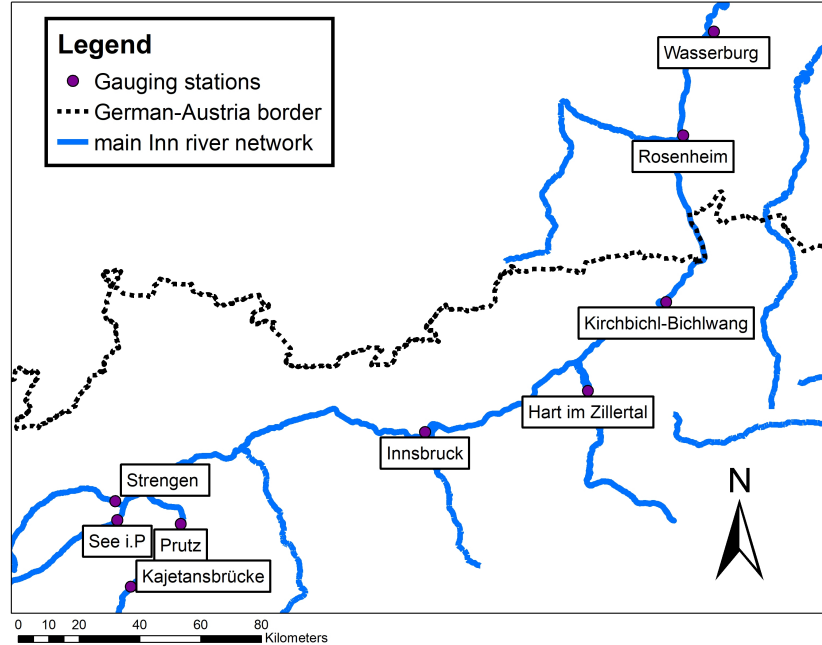


Figure 2.7: Location of the examined gauging stations in the Adige river basin (Sources: Bayerisches Landesamt für Umwelt [www.lfu.bayern.de](http://www.lfu.bayern.de), Geobasisdaten ©Bayerische Vermessungsverwaltung [www.geodaten.bayern.de](http://www.geodaten.bayern.de), offene Daten Österreichs [www.data.gv.at](http://www.data.gv.at)©SAGIS)

one in spring due to snow melt and one in autumn due to cyclonic storms.

The discharge data relative to the Adige river basin were provided by the University of Trento in the framework of the european project GLOBAQUA (<http://www.globaqua-project.eu/en/home/>). The discharge data of the Inn river basin were downloaded from the website of the *Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft, Ministerium für eine Lebenswertes Österreich* (<http://ehyd.gv.at/>) for the Austrian gauging stations, and from the website of the Bavarian Hydrological Service, Bavarian Environmental Agency (<http://www.gkd.bayern.de>) for the German ones.

Since we wanted to have the possibility to compare the results obtained for the two catchments, we decided to choose 16 gauging stations having comparable catchment areas. The location of the stations is shown in Figures 2.6 and 2.7. More details about the

locations of the stations are given in Table 2.7. The elevations and the catchment areas of the gauging stations are given in Table 2.8.

Table 2.7: Location of the gauging stations.

	<b>Station</b>	<b>River</b>	<b>Lat</b>	<b>Long</b>
Adige	Ponte S. Lorenzo	Adige	11.115	46.070
	Bronzolo	Adige	11.315	46.414
	Ponte Adige	Adige	11.304	46.483
	Vandoies	Rienza	11.708	46.816
	Tel	Adige	11.080	46.672
	Mantana	Gadera	11.879	46.776
	Vipiteno	Rio Ridanna	11.431	46.884
Inn	Wasserburg	Inn	12.234	48.059
	Rosenheim	Inn	12.144	47.854
	Kirchbichl-Bichlwang	Inn	12.094	47.523
	Innsbruck	Inn	11.381	47.264
	Prutz	Inn	10.659	47.079
	Kajetansbrücke	Inn	10.512	46.953
	Hart im Zillertal	Ziller	11.862	47.346
	See i.P	Trisanna	10.472	47.087
	Strengen	Rosanna	10.466	47.125

The longest time series of the dataset is the one of Adige - Ponte S. Lorenzo, which starts in 1923. Most of the other time series start between the end of the '40s and the beginning of the '50s. The shortest time series is the one of Rosanna - Strengen with 43 year of data.

Table 2.8: Characteristics of the gauging stations.

	<b>Station</b>	<b>Elevation</b> [m]	<b>Catchment</b> <b>Area [km<sup>2</sup>]</b>
Adige	Ponte S. Lorenzo	186	9800
	Bronzolo	227	6926
	Ponte Adige	237	2705
	Vandoies	733	1920
	Tel	506	1676
	Mantana	814	389
	Vipiteno	940	207
Inn	Wasserburg	420	11960
	Rosenheim	440	10154
	Kirchbichl-Bichlwang	484	9310
	Innsbruck	569	5772
	Prutz	862	2462
	Kajetansbrücke	970	2148
	Hart im Zillertal	531	1095
	See i.P	1018	385
	Strengen	970	271

## 2.5 Lower Inn river discharge dataset

For the lower part of the Inn river basin, we concentrated on the discharge data of two gauging stations: Wasserburg and Passau (see Figure 2.8 and Table 2.9). In chapter 5 we will model the flood events registered at these gauging stations by means of the copula. The data have been provided by the *Bayerische Landesamt für Umwelt* (Bavarian Environment Agency, Germany - LFU).

Table 2.9: Location and dimension of the drainage basin of the gauging stations.

<b>Gauging station</b>	<b>Altitude</b> [m]	<b>Lat</b>	<b>Long</b>	<b>Catchment Area [km<sup>2</sup>]</b>
Passau	289	48.561	13.444	26040
Wasserburg	420	48.059	12.234	11960

The time series of Passau covers the period 1920-2014, while the one of Wasserburg that was made available by the LFU starts in 1964 and ends in 2014. The original time series contain data

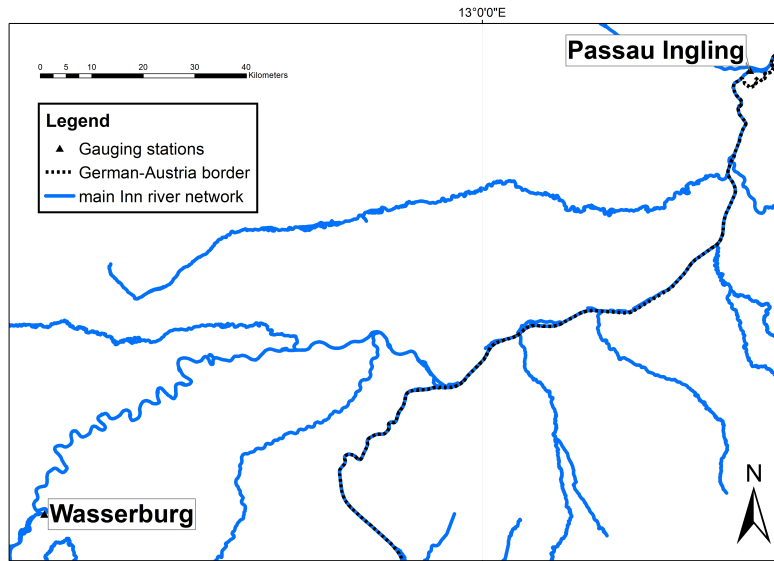


Figure 2.8: Map of the Inn River Basin with the location of the analyzed gauging stations. (Sources: Bayerisches Landesamt für Umwelt [www.lfu.bayern.de](http://www.lfu.bayern.de), Geobasisdaten ©Bayerische Vermessungsverwaltung [www.geodaten.bayern.de](http://www.geodaten.bayern.de), offene Daten Österreichs [www.data.gv.at](http://www.data.gv.at), ©SAGIS )

with different resolution: daily data (up to 31.10.1969), discrete data (up to 31.12.2006) and data collected every 15 minutes. In the discrete data time series, the discharge values have only been saved in those time steps, where a strong change in the discharge has been observed. This brings to a variable number of data per each day.

In order to have a resolution of 15 minutes for all time series, the time series were preprocessed as it follows. The daily time series have been disaggregated on a 15 minutes basis, modeling the discharge of each single day with a third degree polynomial, as suggested by Wagner [2012]. The four conditions for the estimation of the coefficients of the polynomial are the starting value and the conservation of the volume in the modeled day and in the two following days. A critical point of this approach is that the obtained line sometimes shows some fluctuations, which, in case of low streamflow conditions, can introduce negative values. Nevertheless, since we are mainly interested in flood events, this does not have a strong effect on our analysis. The discrete data

have been linearly interpolated on a 15 minutes basis, as suggested by the *Bayerische Landesamt für Umwelt* (Bavarian Environment Agency - Germany). Linear interpolation has also been applied in order to fill the gaps of the data collected every 15 minutes.

The reliability of the time series has been tested investigating the presence of inhomogeneities, trend and change points in the time series of the yearly maximum values. These tests were performed with the statistical software HyStat [Willems, 2013], Version 4.0.8. The homogeneity tests that were performed are the following:

- Outliertest (OUTS)
- Cox-Stuart-Test, Location (COXLO)
- Cox-Stuart-Test, Dispersion (COXDI)
- Noether-Test (NOETH)
- Kolmogoroff-Smirnoff-Test (KS)
- Chi2-Test (CHI)
- Kruskal-Wallis-Test, 2 Samples (KW2SP)
- Kruskal-Wallis-Test, 3 Samples (KW3SP)
- t-Test (TTST)
- F-test (FTST)

For the detection of the trends we applied the Mann-Kendall Test (MK), the Trend to Noise Ratio (TR) and the t-Test for the slope coefficient (TTST) in combination with the least square method (LS), the least absolute error method (LAE) and the 3-group-resistant line (RES).

The Bernier-, Pettit- and Local-Jump-Methods were applied for the change points detection.

For further explanations about these tests we refer to the manual of the HyStat Software [Willems, 2013]. The two analyzed time series passed most of the tests and were as a consequence judged as reliable for further analysis.

## Chapter 3

# Homogenization

### 3.1 Introduction

A prerequisite to investigate possible climatic changes and trends occurring in a region is to check the homogeneity of the available time series. It is important, indeed, to identify changes driven by non-climatic factors, which could have affected some of the time series composing the dataset. Among them are relocations of the station, substitutions of the equipment, changes in the operator collecting the measurements or modifications in the measuring procedure [see e.g., Hanssen-Bauer and Førland, 1994, Brunetti et al., 2006, Aguilar et al., 2003]. The homogeneity of the dataset can be checked considering only metadata, from the Greek *μετὰ* "beyond" and the Latin *datum* "information, data", in other words "data about data", [see e.g. Scherrer et al., 2013], or by applying statistical tests [see e.g., Aguilar et al., 2003], or better combining these two types of information, such as to confirm with metadata breakpoints identified by statistical tests [see e.g., Williams et al., 2012, Domonkos and Štěpánek, 2009, Domonkos, 2013]. Several statistical tests have been developed to homogenize mainly precipitation and temperature time series. Examples of such tests are PRODIGE [Caussinus and Mestre, 2004], the two-phase regression method [Easterling and Peterson, 1995], HOMER [Mestre et al., 2013], and the multiple linear regression method [Vincent, 1998]. For a comparison of these tests see for example Vincent [1998], Ducré-Robitaille et al. [2003], Reeves et al. [2007]. To the best of our knowledge, homogenization of snow depth time series with the aid of statistical tests has been rarely attempted [Koch et al.,

2014].

In the first section of this chapter, we verify whether the Standard Normal Homogeneity Test (SNHT) [Alexandersson, 1986, Alexandersson and Moberg, 1997] is suitable for the homogenization of snow depth time series. This method has been developed for testing time series of precipitation and air temperature and has been shown to perform well, both in detecting time series affected by breakpoints as well as in identifying time series which are homogeneous [Ducré-Robitaille et al., 2003]. Its effectiveness for snow depth time series should be checked, given the different characteristics of these time series with respect to precipitation and temperature, such as the probability distribution, measurement time steps and accuracy. Snow depth time series often contain gaps distributed over several years. We apply the SNHT for the detection of inhomogeneities of the time series of the dataset described in section 2.1. Our main contribution, besides verifying the applicability of SNHT to snow depth time series, is the attention given to the creation of a well representative reference time series. The latter is a challenging point to be faced, especially for the characteristics of snow data (e.g. difficulties in measuring snow depth). Identifying a reliable reference time series is of fundamental importance to avoid the identification of false breakpoints or hiding real ones [Hanssen-Bauer and Førland, 1994, Menne and Williams Jr, 2005]. For these reasons, in our homogeneity test we create two reference time series using the approaches proposed by Alexandersson and Moberg [1997] and Peterson and Easterling [1994], respectively.

As mentioned before, there are several methods that were developed for the homogenization. In the second section of this chapter we present an intercomparison experiment between the SNHT and HOMOP, an algorithm that couples the performance of PRODIGE for the detection of the breakpoints and of INTERP for their correction. This analysis has been performed with the group of the Central Institute of Meteorology and Geodynamic of Vienna, which also kindly made the needed data available.



## 3.2 The Standard Normal Homogeneity Test for the homogenization of mean seasonal snow depth data

### 3.2.1 The algorithm

Let us consider a dataset composed of  $\tilde{k}+1$  time series, that we will indicate with  $\mathbf{S}_j$  where  $j = 1, \dots, \tilde{k}+1$ . We will denote each element of the time series  $\mathbf{S}_j$  with  $\mathbf{S}_{ji}$  with  $i = 1, \dots, m$ , where  $m$  indicates the length of the time series  $\mathbf{S}_j$ . The goal of the homogenization is to detect, and where possible correct, inhomogeneities in the tested time series, i.e., modifications of the behavior of a time series that are not due to natural variations or climatic changes. In doing that it is implicitly assumed that the time series of the same climatic area are influenced by the same natural variations. In this sense, we are talking about the so called relative homogeneity. Each time series will be tested for homogeneity against the reference time series by using the SNHT [Alexandersson, 1986, Alexandersson and Moberg, 1997]. Accordingly, all the  $\tilde{k} + 1$  time series will in turn play the role of the tested time series, which will be indicated as  $\mathbf{Y}$ . The reference time series  $\mathbf{X}$  will be constructed on the basis of  $k$  reference stations, chosen from the  $\tilde{k}$  remaining time series of the dataset. Similarly to the notation for  $\mathbf{S}_j$ , the single elements of the tested time series  $\mathbf{Y}$  and of the reference time series  $\mathbf{X}$  are denoted with  $\mathbf{Y}_i$  and  $\mathbf{X}_i$  respectively.

As shown in Fig. 3.1, the steps of the homogeneity test are the following ones: the choice of the time series  $\{\mathbf{S}_1, \dots, \mathbf{S}_k\}$  (Step 1), which will be used for the construction of the reference time series  $\mathbf{X}$  (Step 2), the detection of the breakpoints (Step 3), their classification (Step 4) and the computation of a correction factor to correct the time series (Step 5).

Notice that the SNHT is more effective in detecting the breakpoints when the time series are composed by mean values computed over one year as described in Aguilar et al. [2003]. Furthermore, as shown by Toreti et al. [2011] the probability that the SNHT detects false breakpoints or, on the contrary, it misses real breakpoints, is higher at the beginning and at the end of the time series. In our case breakpoints located in the first and last 5 years of the tested time series are not corrected [see e.g. Alexandersson and Moberg, 1997, Hanssen-Bauer and Førland, 1994]. In addition, only time series longer than 10 years are tested.

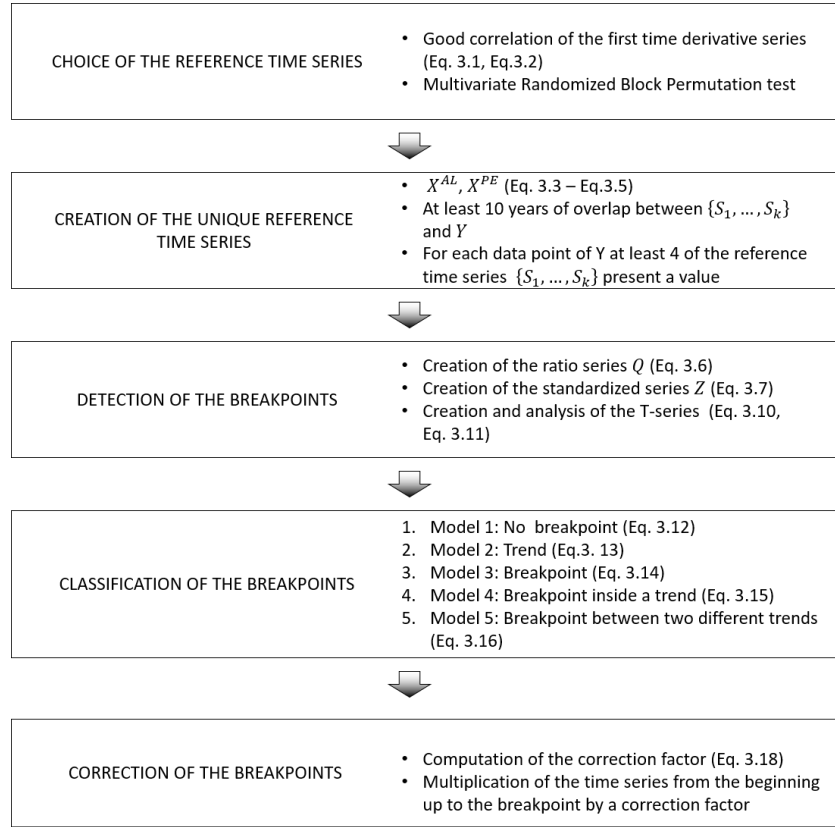


Figure 3.1: Scheme of the homogeneity test for the detection of the breakpoints.

### Step 1: Selection of the $k$ time series $\mathbf{S}$ for the construction of the reference time series $\mathbf{X}$

The first step for the detection of the breakpoints in the tested time series  $\mathbf{Y}$  is the choice of the  $k \leq \tilde{k}$  reference stations that will form the reference time series  $\mathbf{X}$ .

According to Peterson and Easterling [1994], in order to reduce the effect of possible breakpoints of the reference stations on the homogeneity analysis of the tested time series  $\mathbf{Y}$ , the weights  $\rho_j$  for the creation of the reference time series are not computed as the correlation coefficient between each times series  $\mathbf{S}_j$  and  $\mathbf{Y}$ , but as the correlation coefficient between their respective increments.

Let  $n$  be the length of the tested time series  $\mathbf{Y}$ , let  $i = 1, \dots, n$  be the index of the time step and let  $j = 1, \dots, k$  identify one of the

$k$  reference stations selected to construct the reference time series. Following Easterling and Peterson [1995], we require that

$$\text{corr}(\Delta \mathbf{S}_j, \Delta \mathbf{Y}) \geq 0.7 \quad (3.1)$$

$$\text{corr}(\Delta \mathbf{X}, \Delta \mathbf{Y}) \geq 0.8 \quad (3.2)$$

where  $\Delta X_i = X_{i+1} - X_i$ ,  $\Delta S_{j,i} = S_{j,i+1} - S_{j,i}$  and  $\Delta Y_i = Y_{i+1} - Y_i$  for  $i = 1, \dots, n - 1$ .

As suggested by Peterson and Easterling [1994], the significance of the correlation coefficient between two time series is tested applying the Multivariate Randomized Block Permutation (MRBP) test (see for example Mielke Jr [1986] or Mielke Jr [1991]). We require that the probability that a random permutation of the elements of the array of the increments of the reference time series  $\frac{\Delta \mathbf{S}_j}{\Delta t}$  can give a better prediction of  $\frac{\Delta \mathbf{Y}}{\Delta t}$  with respect to the array of increments of the reference time series  $\frac{\Delta \mathbf{S}_j}{\Delta t}$  itself is less than 1%.

We also require that all the reference stations  $\{\mathbf{S}_1, \dots, \mathbf{S}_k\}$  as well as the tested one  $\mathbf{Y}$  have a common overlapping period  $P$  of at least 10 years and that at each time step of  $\mathbf{Y}$  at least 4 of the reference stations  $\{\mathbf{S}_1, \dots, \mathbf{S}_k\}$  show a value [Moberg and Alexandersson, 1997].

With these requirements, the homogeneity test is performed where  $\mathbf{X}$  can be defined.

## Step 2: Creation of the reference time series

We compute the reference time series using two different approaches. This allows us to minimize the probability that the reference time series contains a breakpoint that could negatively influence the homogenization analysis.

In the first approach to construct the reference series we follow Peterson and Easterling [1994]:

$$\mathbf{X}_i^{PE} = \mathbf{X}_{i+1}^{PE} - \Delta \mathbf{X}_i^{PE} \quad \text{for } i = n - 1, \dots, 1. \quad (3.3)$$

with  $\mathbf{X}_n^{PE} = \mathbf{Y}_n$  as last value (the construction proceeds backward) of the time series and the increments time series computed as follows:

$$\Delta \mathbf{X}_i^{PE} = \frac{\sum_{j=1}^k \rho_j^2 \Delta S_{ji}}{\sum_{j=1}^k \rho_j^2}. \quad (3.4)$$

The second approach we use is that proposed by Alexandersson [1986], according to which the reference time series  $\mathbf{X}^{AL}$  is defined as follows:

$$\mathbf{X}_i^{AL} = \frac{\sum_{j=1}^k \rho_j^2 \mathbf{S}_{ji} \bar{\mathbf{Y}} / \bar{\mathbf{S}}_j}{\sum_{j=1}^k \rho_j^2}, \quad (3.5)$$

where the bar indicates the mean values in the common overlapping period  $P$  between all time series selected in Step 1 [Alexandersson and Moberg, 1997].

Notice that in both cases the weights  $\rho_j$  are the correlation coefficients between the increments time series of  $\mathbf{S}_j$  and that of the tested time series  $\mathbf{Y}$  in the common overlapping period  $P$  between all time series [Alexandersson and Moberg, 1997].

### Step 3: Detection of the breakpoints

Once the reference time series  $\mathbf{X}$  is created, the ratio series is defined:

$$\mathbf{Q}_i = \frac{\mathbf{Y}_i}{\mathbf{X}_i}, \quad (3.6)$$

which is then standardized:

$$\mathbf{Z}_i = \frac{(\mathbf{Q}_i - \bar{\mathbf{Q}})}{\sigma_{\mathbf{Q}}}, \quad (3.7)$$

where  $\bar{\mathbf{Q}}$  and  $\sigma_{\mathbf{Q}}$  are the mean and the standard deviation of the time series  $\mathbf{Q}$ , respectively.

The null hypothesis  $H_0$  that  $\mathbf{Y}$  is homogeneous can be formulated as follows

$$H_0 : \mathbf{Z}_i \in N(0, 1) \quad i \in \{1, \dots, n\}, \quad (3.8)$$

where  $N(\mu, \sigma^2)$  denotes the normal distribution with mean  $\mu$  and standard deviation  $\sigma$ . Therefore, the time series  $\mathbf{Y}$  is homogeneous with respect to the reference time series  $\mathbf{X}$ , if their ratio is a white noise [Easterling et al., 1996]. The alternative hypothesis  $H_1$  that  $\mathbf{Y}$  has a breakpoint at the time step  $a$  is formulated as follows

$$H_1 : \begin{cases} \mathbf{Z}_i \in N(\mu_1, 1) & i \in \{1, \dots, a\} \\ \mathbf{Z}_i \in N(\mu_2, 1) & i \in \{a + 1, \dots, n\} \end{cases} . \quad (3.9)$$

After these preparatory steps, the following times series, hereafter called T-series, is computed

$$T(a) = a\bar{z}_1^2 + (n - a)\bar{z}_2^2, \quad (3.10)$$

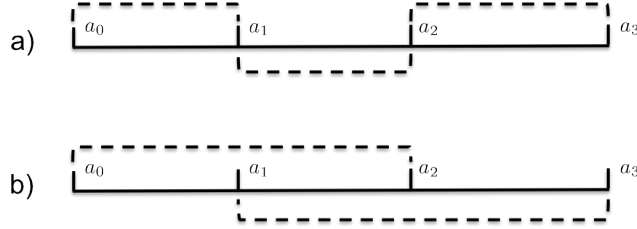


Figure 3.2: In the splitting phase of the semihierarchical splitting and merging algorithm each single segment between two break-points is tested (a), while in the merging phase (b) we analyze each couple of subsequent segments (containing only one break-point).

where  $\bar{z}_1$  and  $\bar{z}_2$  are the averages of the time series  $\mathbf{Z}$  respectively before and after the shift  $a$ . The  $\mathbf{T}$ -series is then used to compute the following test statistic:

$$T_{max} = \max_{1 \leq a \leq n-1} \{\mathbf{T}(a)\}, \quad (3.11)$$

in order to determine which of the two hypothesis is the most probable. The choice of this particular test quantity  $T_{max}$  is detailed in Alexandersson [1986]. The general idea behind this choice is to find the point  $a$ , where the two averages  $\bar{z}_1$  and  $\bar{z}_2$  depart most largely from 0, which makes the homogeneity hypothesis  $H_0$  unlikely [Hanssen-Bauer and Førlund, 1994].

If  $T_{max}$  is above a critical level  $L(n, \alpha)$ , where  $n$  is the length of the time series  $\mathbf{Z}$  and  $\alpha$  is the chosen confidence level (typically  $\alpha = 95\%$ ) [Khaliq and Ouarda, 2007], the position  $a$  corresponding to  $T_{max}$  is considered a breakpoint and the null hypothesis  $H_0$  (homogeneity) is rejected.

**Multiple breakpoints** In the tested time series there might be multiple breakpoints. In order to maximize the probability of detecting multiple breakpoints, we first apply the SNHT algorithm to the whole time series  $\mathbf{Y}$ . If the test detects a breakpoint, we proceed iteratively with the so-called semihierarchical splitting and merging algorithm [Menne and Williams Jr, 2005, 2009]. In each iteration two checks will be performed.

Let  $\mathbf{A} = \{a_0 = 0, a_1, \dots, a_{p-1}, a_p = n\}$  be the set of all detected breakpoints, then we proceed as follows:

**SPLITTING** We apply the SNHT algorithm to each segment  $a_{i-1} + 1 : a_i$  for  $i = 1 : p$  (see Fig. 3.2 a). If a breakpoint is detected, it is added to the breakpoint set  $\mathbf{A}$ .

**MERGING** For each breakpoint  $a_i$  with  $i = 1 : p - 1$  the SNHT algorithm is applied to the segment  $a_{i-1} + 1 : a_{i+1}$  (see Figure 3.2 b). If no breakpoint is detected, the breakpoint  $a_i$  is deleted from the breakpoint set.

The procedure stops when no breakpoint is added or deleted from the breakpoint array  $\mathbf{A}$  or when the segments are too small to be divided because shorter than 10 years [see e.g., Moberg and Alexandersson, 1997, Toreti et al., 2011]. This choice represents a compromise between the minimum distance of the detectable breakpoints and the performance of the test. Indeed, analyzing periods shorter than 10 years, would allow us to detect breakpoints that are closer in time, but would also increase the probability of false breakpoint detections.

#### Step 4: Classification of the breakpoints

Once a breakpoint is detected, it is important to verify whether the associated inhomogeneity is in the form of a trend or a shift and, in the latter case, if it takes place between two trend periods. This check is performed following the procedure suggested by Menne and Williams Jr [2009] and summarized here. Each segment  $a_{i-1} + 1 : a_{i+1}$ ,  $i = 1 : p - 1$  of the time series  $\mathbf{Q}$  is fitted with one of the following models:

where  $\epsilon_i$  represents a random error term.

Differently from Menne and Williams Jr [2009], we choose as representative model of the type of inhomogeneity the one that minimizes the reduced chi-square index

$$\chi_{red}^2 = \frac{SSE}{m - pn}, \quad (3.17)$$

where  $SSE$  indicates the sum of the squared errors,  $m$  is the length of the time series  $\mathbf{Q}$  and  $pn$  is the number of parameters of the model. We decided to minimize the reduced chi-square index instead of the Bayesian information criterion, used by Menne and Williams Jr [2009], because in our case the latter seemed to give too much weight to the number of parameters and so tended to classify the inhomogeneity according to Model 1 and 2 even in presence of actual shifts.

**Model 1** no breakpoint

$$\mathbf{Q}_i = \mu + \epsilon_i \quad (3.12)$$

**Model 2** trend

$$\mathbf{Q}_i = \mu + \beta i + \epsilon_i \quad (3.13)$$

**Model 3** shift

$$\mathbf{Q}_i = \begin{cases} \mu_1 + \epsilon_i & i \leq a \\ \mu_2 + \epsilon_i & i > a \end{cases} \quad (3.14)$$

**Model 4** shift inside a trend

$$\mathbf{Q}_i = \begin{cases} \mu_1 + \beta i + \epsilon_i & i \leq a \\ \mu_2 + \beta i + \epsilon_i & i > a \end{cases} \quad (3.15)$$

**Model 5** shift between two different trends

$$\mathbf{Q}_i = \begin{cases} \mu_1 + \beta_1 i + \epsilon_i & i \leq a \\ \mu_2 + \beta_2 i + \epsilon_i & i > a \end{cases} \quad (3.16)$$

### Step 5: Correction of the inhomogeneities

Once an inhomogeneity is detected,  $\mathbf{Y}$  can be corrected applying a correction factor. We emphasize here that the use of homogenized time series for climatological analysis is still matter of debate [e.g. Beniston et al., 2003, Marty, 2008]. Another argument that should be considered is that depending on the goal of the analysis to be done on the homogenized dataset, one can choose to take into account the homogenized time series or the original ones [Peterson et al., 1998], so that it is important to keep both of them in separate datasets. In some cases, it could also be advisable to divide the original time series into distinct time series where the breakpoints occur.

The correction is applied as follows. Let us suppose that the time series  $\mathbf{Y}$  displays a breakpoint at the position  $a$  of the time series. We will maintain unchanged the most recent part (from

$a + 1$  to  $n$ ) of the tested time series  $\mathbf{Y}$  [Aguilar et al., 2003] and multiply the older part (from the beginning of the time series up to breakpoint  $a$ ) by the following correction factor:

$$cf = \frac{\bar{q}_2}{\bar{q}_1}, \quad (3.18)$$

with  $\bar{q}_1$  and  $\bar{q}_2$  defined as follows:

$$\bar{q}_1 = \sigma_{\mathbf{Q}}\bar{z}_1 + \bar{\mathbf{Q}} \quad (3.19)$$

$$\bar{q}_2 = \sigma_{\mathbf{Q}}\bar{z}_2 + \bar{\mathbf{Q}} \quad (3.20)$$

where again  $\bar{z}_1$  and  $\bar{z}_2$  are the average of the time series  $\mathbf{Z}$  before and after the shift  $a$  respectively, and  $\sigma_{\mathbf{Q}}$  and  $\bar{\mathbf{Q}}$  are the standard deviation and the mean of the time series  $\mathbf{Q}$  [Alexandersson and Moberg, 1997].

### 3.2.2 Testing procedure of the algorithm

Figure 3.3 shows the general structure of the homogenization procedure, whose application is performed in two phases: breakpoint detection and breakpoint confirmation.

#### Breakpoint detection

In the breakpoint detection phase, only the first three steps described in section 3.2.1 are performed, as also illustrated in Figure 3.3.

The time series  $\mathbf{Y}$  is tested against the two reference time series  $\mathbf{X}^{AL}$  and  $\mathbf{X}^{PE}$  formed with the reliable time series  $\{\mathbf{S}_1, \dots, \mathbf{S}_k\}$ , as described in section 3.2.1.

In the first phase of the algorithm, we keep all the breakpoints identified by comparing  $\mathbf{Y}$  with the two reference time series  $\mathbf{X}^{AL}$  and  $\mathbf{X}^{PE}$ , which are collected in two distinct sets indicated in Figure 3.3 as  $BP^{AL}$  and  $BP^{PE}$ , respectively.

In order to test the reliability of the identified breakpoints, the homogeneity test is applied several times considering a different minimum overlapping percentage  $MOP$  (e.g. 30%, 50%, 70%) between each single reference station's time series  $\mathbf{S}_j$ , used to construct  $\mathbf{X}^{AL}$  and  $\mathbf{X}^{PE}$ , and the tested time series  $\mathbf{Y}$ .



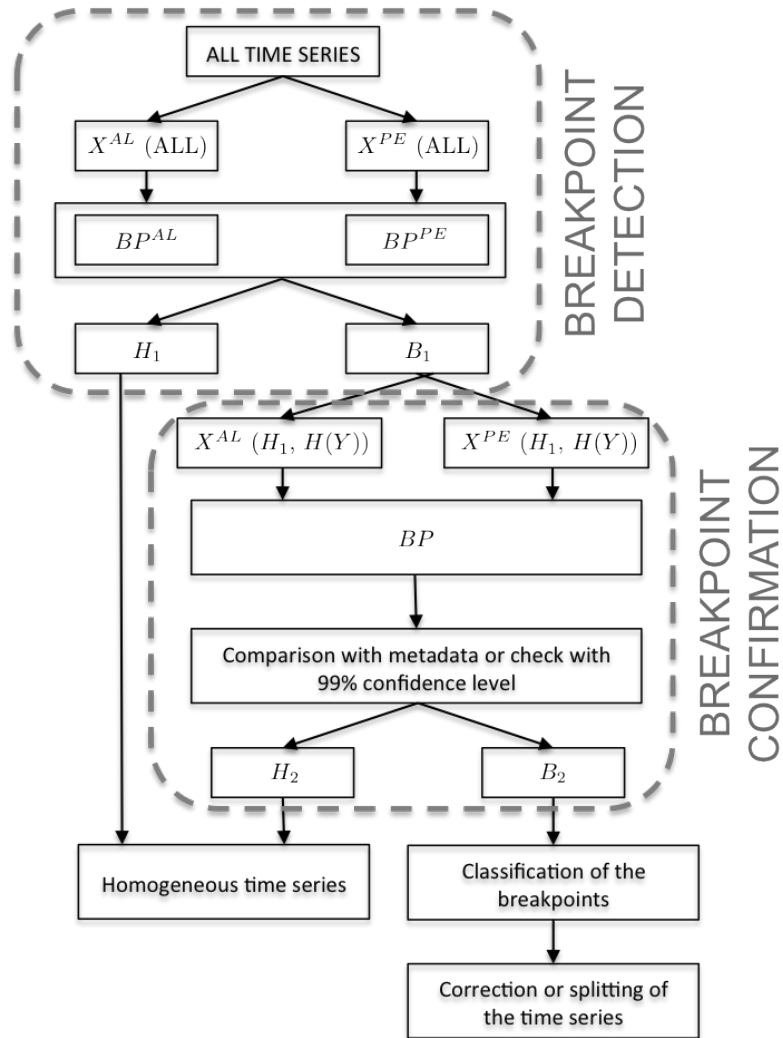


Figure 3.3: General scheme of the testing procedure.  $X^{AL}$  and  $X^{PE}$  indicate the reference time series according to [Alexandersson and Moberg, 1997] and Peterson and Easterling [1994], respectively.  $BP$  indicates the breakpoint set.  $H$ ,  $B$  and  $H(Y)$  indicate the dataset of the homogenous, inhomogenous and locally homogenous time series, respectively, as defined in section 3.2.2.

Table 3.1: Scheme of the procedure for the application of the algorithm for the detection of the breakpoints.

phase	tested	candidate	reference	output
		reference stations	time series	
breakpoint detection	ALL	ALL	$\mathbf{X}^{AL}, \mathbf{X}^{PE}$	$H1$ $B1$
breakpoint confirmation	$B1$	$H1$ $H(\mathbf{Y})$	$\mathbf{X}^{AL}, \mathbf{X}^{PE}$	$H1, H2$ $B2$

### Breakpoint confirmation

In the first phase of the algorithm all the time series of the dataset are tested in turn, and each reliable time series of the dataset is taken into account in order to create the reference time series  $\mathbf{X}^{AL}$  and  $\mathbf{X}^{PE}$ . The output will be two sets of time series: the homogeneous ones (dataset  $H1$ ) and those, contained in dataset  $B1$ , that have shown to have a breakpoint during the comparison with one or both the reference time series  $\mathbf{X}^{AL}$  and  $\mathbf{X}^{PE}$  (see Table 3.1).

In the second phase, we test only the time series containing potential breakpoints, which are stored in the dataset  $B1$  (see Figure 3.3). The constraints for the choice of the reliable time series for the creation of the reference time series are more stringent. The ideal case would be that the reference stations' time series were all homogeneous. It is difficult to meet this condition, because there is the possibility that only a few, or even none, of the time series contained in the dataset  $H1$  fulfill the requirements illustrated in section 3.2.1. For this reason in this second phase we require that a reference station is homogeneous (dataset  $H1$ ) or at least that it is locally homogeneous in an interval of 10 years centered on the breakpoints detected in the first phase in the tested time series  $\mathbf{Y}$  (dataset  $H(\mathbf{Y})$ ). Figure 3.4 shows an example of the acceptable candidate reference stations for the second phase.

### Post-processing

An important last step in the algorithm is the comparison of our results with the metadata, in order to identify which breakpoints correspond to documented changes that could have affected the

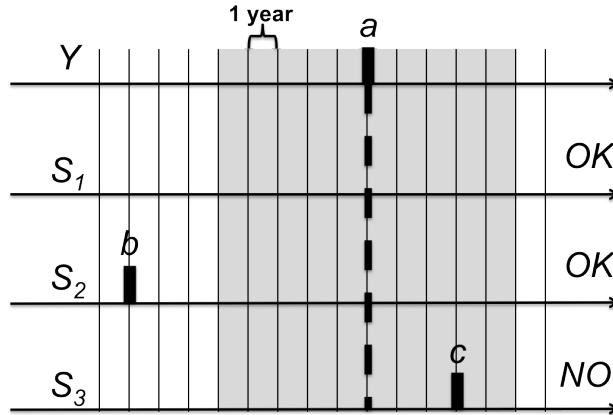


Figure 3.4: Example of candidate reference time series for the second phase of the homogenization analysis.  $\mathbf{S}_1$  is homogeneous and is part of the set  $H1$ .  $\mathbf{S}_2$  presents a breakpoint  $b$  8 years before the breakpoint  $a$  and is therefore contained in the set  $H(\mathbf{Y})$ .  $\mathbf{S}_3$  has shown to have a breakpoint  $c$  only 3 years after the breakpoint  $a$ . For this reason, neither  $H(\mathbf{Y})$  nor  $H1$  contain  $\mathbf{S}_3$ . In this example,  $\mathbf{S}_1$  and  $\mathbf{S}_2$  will be taken into account for the choice of the reference station for the creation of the reference time series for  $\mathbf{Y}$ , but  $\mathbf{S}_3$  will not be considered.

time series [Peterson et al., 1998]. Metadata are all the documents concerning the history of the station (e.g. relocations, change in the operator, measurements procedure, substitution of the equipment, surrounding environment) and any other information about the factors that could have influenced the corresponding time series. Unfortunately, this documentation is not always complete, and sometimes it can also be inaccurate or contradictory [Menne and Williams Jr, 2005, Domonkos and Štěpánek, 2009].

A breakpoint is confirmed if identified by both the reference time series  $\mathbf{X}^{AL}$  and  $\mathbf{X}^{PE}$  with a confidence level of 0.95 or if it is identified by either  $\mathbf{X}^{AL}$  or  $\mathbf{X}^{PE}$  considering in addition either a confidence level of 0.99 or with the support of the metadata.

As illustrated in Figure 3.3 and in Table 3.1, at the end of the second phase, we obtain three datasets: the dataset  $H1$  formed by the time series classified as homogeneous in the first phase, the dataset  $H2$  containing the time series classified as homogeneous in the second phase and the dataset  $B2$  of the time series containing

at least one breakpoint (detected using both  $\mathbf{X}^{AL}$  and  $\mathbf{X}^{PE}$ ) and satisfying the criteria presented before. The time series contained in the dataset  $B2$  can then be homogenized applying the correction factor defined in Eq. (3.18) as illustrated in section 3.2.1. Only the time series showing a shift or a shift between trends are considered for corrections in this work. Depending on the type of analysis that should be performed, each operator will then choose which time series should be used [Peterson et al., 1998]. For example, one could choose to use only the homogeneous time series from subset  $H1$  and  $H2$ , without taking into account the homogenized time series, which were contained in dataset  $B2$ , or the operator could also decide to involve in the analysis these latter time series too. It is worth noting, that since the correction factor is a multiplicative factor, it can only modify existing values, also, it can not create new days with positive snow depth. For this reason, the application of this multiplicative factor provides reasonable results only for the mean seasonal snow depth time series.

### 3.3 Application of the procedure to the Trentino snow depth dataset

In this section we present the application of the algorithm introduced in section 3.2 to the mean seasonal snow depth dataset relative to the Province of Trento, which was described in section 2.1.

#### 3.3.1 Detected breakpoints

We analyzed only the 81 time series longer than 10 years [Marcolini et al., 2017]. The application of the first phase of the procedure as described in section 3.2.2 identified 40 time series with at least one breakpoint. Among them, the second phase reclassified as homogeneous 22 time series and in 5 cases it was not possible to analyze the homogeneity of the time series in the second phase because of lack of appropriate reference stations. At least one breakpoint was detected in 13 time series (Table 3.2). Table 3.3 summarizes the temporal location, the cause of the breakpoint, the available metadata and the magnitude of the correction factor. At site 24 the correction factor was larger than 4, so it was more advisable to consider two separate time series instead of correcting

Table 3.2: Site number, name, altitude (expressed in m a.s.l.), coordinates (WGS 84) and available measurement period (starting and ending year) for the inhomogeneous time series.

Site	Name	Altitude	Long	Lat	Period
5	Andalo	1047	11.0010	46.1715	1981-2012
7	Predazzo Gardonè	1675	11.5758	46.3403	1981-2012
12	Canal San Bovo - Calaita	1318	11.7732	46.1837	1975-2012
13	Vallarsa Pian Fugazze	1170	11.1671	45.7431	1981-2012
18	Madonna di Campiglio	2015	10.8015	46.2273	1983-2012
19	Pancugolo Val Noana	1027	11.8392	46.1388	1965-2012
23	Rumo	1100	11.0085	46.4539	1990-2012
24	Folgarida	1890	10.8467	46.2902	1975-2012
63	Pinzolo Doss del Sabion	1899	10.8127	46.1723	1981-2012
631	Fondo Val Cadino	987	11.1380	46.4396	1953-1999
649	(Segheria Canton)	964	11.4145	46.2457	1949-1985
675	Ala (Ronchi) Daone	692	11.0654	45.7392	1974-2003
679	(Diga di Malga Boazzo)	1200	10.5238	45.9960	1959-2001

it.

Sites 5, 7, 23 and 63 show a breakpoint close to the edges of the time series (less than 5 years), and for this reason the corresponding breakpoints have not been corrected. Notice that, due to a smaller availability of reference stations, the analysis of the site 631 started in 1974, and for this reason its breakpoint lays on the edge of the reference time series and has not been corrected. Sites 13, 649 and 675 have been corrected even in absence of meta-data because their breakpoints were confirmed by the comparisons

Table 3.3: Results obtained with the two reference time series  $\mathbf{X}^{AL}$  and  $\mathbf{X}^{PE}$  for those time series that resulted non-homogeneous both in the first and second phase. When "(Edge)" is indicated, it means that a breakpoint has not been corrected because its distance to the edge of the time series is less than 5 years. The abbreviations "sr" and "df" mean station relocation and different sources, respectively. The correction factor is indicated between brackets next to the breakpoint location in the column "Corrected".

Site	$\mathbf{X}^{AL}$	$\mathbf{X}^{PE}$	Documented	Corrected
5	2009/10	2009/10	2009/10 (sr)	NO (Edge)
	1985/86	1985/86	1985/86(sr)	1985/86 (0.68)
7	1994/95	1994/95	1994/95 (sr)	1994/95 (2.18)
	2010/11	2010/11		NO (Edge)
12	2004/05	2002/03	2002/03 (sr)	2002/03 (2.01)
13	1997/98	2000/01	-	2000/01 (1.53)
18	1996/97	1996/97	1997/98 (sr)	1997/98 (1.29)
19	1987/88	1981/82	1983/84 (ds)	1983/84 (0.68)
23	1992/93	1992/93	-	NO (Edge)
24	1999/2000	1999/2000	2002/2003 (ds)	2002/2003 (Divided)
63	1989/90	1988/89	1987/88 (sr),	1987/88 (0.65)
	2007/08	2007/08	2007/08 (ds)	NO (Edge)
631	1974/75	1975/76	-	NO (Edge)
649	1962/63	1962/63	-	1962/63 (0.56)
675	1983/84	1984/85	-	1984/85 (1.34)
679	1992/93	-	-	1992/93 (1.8)

with both  $\mathbf{X}^{AL}$  and  $\mathbf{X}^{PE}$ . At station 679, the comparison between  $\mathbf{X}^{AL}$  and  $\mathbf{X}^{PE}$  provided contradictory results. However, since the analysis performed with  $\mathbf{X}^{AL}$  consistently detected the occurrence of a breakpoint in year 1992 with confidence level of 0.99, the time series has been corrected. At two sites (site 13 and site 675), the comparison with  $\mathbf{X}^{AL}$  and  $\mathbf{X}^{PE}$  identified a breakpoint in two different, but very close years. In these cases, since no metadata are available, we corrected the breakpoint identified using the refer-

Table 3.4: Homogeneity analysis results that were not consistent comparing the tested times series  $\mathbf{Y}$  to the two reference time series  $\mathbf{X}^{AL}$  and  $\mathbf{X}^{PE}$  or varying the minimum percentage of overlap required between each one of the reference stations  $\mathbf{S}$  and the tested one  $\mathbf{Y}$ . In these cases, we did not apply any correction to the tested time series.

Site	$\mathbf{X}^{AL}$	$\mathbf{X}^{PE}$	Documented	Corrected
25	1976	2010	-	NO
51	1994	1981,2010	-	NO
61	2002	-	-	NO

ence time series with the highest correlation coefficient. At three sites (site 25, site 51 and site 61) the homogeneity test revealed the presence of breakpoints, but the results were ambiguous (see Table 3.4). In these cases the breakpoint was identified only with one of the two reference time series ( $\mathbf{X}^{AL}$  or  $\mathbf{X}^{PE}$ ) or its identification resulted dependent on the reference times series or on the *MOP*. Furthermore, metadata did not confirm the breakpoints. For these reasons, the breakpoints contained in Table 3.4 are not considered reliable and have not been corrected.

As shown in Figure 3.5 and in Table 3.3, the detected breakpoints are mainly distributed between 1980 and 2009. This is due to a higher number of time series covering this period of time and hence to a higher number of reference stations and of time series to be analyzed. Being the detected inhomogeneities uniformly distributed in this period, the procedure does not suffer for abrupt increases in the number of available reference stations observed in Figure 2.3.

Furthermore, Figure 3.5 shows the results of a self-homogeneity test for the homogenous time series of the dataset. We identified the location of the peak in the T-series, independently on its statistical significance. We can observe that 35% of the stations show a peak around 1990 indicating a possible regional change in the mean seasonal snow depth time series. This observation would be in accordance with what has been reported for the same period for other regions of the European Alps [see e.g. Beniston et al., 2003, Marty, 2008, Valt and Cianfarra, 2010, Reid et al., 2016]. Although 4 breakpoints are found in the vicinity of that period, only

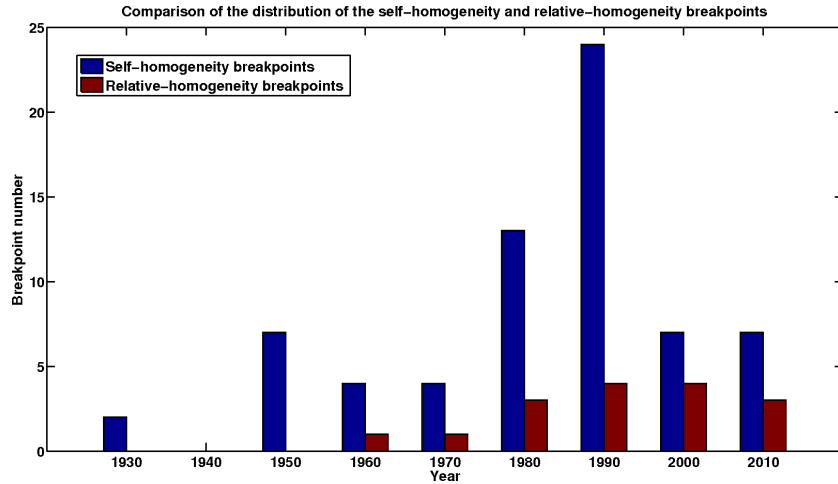


Figure 3.5: Histogram showing the distribution of the self-homogeneity breakpoints in comparison with that of the detected relative breakpoints. The bins are centered in the middle of a period of 10 years (e.g., 1990 defines the period 1986-1995).

2 of them (at sites 7 and 63) have been confirmed by metadata (relocation of the station). Therefore, the test provides robust results also in case of regional climate effects.

About 56% of the detected breakpoints were confirmed by metadata. The inhomogeneous time series are located at different altitudes and are distributed in the whole Province of Trento (Table 3.3), so that we can conclude that the proposed procedure is not biased by the location of the measuring stations. Metadata information are incomplete for this dataset. Hence, it is not possible to identify how often potential sources of breakpoint did not lead to an effective occurrence of a breakpoint. Moreover, we did not observe a statistically significant correlation between the estimated correction factor and the elevation of the stations.

As shown in Table 3.3, three of the detected breakpoints were the result of merging time series of different sources (sites 19, 24 and 63). An example of this type of inhomogeneity is in the Val Noana time series (site 19 in Table 3.3). As shown in Figure 3.6(a) the computation of the mean seasonal values for this time series includes mainly manual and historical data. Figure 3.6(a) and Table 3.5 show the different sources available for the Val Noana



site. Even though the manual and the historical stations were located very close to each other (see Table 3.5) the difference in the source introduced a breakpoint. The breakpoint was not identified as coincident at the time manual substituted the historical data, and this is because the temporal location of breakpoints detected using statistical methods, such as SNHT in this study, is affected by a degree of uncertainty, as shown in Lindau and Venema [2016].

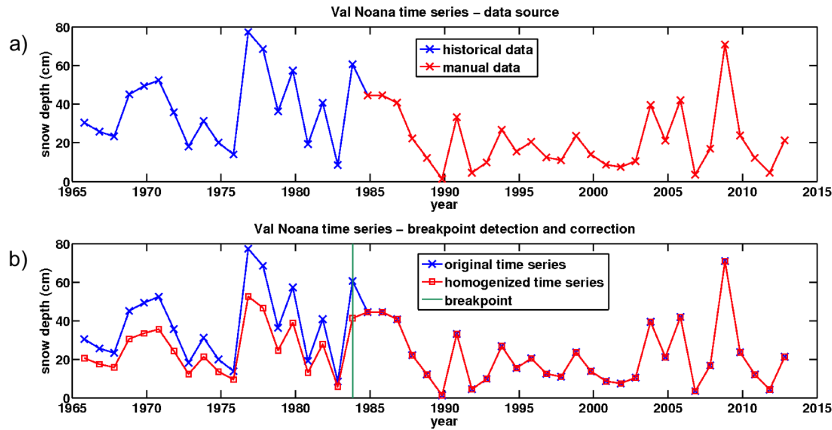


Figure 3.6: Time series of Val Noana (site 19). We can observe the different sources of the data composing the time series (a), the location of the breakpoint and the time series after its correction (b). The breakpoint due to the merging of data coming from two different sources is indicated with the green line.

Table 3.5: Stations forming the VAL NOANA site.

Source	Altitude	Latitude	Longitude	Period
manual	1020	46.13878	11.83925	1984-2012
historical	1030	46.13878	11.83917	1965-2000
automatic	1030	46.13878	11.83917	2012

Relocation of the station is the cause of 5 of the detected breakpoints. Figures 3.7a and 3.7b illustrate the application of the homogenization procedure to the snow depth time series of Madonna di Campiglio-Pancugolo (site 18 in Table 3.3). The two T-series, obtained by applying Eq. (3.10) to both the original and cor-

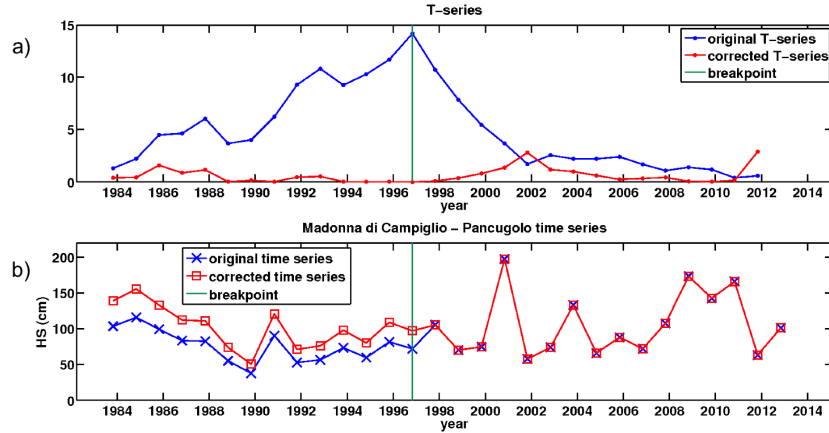


Figure 3.7: Time series relative to Madonna di Campiglio - Pancugolo (site 18) before and after the correction of the detected breakpoint (b) and the corresponding T-series (a).

rected time series are shown in Figure 3.7a with a blue and red line, respectively. The T-series computed with the original snow depth time series peaks at the season 1996/97, where a breakpoint is identified (the value of the T-series corresponding to season 1996/97 is above the critical threshold ( $L(29, 0.95) = 7.695$ )), which disappears in the corrected time series. Both the original and corrected time series are shown in the Figure 3.7b. The applied correction factor is 1.29.

### 3.3.2 Validation using artificially created breakpoints

To further analyze the reliability of the SNHT for the detection of breakpoints in case of mean seasonal snow depth data, we performed some additional tests [Marcolini et al., 2017].

We created an artificial breakpoint at a given time step  $a$  in a homogeneous time series of the datasets  $H1$  and  $H2$  by multiplying the snow depth values before that time by a factor  $c$ , which varies between 0.3 and 3. The location  $a$  of the breakpoint also varies and is located between  $a = 3$  and  $a = n - 3$ . Considering all possible combinations of  $c$  and  $a$  values applied to the available homogeneous time series (subsets  $H1$  and  $H2$ ), we generated a total of 4102 breakpoints. We then tested the homogeneity of

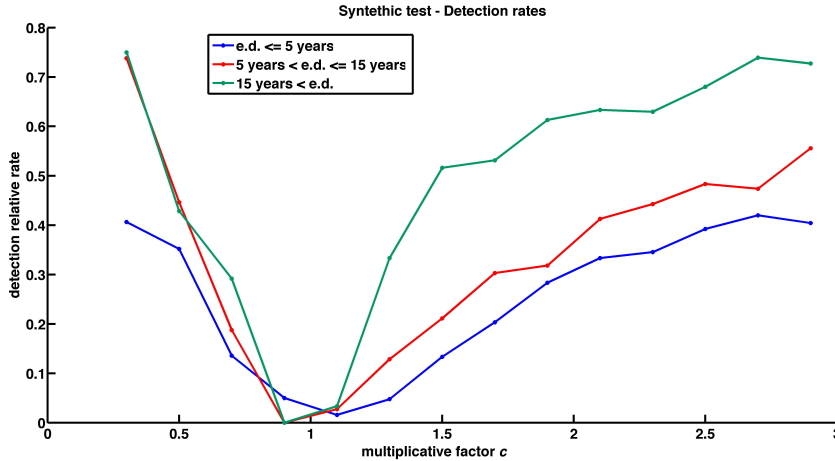


Figure 3.8: Percentage of detected artificial breakpoints. The applied inhomogeneities have varying entities and location respect to the edges of the time series (e.d. = edge distance).

each modified time series, choosing the reference stations among the homogeneous time series of our dataset. We considered as correctly detected those breakpoints identified in the interval of  $\pm 2$  years centered on  $a$ .

The proposed procedure shows better performances in the detection of the breakpoints located far from the edges of the time series (Figure 3.8). We find the highest rates of detections for breakpoints with values  $c$  lower than 0.7 or higher than 1.42. In particular, it is interesting to notice the low rate of breakpoint detections when  $c$  is between 0.9 and 1.1. This shows on one hand that the SNHT is less effective in detecting breakpoints with a small magnitude and, on the other hand, that the SNHT has a very small probability of false breakpoint detections. Moreover, 62% of the detected breakpoints were correctly located, 27% have a temporal shift of  $\pm 1$  year and 9% have a temporal shift of  $\pm 2$  years.

The occurrence of a breakpoint in a time series influences the analysis of the anomalies (defined as the difference between the mean seasonal snow depth of a given year and the mean seasonal snow depth computed over a reference period) in two ways. First the snow depth values before the occurrence of the breakpoint are

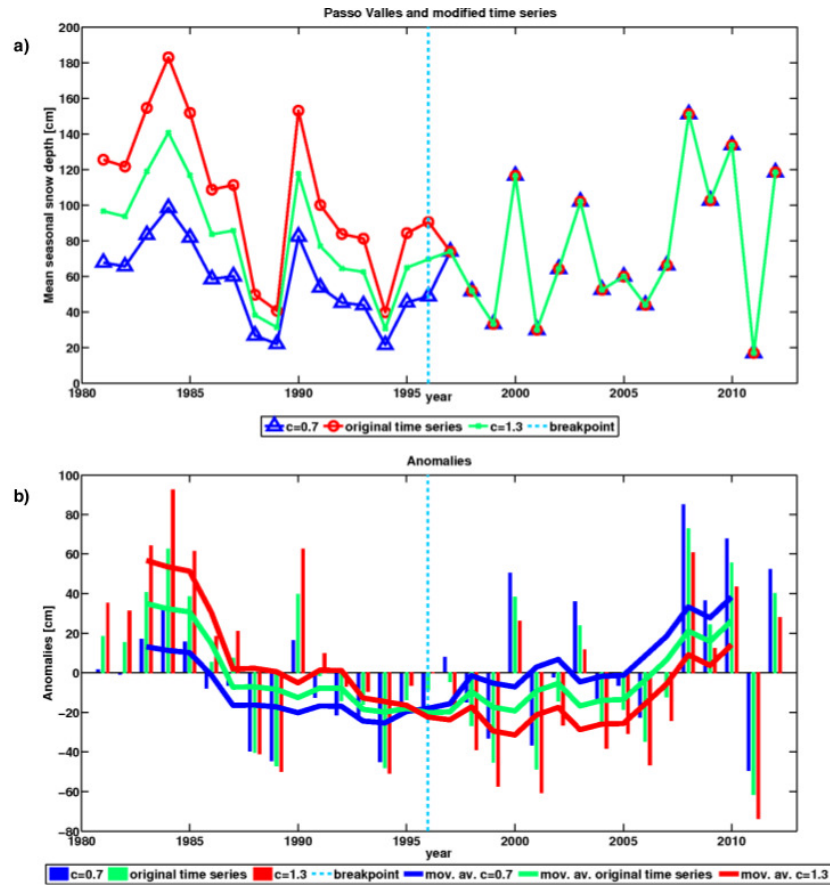


Figure 3.9: Changes of the time series of Passo Valles (2032 m, Lat: 46.3384, Long: 11.7998) if a breakpoint had occurred in 1996/97 (a) and the consequent changes in the anomalies respect to the mean of the whole time series (b).

either over or under estimated. Second the reference mean value is different from that of the homogeneous time series. To exemplify the influence of a breakpoint on the study of anomalies in the time series, we performed the following synthetic test. We set a breakpoint in 1996/97 to the homogeneous time series of Passo Valles (green line in Figure 3.9 a) ) and we generate two inhomogenous time series considering two different values of  $c$  (0.7 and 1.3, represented with a blue and a red line in Figure 3.9 a) respectively). In Figure 3.9 b) we show the computed anomalies for the three time series using the entire length of the series as

reference period. To better illustrate the changes in the observed behavior we also plotted the corresponding 5 years moving average. We observe that the occurrence of a breakpoint changes the entity of the anomalies. Even if the general behavior of the time series remains similar, the magnitude of the anomalies changes and their sign may also change (e.g., in 1986, 1991 and 1997). For the case  $c = 1.3$  we observe more positive anomalies in the '80s and more negative anomalies after the end of the '90s in respect to the anomalies identified by using the homogeneous time series. The anomalies of the time series obtained with  $c = 0.7$  are generally smaller than the anomalies computed using the homogeneous time series, with the exception of the positive anomalies starting from the year 2000. These changes are even more evident observing the difference between the moving average of the original and of the two modified time series. The '90s are of particular interest since, as shown in several works [see e.g., Marty, 2008, Valt and Cianfarra, 2010], they mark a period of particularly persistent and large negative anomalies in snow depth time series in the Alps. The 5 years moving averages show how a breakpoint can influence the temporal extension and location of this period. The red line shows a period of consecutive mean negative anomalies of 15 years between 1992 and 2007, the green line between 1987 and 2006 (19 years), the blue line between 1986 and 2001 (15 years). Therefore, even if all three time series show a period with lower mean seasonal snow depth in the '90s, its starting and ending point, its duration and its intensity are influenced by the occurrence of a breakpoint in the time series.

### 3.4 Intercomparison Experiment

As we already mentioned, to the best of our knowledge only two algorithms have been used for the homogenization of mean seasonal snow depth time series: the procedure based on the SNHT shown in sections 3.2 and 3.3, and HOMOP [Koch et al., 2014]. For the homogenization of other variables, such as temperature and precipitation, several intercomparison studies have been performed [see e.g. Easterling and Peterson, 1995, Ducré-Robitaille et al., 2003, Venema et al., 2012]. We decided to perform an intercomparison experiment where 25 snow depth time series have been homogenized on a seasonal basis with the procedure based on the

SNHT introduced in section 3.2 and with HOMOP (see section 3.4.1). The time series belong to the Austrian dataset presented in section 2.2. This kind of comparison is particularly interesting since the SNHT and PRODIGE (the homogenization algorithm inside HOMOP) belong to two different classes of homogenization algorithms, in fact the SNHT is based on the comparison of the tested time series with a reference one, while PRODIGE is based on a pairwise comparison between the time series.

### 3.4.1 HOMOP

The homogenization analysis performed by the Central Institute of Meteorology and Geodynamics of Vienna is based on the integrated software package HOMOP [Nemec et al., 2013]. This is composed by the method PRODIGE [Caussinus and Mestre, 2004] for the detection of multiple inhomogeneities in the time series, and by the method INTERP [Vincent et al., 2002] for the computation of the corresponding correction factors. For the application of the method PRODIGE, stations with a correlation coefficient larger than 0.7 on a daily scale were selected. An additional distance criteria has been introduced in order to guarantee that the stations have experienced the same climatic conditions, therefore only stations with a maximum horizontal distance of 100 km and a maximum vertical distance of 300 m have been considered. PRODIGE analyzes the ratio time series between the tested time series and each reference station and it is based on a penalized log-likelihood criterion [Mestre et al., 2011]. Moreover, the criteria of Caussinus and Lyazrhi [1997], Jong et al. [2003] and Lebarbier [2005] are used in the detection process.

The analysis has been performed for two kind of mean seasonal values: the ones computed between December and February and the ones computed between November and March. A time series is classified as inhomogeneous according to one of the above mentioned criteria, if more than half of the reference stations detected the break. A breakpoint is considered confirmed, if it has been detected by at least two criteria in both seasons. Once a breakpoint is detected, its temporal location can be adapted according to the metadata.

For the correction of the detected breakpoint, a modified version of the INTERP method [Vincent et al., 2002] has been used. Similarly to the SNHT, a multiplicative factor is applied, which

was calculated as

$$\text{correction factor} = \frac{\text{median} \left\{ \frac{C_2}{R_2} \right\}}{\text{median} \left\{ \frac{C_1}{R_1} \right\}}, \quad (3.21)$$

where indices 1 and 2 represent the time period after and before the detected inhomogeneity and  $C$  ( $R$ ) is the mean seasonal (ND-JFM) snow depth of the candidate (reference) time series. The most recent part of the time series (after the breakpoint), is left unchanged, while the oldest part is multiplied times the correction factor.

### 3.4.2 Breakpoint detection

As shown in Table 3.6, in 56% of the cases we had an agreement on the breakpoint detection using the two homogenization algorithms, i.e. both tests judged a time series as homogeneous or individuated the same breakpoints. We considered that the two homogenization algorithms have individuated the same breakpoint, if the difference between the time location of the detected breakpoints was maximum two years, since the SNHT can have some uncertainty in the identification of the correct year, as also discussed in section 3.3, and since the temporal location of breakpoints identified by PRODIGE have been sometimes adjusted using the metadata (see section 3.4.1). In 6 cases, the time series were judged as homogeneous by the SNHT algorithm, while PRODIGE detected a suspicious breakpoint. The breakpoints found in the time series of Holzgau, Umhausen, Mayrhofen, Innerkrems, Kals and Weitra were considered not reliable due to low snow depth. In fact, they all occur for winter seasons with on average low snow depths (beginning of the 1970s, end of the 1980s, beginning of the 1990s, see also [Marty, 2008]) and are hence considered suspicious breaks since they cannot be linked to changes in the observational environment. Apparently, the detection algorithm PRODIGE is more sensitive to changes at low snow depths. The breakpoints found in the time series of Weitra and of Wien Hohe Warte were classified as suspicious because they are close to the end of the tested time series (1 and 3 years from the edge, respectively).

The analysis of the sites Weitra, St.Leonhard i.P., Tamsweg and Göstling an der Ybbs gave different results for the two meth-

Table 3.6: Results of the homogeneity analysis. In the cases in which the result is indicated with Y, PRODIGE detected suspicious breaks at low snow depths.

Site	Results	
	PRODIGE	SNHT
Kufstein	-	-
Jenbach	-	-
Klagenfurt	-	-
Ötz	-	-
Kelchsau	-	-
Untertauern	-	-
Schladming	-	-
Frankenfels	-	-
Almsee (Forsthaus)	-	-
Schönberg im Stubaital	-	-
Rauris	1973	1973
Bad Gastein	1993	1995
Bad Gastein	1972	1974
Galtür	1988	1987
Oed	1996	1996
Holzgau	Y	-
Umhausen	Y	-
Mayrhofen	Y	-
Innerkrems	Y	-
Kals	Y	-
Weitra	Y	1975
Wien Hohe Warte	Y	-
Weitensfeld	Y	1998
St.Leonhard i.P.	1985	-
Göstling an der Ybbs	-	2001
Tamsweg	1983,1998	-

ods. We did not find any correlation between the altitude of the site or its position and the difference in the results of the algorithms.

In Figure 3.10 we can observe three time series. The first two plots show the time series of Galtuer and Untertauen, where the



Table 3.7: Correction factors resulting from the homogeneity analysis.

Site	Correction factors	
	INTERP	SNHT
Rauris	1.57 - 1.64	2.18 - 1.58
Bad Gastein	1.38	1.84
Galtür	1.22	1.24
Oed	1.22	1.81
Weitra	-	0.50
Weitensfeld	-	0.66
St.Leonhard i.P.	0.95	-
Göstling an der Ybbs	-	1.36
Tamsweg	0.83-0.73	

results of the SNHT and of PRODIGE agree. The time series relative to St. Leonhard gave instead different results with the two homogenization methods.

### 3.4.3 Correction of the inhomogeneities

As described in sections 3.2.1 and 3.4.1, the correction of a breakpoint in a time series in this work was performed by multiplying the older part of the time series (i.e. from the beginning up to the breakpoint) times a correction factor. Due to the multiplicative approach, it is not possible to add new snow days to the corrected time series, but it is possible just to modify the registered mean seasonal snow depth. The correction factors of the time series where the two algorithms detected only one breakpoint (Bad Gastein, Oed and Galtür) are in good agreement (see Table 3.7). The correction factors of the time series of Rauris, where both algorithms detected two breakpoints, are in good agreement for the correction of the most recent inhomogeneity, while the correction factors for the oldest breakpoints are quite different. Notice that the SNHT is primarily thought to be applied for the detection and correction of one single breakpoint [Alexandersson and Moberg, 1997], and hence its performance in case of multiple breakpoint detection and correction is generally poor.

There are several possible reasons to justify the variability in

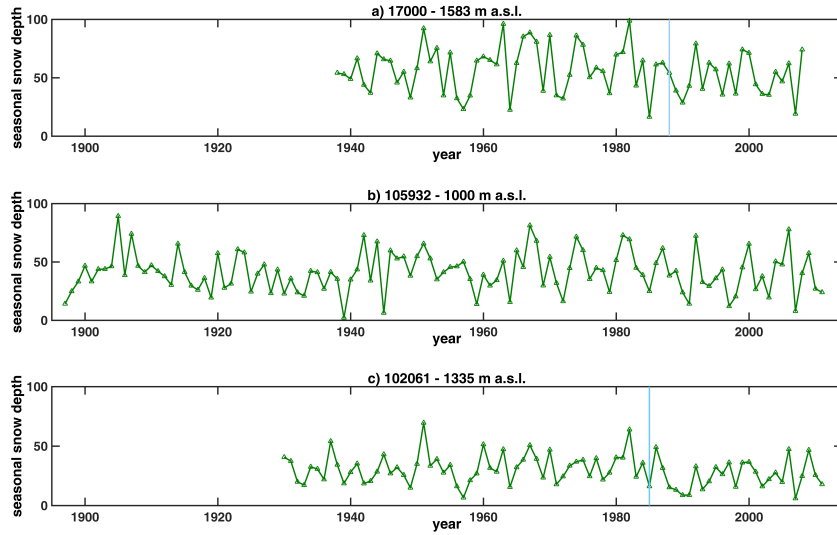


Figure 3.10: Example of the time series used for the intercomparison experiment. In the upper panel (a) the time series of Galtür where both the used algorithm found the same unique breakpoint is shown. The time series of Untertauern, shown in the second panel (b), is homogeneous according to the results of both the algorithms. The last plot (c) shows the time series of St. Leonhard im Pitztal, which was classified as homogenous by the SNHT and as inhomogeneous by PRODIGE. The vertical blue lines show the location of the breakpoints.

the computed correction factors. First, the two mathematical expressions reported in Eq. 3.21 and Eq. 3.18 are not equivalent. The second and probably most important factor, is that the candidate station is not necessarily compared with the same set of reference stations, since the procedure used by the two methods to select them is different.

### 3.4.4 Implication for time series analysis

It is interesting to observe some of the possible effects of the correction of an inhomogeneous time series on its climatological analysis. A first interesting point is the one shown in figure 3.11. The upper panel (Figure 3.11a) shows the anomalies respect to the period 1961-1990 of the original time series after having smoothed them with a 5-years moving average. The second and third panels

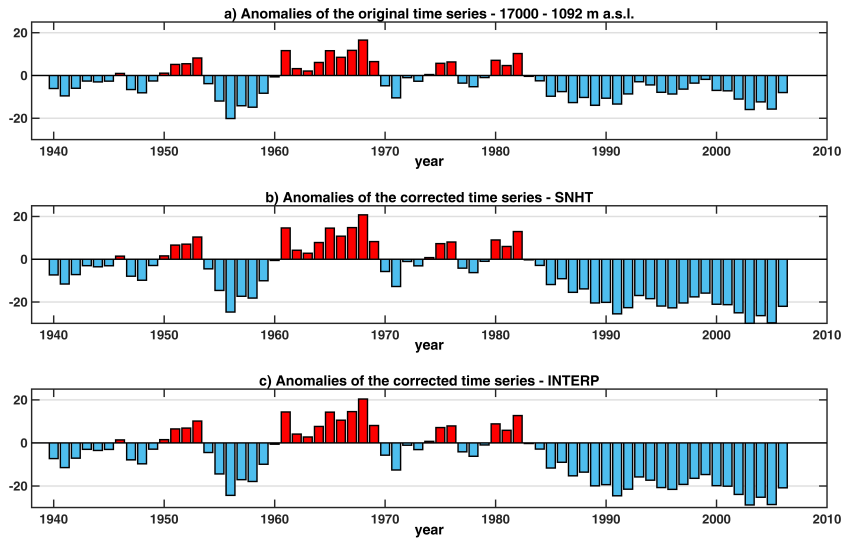


Figure 3.11: Anomalies respect to the period 1961-1990 of the time series of Galtür before (a) and after the correction of the detected breakpoint in 1988 (b for SNHT and c for INTERP). The bars show the anomalies of the same time series, the color should only differentiate between positive (red) and negative (light blue) anomalies.

(Figures 3.11b and 3.11c) show the same figure for the time series corrected with the correction factors computed by the SNHT and INTERP respectively. The corrected time series of Galtür show much more pronounced negative anomalies during the '90s, than the uncorrected one. The Alpine region comprises individual regions that are characterized by different variations of the snow depth (due to several factors such as location, elevation, influence of different weather patterns). Anyway, it is worth noting that the behavior of the corrected time series described above is more consistent with what reported in literature about the behavior of the snow depth in this period in the Alpine region [see e.g. Beniston et al., 2003].

Another interesting example of how the correction of a time series can influence a climatological analysis is shown in figure 3.12. This figure shows the anomalies, smoothed with a 5-years moving average, of the time series of Bad Gastein before and after the correction. In the original time series, we can observe two periods with large anomalies in the '70s and in the '80s, where

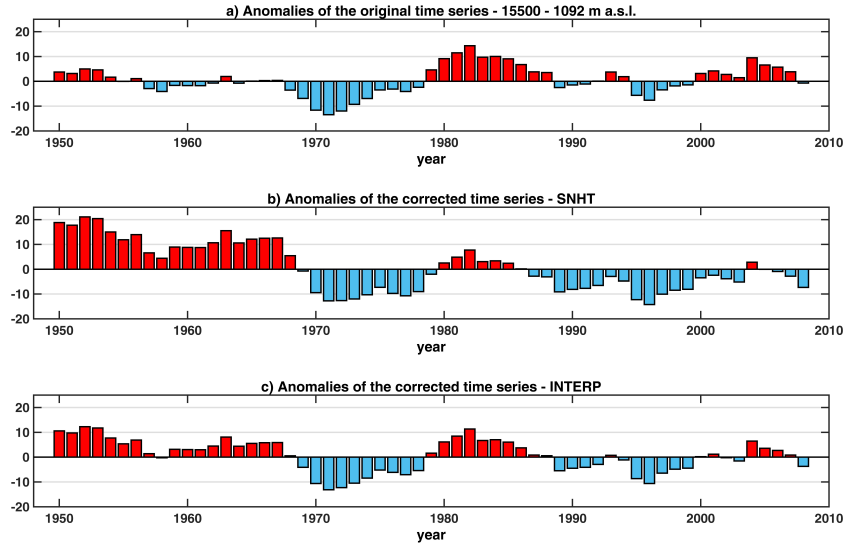


Figure 3.12: Anomalies respect to the period 1961-1990 of the time series of Bad Gastein before (a) and after the correction of the detected breakpoint in 1972 (b for SNHT and c for INTERP). The bars show the anomalies of the same time series, the color should only differentiate between positive (red) and negative (light blue) anomalies.

there is respectively a decrease and an increase in the snow depth respect to the mean value. Furthermore, there seems to be a slight increase in the snow depth after 2002. After the correction of the time series with the factors computed by INTERP, the behavior of the time series seems to be quite similar, even if the amplitude of the anomalies are smaller in the '70s and '80s and we find bigger positive anomalies before 1967. These changes are more evident in the time series corrected with the correction factors computed with the SNHT. Also the negative anomalies at the end of the '80s and in the '90s are more evident. The positive anomalies in the beginning of the '80s are instead very small.

Correction factors can also have little influence for the climatological analysis, when they are close to 1, as we can observe for example in figure 3.13, where the anomalies, again smoothed with a 5-years moving average, for the time series of St.Leonhard i.P. are shown. This time series was classified as homogeneous by the SNHT.

Even after the application of the correction factor, the site

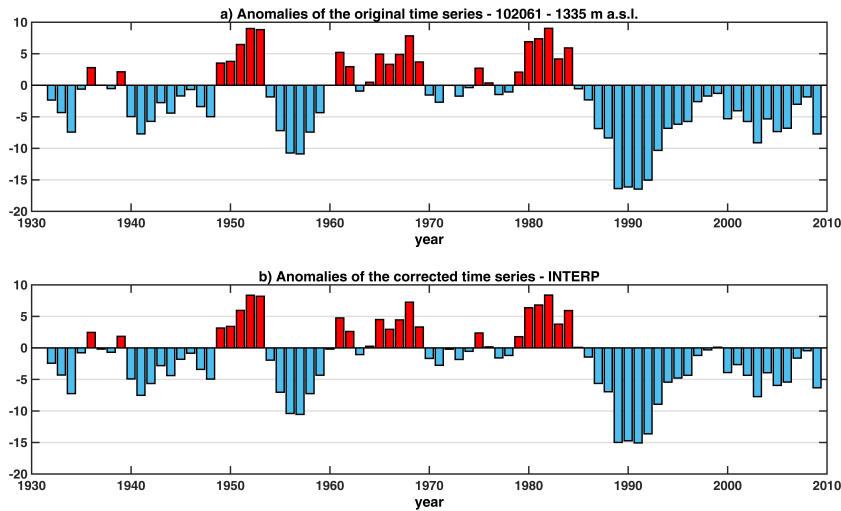


Figure 3.13: Anomalies respect to the period 1961-1990 of the time series of St.Leonhard i.P. before (a) and after the correction of the detected breakpoint in 1985 with INTERP (b). The bars show the anomalies of the same time series, the color should only differentiate between positive (red) and negative (light blue) anomalies.

specific behavior of the time series of Galtür, Bad Gastein and St.Leonhard i.P. (figures 3.11, 3.12 and 3.13), related to their specific aspects (e.g. the altitude, the regional variability), is preserved. This is important in order to characterize the relationship between the snow depth and other related climatic and environmental factors.

### 3.5 Conclusions

In section 3.2 we presented an algorithm for the homogenization of mean seasonal snow depth time series based on the Standard Normal Homogeneity Test [Alexandersson, 1986, Alexandersson and Moberg, 1997]. Most of the homogenization algorithms developed up to now, indeed, are mainly suited for the analysis of other variables, such as precipitation and temperature. In this chapter we also showed its application to the mean seasonal snow depth dataset of the Province of Trento, presented in section 2.1. This has shown good agreement with the available metadata. We

further showed an application of the algorithm to synthetic time series. Moreover, the possible influence of a breakpoint was investigated analyzing the anomalies of an homogeneous time series to which two different kinds of breakpoint had been applied and comparing them to the anomalies of the original time series. We observed that the presence of a breakpoint can influence both the evaluation of the magnitude of the changes happened in the considered time interval and the temporal location of these variations.

To the best of our knowledge, only another study approached the homogenization of mean seasonal snow depth on a statistical basis, the one performed by Roland Koch at the Central Institute of Meteorology and Geodynamic of Vienna [Koch et al., 2014]. We then decided to perform an intercomparison study in order to evaluate the performances of the two methods. We found good agreement between the two algorithms. Moreover, further investigations of the effects of the breakpoints on the evaluation of the variations of the time series also proved the importance of testing the homogeneity of hydrological time series.

## Chapter 4

# Wavelet analysis

### 4.1 Introduction

Wavelet analysis has been shown to be a useful tool for detecting localized variations of power within a time series [e.g., Lau and Weng, 1995, Coulibaly and Burn, 2004, Guan et al., 2011, Carey et al., 2013]. The wavelet analysis is performed by decomposing the time series into a transformed variable in time-frequency space to determine both the dominant modes of variability and how these modes vary in time. We performed wavelets analysis by using the Matlab toolboxes developed by Torrence and Compo [1998] and by Grinsted et al. [2004]. Object of the analysis in this chapter are the snow depth dataset of the Trentino - Alto Adige within the Adige catchment presented in section 2.3 and the discharge dataset of the Adige and upper Inn river basins presented in section 2.4.

### 4.2 Continuous wavelet transform

Similarly to the Fourier transform, the continuous wavelet transformation looks for similarity between a signal and a well-known mathematical function [Labat, 2005]. The major difference to the Fourier transform, which is also the strength of this kind of analysis, is that the well-known mathematical function, which for the wavelet analysis is a wavelet function, is applied several times with different scales to the analyzed time series and at different temporal positions. This allows to determine not only the frequency content of a signal, as also the Fourier analysis can do, but also the frequency time - dependence [Labat, 2005].

The continuous wavelet transformation of a discretized signal  $x_n$ ,  $n = 1, \dots, N$ , sampled at the time interval  $\delta t$ , is defined as it follows [Grinsted et al., 2004]:

$$W_n(s) = \sum_{n'=0}^{N-1} x_{n'} \psi^* \left[ \frac{(n' - n)\delta t}{s} \right] \quad (4.1)$$

where  $\psi(\eta)$  is a band-pass filter and the superscript  $*$  indicates the complex conjugate. Mathematically, the transform (4.1) is the convolution of the signal  $x_n$  with the scaled version of the wavelet function  $\psi$ , which is obtained from a “mother wavelet”  $\psi_o(\eta)$ , normalized at each scale  $s$  to fulfill the following condition:

$$\sum_{k=0}^{N-1} |\psi(s\omega_k)|^2 = N \quad (4.2)$$

where  $\omega_k = 2\pi k/(N\delta t)$  for  $k \leq N/2$ , and  $\omega_k = -2\pi k/(N\delta t)$  for  $k > N/2$  is the angular frequency and  $s$  is the wavelet scale [Torrence and Compo, 1998]. Similarly to other studies analyzing variability in climate and hydrological signals [see e.g., Lau and Weng, 1995, Coulibaly and Burn, 2004, Carey et al., 2013, Guan et al., 2011], the Morlet function was used as “mother wavelet” for its ability to evidence fluctuations in the time series:

$$\psi_o(\eta) = \pi^{-1/4} e^{i\omega_o\eta} e^{-\eta^2/2} \quad (4.3)$$

where  $\omega_o$  is the dimensionless frequency [Grinsted et al., 2004] and  $\eta$  is the dimensionless time. Following Torrence and Compo [1998]  $\omega_o$  is set to 6, such that the wavelet scale is almost identical to the corresponding Fourier period.

The wavelet transform (4.1) is then computed for a selection of scales:

$$s_j = s_0^{2^j \delta_j}, (j = 1, 2, \dots, J) \quad (4.4)$$

where  $s_0$  is the smallest scale considered in the analysis and  $J = \delta_j^{-1} \log_2(N\delta_t/s_0)$  determines the largest scale. In addition,  $s_0$  should be chosen such that the equivalent Fourier period is approximately  $2\delta t$ .

The wavelet power spectrum (WPS)  $|W_n(s)|^2$  is defined as the product of wavelet transform,  $W_n(s)$ , by its conjugate,  $W_n^*(s)$  and it represents the energy of the scale  $s$ . It is useful in the



identification of fluctuation scales with the largest influence on the signal. In particular, a larger positive amplitude implies a higher positive correlation between the signal and the wavelet and a large negative amplitude implies a high negative correlation.

The transform provided by eq. (4.1) assumes that the time series is periodic. Since hydrological time series are a-periodic and only a fraction of the entire time series is available, a bias is introduced at the beginning and at the end of the time series. To overcome this problem, Torrence and Compo [1998] suggested to pad the end of the time series with zeroes. The transformed signal is then considered only in the time windows of the recorded signal, but since the discontinuities at both edges of the time series (due to the introduction of the zero padding) propagate inside the time series, a cone of influence where edge effects are significant is identified [Torrence and Compo, 1998]. The cone of influence is shown by fogging the portion of the contour plot showing the wavelet transform affected by edge effects.

#### 4.2.1 Wavelet coherence

The cross wavelet transform of two time series  $X$  and  $Y$  is defined as  $W^{XY} = W^X W^{Y*}$ , where  $W^X$  and  $W^Y$  are the continuous wavelet transform of  $X$  and  $Y$  respectively and  $*$  indicates the complex conjugation [Grinsted et al., 2004]. It is interesting to observe how coherent the cross wavelet transform is in time frequency domain. Following Grinsted et al. [2004], the wavelet coherence of two time series can be defined as

$$R_n^2(s) = \frac{|S(s^{-1}W_n^{XY}(s))|^2}{S(s^{-1}|W_n^X(s)|^2)S(s^{-1}|W_n^Y(s)|^2)}, \quad (4.5)$$

where  $S$  is a smoothing operator defined as

$$S(W) = S_{scale}(S_{time}(W_n(s))). \quad (4.6)$$

The wavelet coherence varies between 0 and 1, as also shown by the definition in equation 4.5. It can be interpreted as a localized correlation coefficient in time frequency domain [Grinsted et al., 2004].

### 4.3 Application to snow data

As a first application of the wavelet transform and of the wavelet coherence, we show the analysis of the snow dataset relative to the stations of the Adige river basin described in section 2.3.

#### 4.3.1 Analysis of the Trentino - Alto Adige snow depth dataset

##### Variable definition

Snow cover duration ( $SCD$ ) is defined as the number of days in a given snow season (from 1 November of a given year to 30 April of the following year) with snow depth higher than 30 cm [Durand et al., 2009, Valt and Cianfarra, 2010]. This threshold was selected since it has a practical implication for winter tourism [Marty, 2008]. The mean seasonal snow depth is computed by averaging the daily snow depth between 1 November of a given year and 30 April of the following year.

Another quantity useful to investigate the sensitivity of the fluctuation of the mean seasonal snow depth  $HS$  is:

$$HS_m = \frac{HS - \langle HS \rangle}{\langle HS \rangle}, \quad (4.7)$$

where  $HS_m$  quantifies the relative fluctuation of  $HS$  with respect to its mean value  $\langle HS \rangle$ . The use of a standardized variable allows to compare the results for data collected at different elevations. Similarly we can define the fluctuation of  $SCD$  as

$$SCD_m = \frac{SCD - \langle SCD \rangle}{\langle SCD \rangle}. \quad (4.8)$$

##### Statistical analysis

In Figures 4.1 and 4.2 we observe some plots for the statistical analysis of the snow depth and snow cover at different altitude in the Adige river basin: stations below 1350 m a.s.l., stations between 1350 m and 1650 m a.s.l., stations between 1650 m and 2000 m a.s.l. and stations above 2000 m a.s.l.. The Hovmöller-type diagrams in Figures 4.1a and 4.2a display the interested variable, using a color code, as a function of time and elevation which are reported in the abscissa and the ordinate, respectively. In particular, in plot 4.2a the colormap changes at a threshold of 100

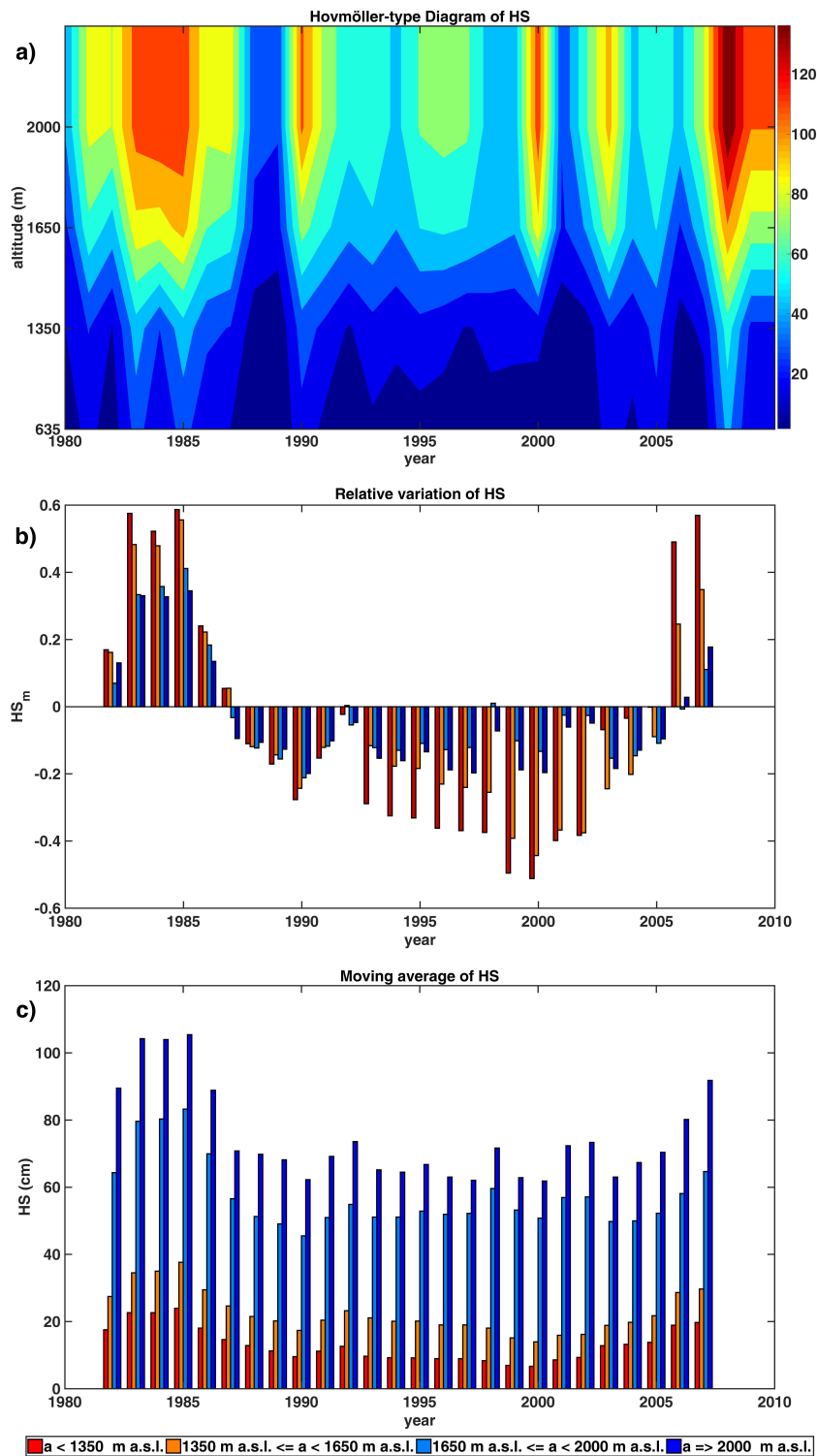


Figure 4.1: Hovmöller-type diagramm of the mean seasonal snow depth (a), 5-year moving average of  $HS_m$  (b) and 5-year moving average of the snow depth (c). The units of the colormap of the Hovmöller-type diagramm is centimeters.

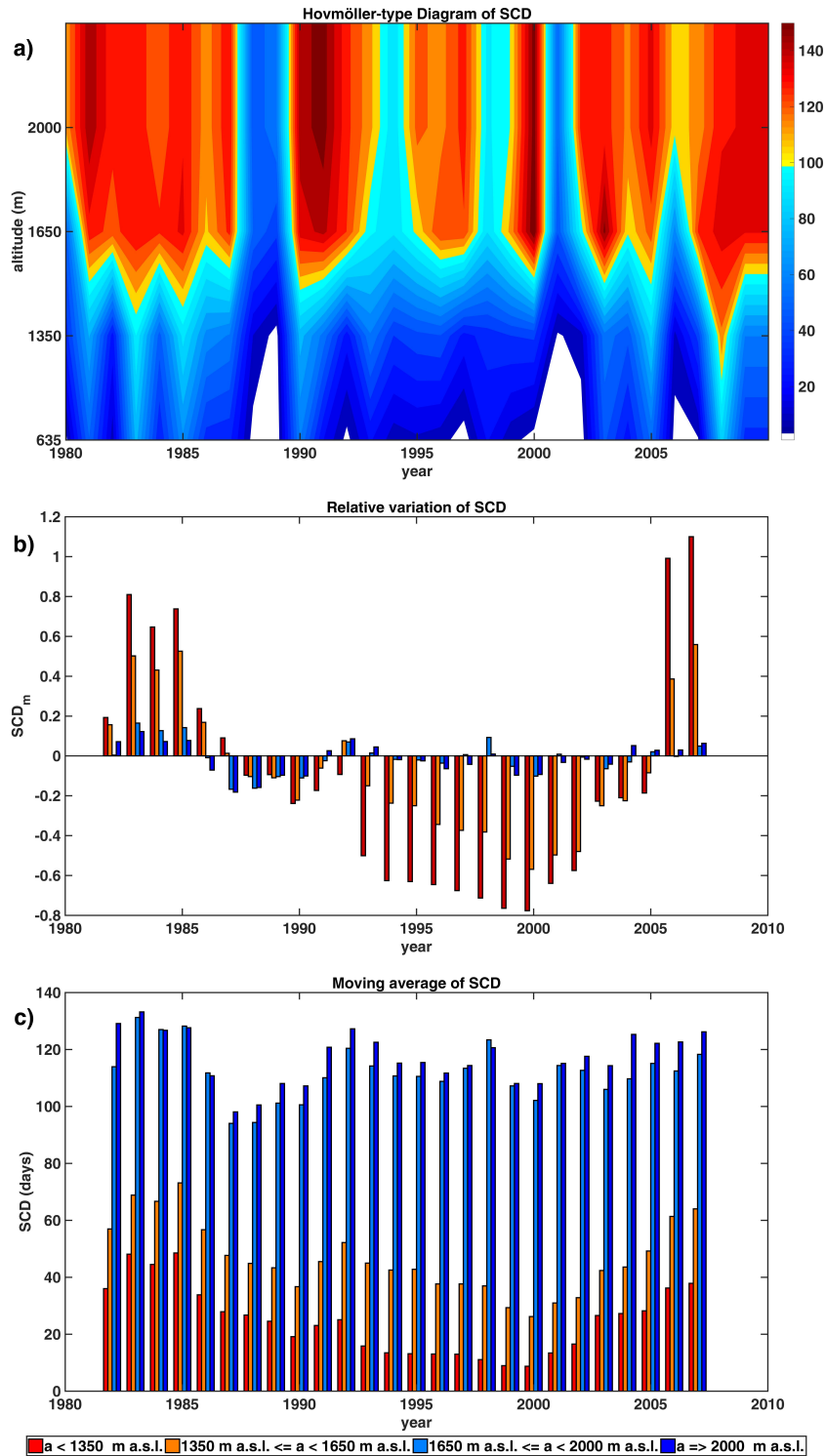


Figure 4.2: Hovmöller-type diagramm of the snow cover duration (a), 5-year moving average of  $SCD_m$  (b) and 5-year moving average of the snow cover (c). The units of the colormap of the Hovmöller-type diagramm is days.

days with more than 30 cm of snow on the ground during the winter season, since this is suggested to consider a site economically profitable for winter activities [Valt and Cianfarra, 2010]. In both figures, a sharp change in the color pattern occurs at about 1650 m a.s.l., which is within the sensitive elevation range observed by Beniston [1997] and by Laternser and Schneebeli [2003b] for the Swiss Alps.

Considering the temporal variability, we identify three periods. The first one ranges between 1980 and 1987, the year in which Marty [2008] identified a regime shift (based on the sequential t-test analysis of regime shifts) in the snow cover duration in Switzerland. The second one (from 1988 to 1999) and the third one (from 2000 to 2009) are identified observing the increase in the mean seasonal snow depth in Figure 4.1a starting from year 2000. To verify the statistical significance of our considerations both Kolmogorov-Smirnow and Mann-Whitney tests have been conducted with a significance level of 5%. The results of the tests show that the period 1988-1999 is characterized by significantly lower values of mean seasonal snow depth than the period 1980-1987 (see results of the statistical tests for comparison  $P1 - P2$  in Tables 4.1 and 4.2). On the contrary, in the period between 2000 and 2009, the increase in mean seasonal snow depth was not identified as statistically significant with respect to the period 1988-1999 (see results of the statistical tests for comparison  $P2 - P3$  in Tables 4.1 and 4.2). The same analysis performed on snow cover duration time series identify a statistically significant (10% confidence level) change only for the two lowest elevation classes for comparison  $P1 - P2$ .

Table 4.1: P-values of the Kolmogorov-Smirnoff applied to the average mean seasonal snow depth at different altitudes in three different periods: P1 stands for the period from 1980 to 1987, P2 stands for the period from 1988 to 1999, P3 stands for the period from 2000 to 2009.

Altitude class	Kolmogorov-Smirnoff	
	P1 - P2	P2 - P3
$a < 1350$	0.01	0.23
$1350 \leq a < 1650$	0.01	0.56
$1650 \leq a < 2000$	0.01	0.27
$2000 \leq a$	0.00	0.23

Table 4.2: P-values of the Mann-Whitney Tests applied to the average mean seasonal snow depth at different altitudes in three different periods: P1 stands for the period from 1980 to 1987, P2 stands for the period from 1988 to 1999, P3 stands for the period from 2000 to 2009.

Altitude class	Mann-Whitney	
	P1 - P2	P2 - P3
$a < 1350$	0.03	0.63
$1350 \leq a < 1650$	0.03	0.67
$1650 \leq a < 2000$	0.03	0.35
$2000 \leq a$	0.01	0.25

We further analyzed mean seasonal snow depth and snow cover data considering also  $HS_m$  and  $SCD_m$  (Figures 4.1b and 4.2b) and  $HS$  and  $SCD$  (Figures 4.1c and 4.2c), aggregated according to the four elevation classes and smoothed using a 5-years moving average to eliminate short term fluctuations. The choice of a 5-years moving average is appropriate considering the 30-years length of the investigated period and the length of the time intervals with low and high snow depth. Figures 4.1b and 4.2b highlight that stations located below 1650 m a.s.l. are affected by a larger variability than stations above this limit. In particular, in the period 1988-2005 fluctuations with respect to the mean both of the snow depth  $HS_m$  and of the snow cover duration  $SCD_m$ , as defined by Eq. (4.7) and by Eq. (4.8) are in general more negative below 1650 m a.s.l.. Furthermore, in the two periods with positive fluctuations with respect to the mean (i.e., 1980-1987 and 2006-2010) they are in general larger at lower elevations.

### Correlation with temperature

Temperature exerts a strong influence on the mean seasonal snow depth and on the snow cover duration. The 5-years moving average of the daily maximum temperature averaged over the winter season (from 1 November to 30 April) is shown in Figure 4.3. The maximum temperature increases in the period 1985-1990. Anomalies with respect to the mean of the maximum temperature of the winter season in the period 1980-2010 are shown in Figure 4.4, which evidences more clearly the positive anomalies in the maximum winter temperature of the '90s, which attenuate around the

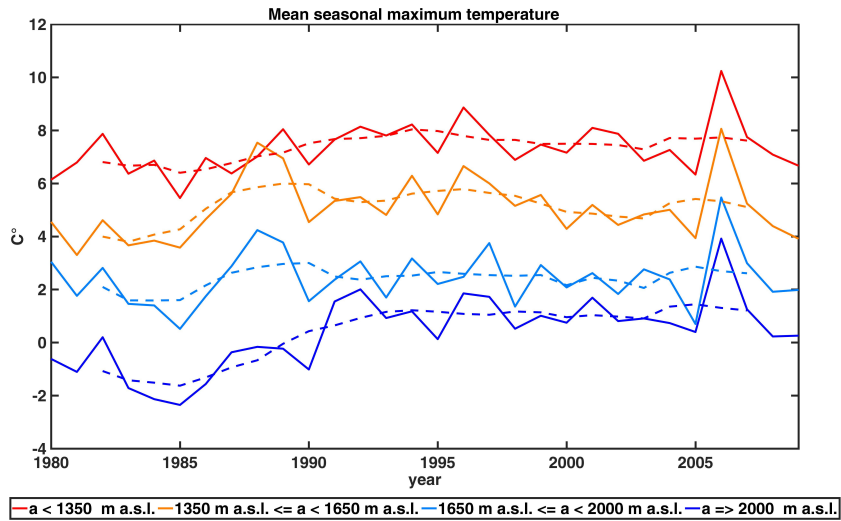


Figure 4.3: Moving average of the mean seasonal maximum temperature at different altitude classes.

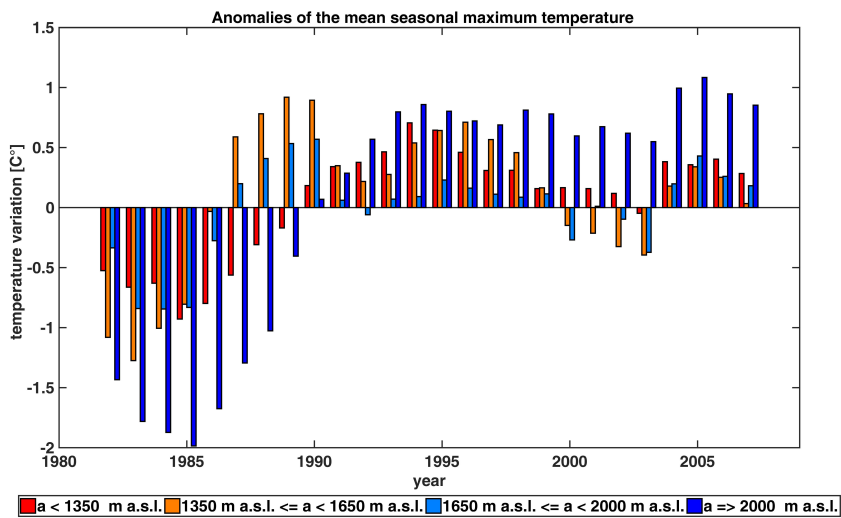


Figure 4.4: Anomalies of the mean seasonal maximum temperature at different altitude classes respect to the average computed on the whole available period.

year 2000, though not at high elevation, to increase again after 2003. At high elevations ( $> 2000$  m a.s.l.) anomalies are always positive after 1988 and in the last three years of the time series, 2005-2007, they are the largest of the entire observation period. Figures 4.5 and 4.6 show that the mean seasonal snow depth and the snow cover duration are negatively correlated with the seasonal mean of the maximum daily temperature. Table 4.3 reports the results for the correlation analysis performed both using Kendall's tau  $\tau_k$  as well as Spearman's rho  $\rho_S$  correlation coefficients. These are correlation coefficients based on the ranks of the data of the compared time series. They vary between -1, meaning perfect negative dependence, and 1, indicating perfect positive dependence. Both correlation coefficients are 0 in case of independence, but  $\tau_k$  or  $\rho_S$  equal to 0 does not necessarily imply that the compared variables are independent.

The Kendall's tau  $\tau_k$  is defined as

$$\tau_k[X, Y] = P\{(X_1 - Y_1)(X_2 - Y_2) > 0\} - P\{(X_1 - Y_1)(X_2 - Y_2) < 0\}, \quad (4.9)$$

where  $X = (X_1, X_2)$  and  $Y = (Y_1, Y_2)$  are i.i.d. vectors of continuous random variables [Embrechts et al., 2001, Nelsen, 2003]. Two observations  $(x_1, x_2)$  and  $(y_1, y_2)$  are concordant if  $(x_1 - x_1)(y_2 - y_2) > 0$ , discordant if  $(x_1 - x_1)(y_2 - y_2) < 0$ . In this way, the Kendall's tau can be seen as a measure of concordance for bivariate random vectors.

As shown in Embrechts et al. [2002], given two random variables  $X$  and  $Y$  with marginals c.d.f.s  $F_1$  and  $F_2$ , the Spearman's rho  $\rho_S$  is defined as

$$\rho_S[X, Y] = \rho[F_1(Y), F_2(Y)]. \quad (4.10)$$

The Spearman's rho can be hence interpreted as the linear correlation of the probability transformed random variables.

As shown in Table 4.3, the negative correlation between temperature and  $HS$  is strongest for the stations located in the lowest elevation range ( $\tau_k = -0.52$  and  $\rho_S = -0.71$ ) and it is weakest at the highest elevation ( $\tau_k = -0.30$  and  $\rho_S = -0.42$ ). The loss of dependency in correlation with the altitude of the stations is even more evident when looking at the snow cover duration. In this case the values of the correlation coefficients for the stations above 2000 m a.s.l. are  $\tau_k = -0.15$  and  $\rho_S = -0.22$ , respectively.



This behavior is consistent with the control exerted by temperature on snowmelt and on the transition between solid to liquid precipitations.

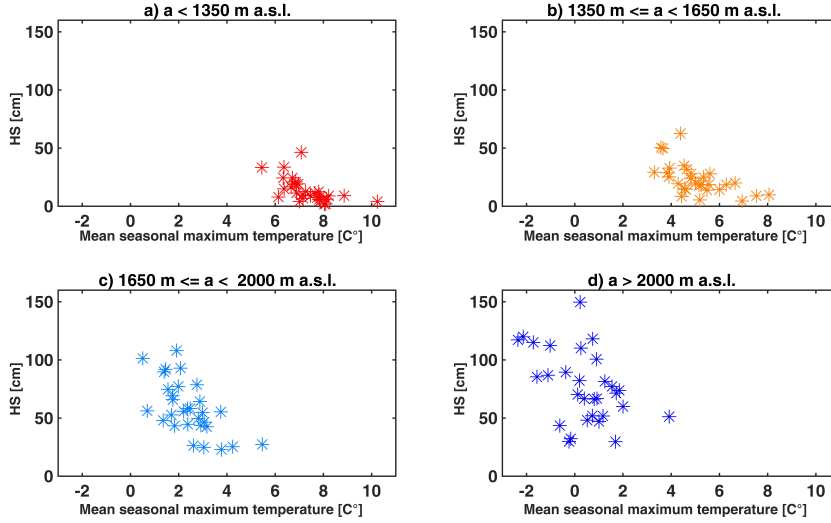


Figure 4.5: Correlation between the mean seasonal maximum temperature and the mean seasonal snow depth at different altitude.

Table 4.3: Correlation coefficients of the mean seasonal snow depth and of the snow cover duration with the mean maximum temperature at different altitude classes.

Altitude class	Snow depth		Snow cover	
	$\tau_k$	$\rho_S$	$\tau_k$	$\rho_S$
$a < 1350$ m a.s.l.	-0.52	-0.71	-0.54	-0.71
$1350 \text{ m} \leq a < 1650$ m a.s.l.	-0.46	-0.64	-0.46	-0.63
$1650 \text{ m} \leq a < 2000$ m a.s.l.	-0.46	-0.64	-0.24	-0.39
$2000 \text{ m a.s.l.} \leq a$	-0.30	-0.42	-0.15	-0.22

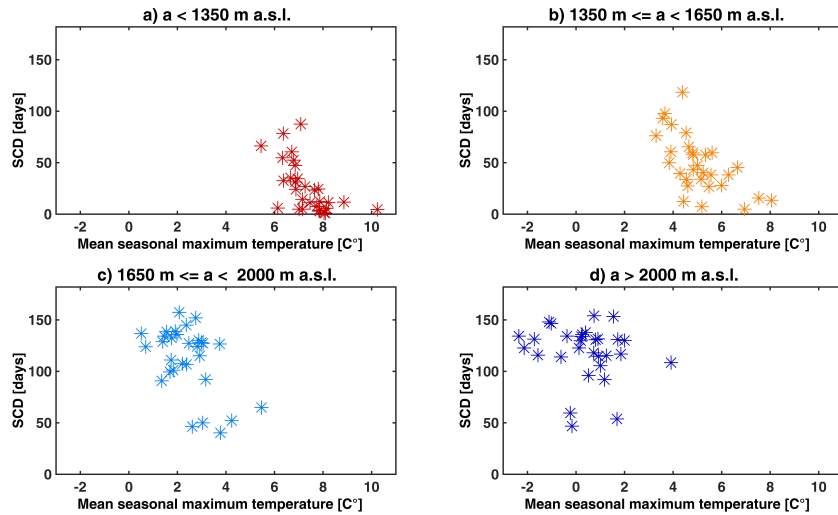


Figure 4.6: Correlation between the mean seasonal maximum temperature and the snow cover duration.

### 4.3.2 Wavelet transform

We analyzed the average of the snow depth time series for the four elevation classes defined above. Figure 4.7 shows the wavelet spectrum of the four time series pertaining to each elevation band between the periods of 2 and 10 years to focus on the characteristics of the signal beyond the annual periodicity. The black contours are the 5% significance regions, using a red-noise background spectrum. Warmer colors represent high power (i.e., high positive correlation with the wavelet), while cold colors represent low power (i.e., high negative correlation with the wavelet) and areas affected by edge effects, within the cone of influence, are fogged. By looking at the regions with the wavelet power spectrum of large power, it is possible to determine the most relevant features of the signal. Low elevation stations, i.e. those at elevations smaller than 1650 m a.s.l., show high power fluctuations at the period of two years from 1983 to 1992, with some spreading to the period of four years during the last 5 years of this interval (Figures 4.7a and 4.7b). The power then reduces, to increase again after 2001. The average behavior for stations above 1650 m a.s.l. (Figure 4.7c-d) show high power at the 2-4 years periods only in the interval from 1987 to 1992 then it decreases, to increase again

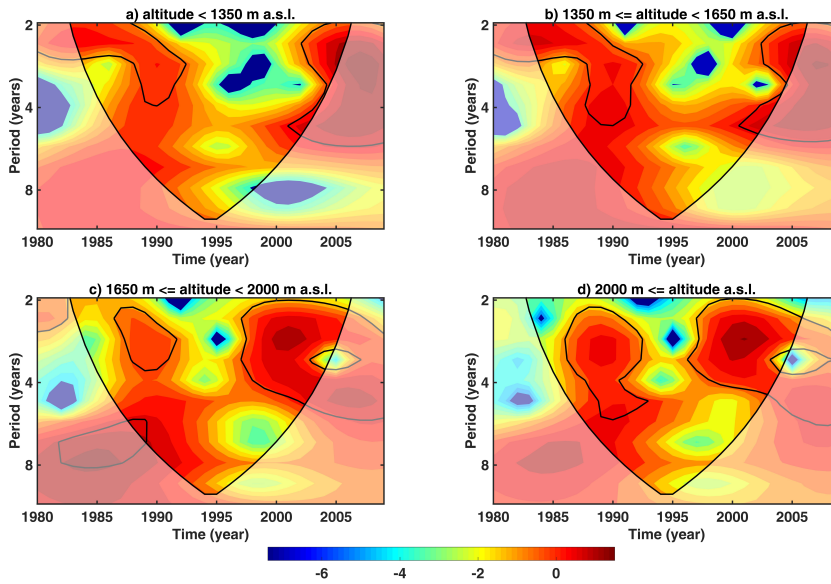


Figure 4.7: Wavelet spectrum of the average mean seasonal snow depth for different altitude classes.

in 1996, hence earlier than for stations at low elevations ( $\leq 1650$  m a.s.l. ).

Figure 4.8 shows the wavelet power spectrum for the snow cover time series analysis. While for the two classes representative for low elevation sites (below 1650 m a.s.l.) the outcomes are similar to those obtained for the mean seasonal snow depth, a difference can be appreciated for the two elevation classes representative of high elevation sites (above 1650 m a.s.l.). For the highest elevation classes, the high power region between the 2 and 4 years periods is almost continuous between 1984 and 2005.

These results show that the snow scarce period observed in the in the late '80s and '90s can be identified at the 2-4 year scale for stations below 1650 m a.s.l., both considering HS as well as SCD. On the contrary, stations at elevations above 1650 m a.s.l. display a distinct pattern. The decrease in the wavelet power spectrum at the 2-4 year scale of HS occurs for a shorter period of time than for low elevation stations. Moreover, the analysis of the SCD shows a much weaker reduction of the wavelet power spectrum than HS for the same sites and SCD for low elevation sites. In particular, snow cover duration at high elevation sites shows a greater resilience

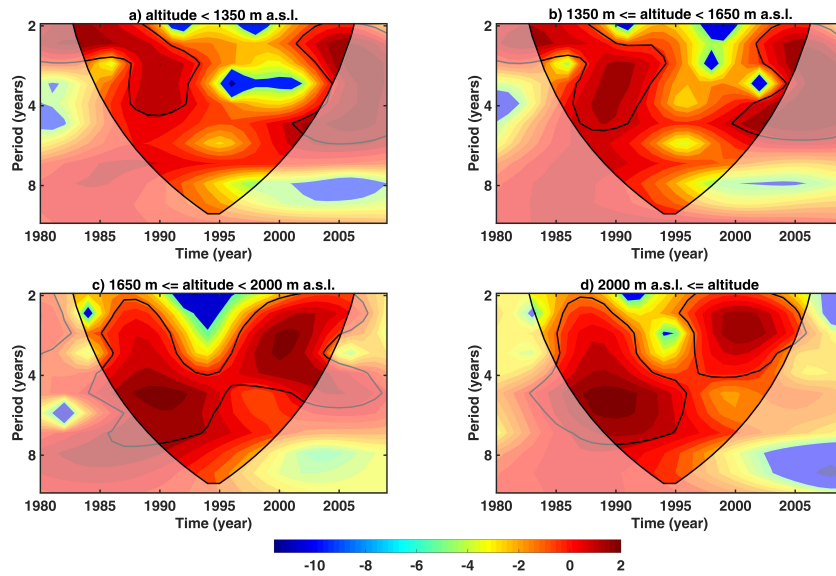


Figure 4.8: Wavelet spectrum of the average snow cover for different altitude classes.

towards an increase in temperature. This analysis has two main implications. The former is that snow time series for low and high elevation sites are not only different at the yearly scale, but they also have remarkable differences at larger scales indicating a different response towards climatic changes. The latter is that the analysis of HS and SCD bear different and complementary information and are worth being analyzed as two distinct variables. This can be explained considering the following extreme situations. In a snow rich year, with high HS values, the SCD can still not exceed the maximum value of 182 days, since this value represent a snow depth larger than 30 cm from November 1st to April 30th (see Section 4.3.1). In a dry year, we may have a single snowfall event at the beginning of the season, leading to a low HS value, yet if the temperature is low enough to avoid melting the value of SCD can still be large.

Figure 4.9 shows the global wavelet spectrum for the mean seasonal snow depth of stations belonging to the four classes of elevations identified previously. This quantity represents the average of  $W_n(s)$  (see eq. 4.1) computed for all times and it can be interpreted in a similar way to the Fourier transform. Difference

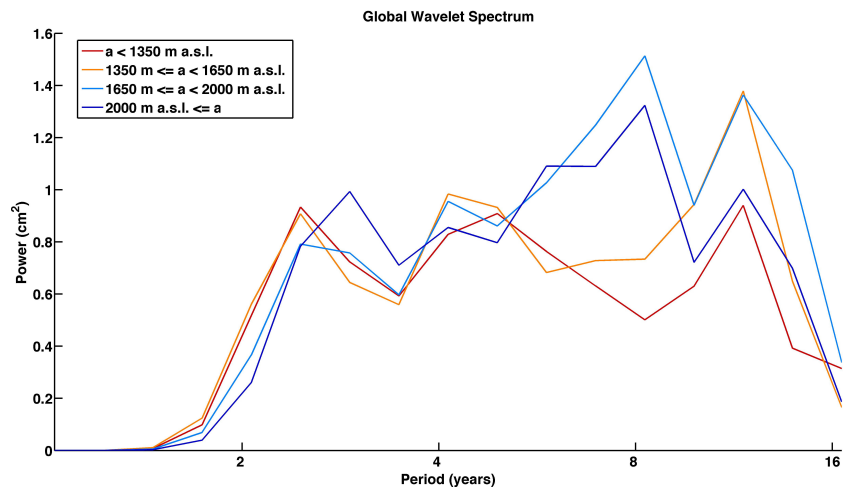


Figure 4.9: Global wavelet spectrum for different altitude classes.

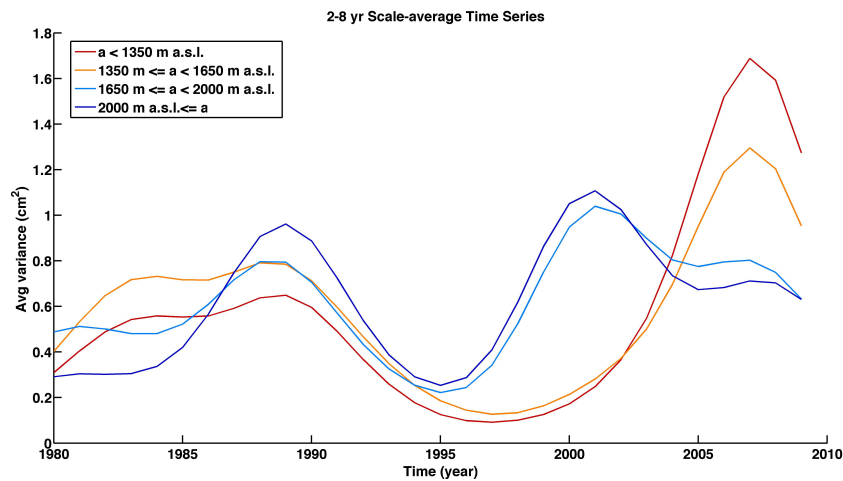


Figure 4.10: Scale average time series between the periods of 2 and 8 years for different altitude classes.

between stations above and below 1650 m a.s.l. are evident in the periods ranging from 6 to 9 years. The global wavelet spectrum of high elevation time series shows a peak at 8 years which is not present in the low elevation time series. Figure 4.10 shows the average of the power spectrum between 2-8 year periods (i.e., the average of  $W_n(s)$  between  $s = 2$  and  $s = 8$  years), which represents the variance of the signal components comprised within this range. Low elevation time series show a first peak in the variance in 1989 followed by a reduction, which leads to a minimum in the late '90s, followed by a strong rise to the highest peak of the time series, which is reached in 2008, followed by a steep reduction in the following years. High elevation time series, show a different behavior with a first peak larger than for low elevations reached in 1989, also followed by a reduction with the minimum reached earlier (in 1995) and a following peak reached earlier (in 2000) and with less intensity than at low elevation.

This analysis confirms some of the conclusions drawn by analyzing the wavelet spectrum. Snow time series at low and high elevation sites display important differences which go beyond the expected seasonal and elevation dependent behavior. They display different dominant frequencies (i.e., they have a different periodic behavior), as evidenced by the global wavelet spectrum. Moreover, the variance of the signal for low and high elevation sites has a different temporal distribution and differs also in magnitude. Overall, fluctuations of the variance of the signal are milder at higher elevations than at low elevation.

### 4.3.3 Wavelet coherence with the NAOI and the MOI

In this section we analyze the correlation between the variations observed for the mean seasonal snow depth and two global indexes: the North Atlantic Oscillation Index (NAOI) and the Mediterranean Oscillation Index (MOI). The North Atlantic Oscillation Index (NAOI) is defined as the normalized pressure difference between Stykkisholmur (Island) and Lisbon (Portugal) [Hurrell, 1995, 1996]. The NAOI has a strong influence on European precipitation patterns. In the southern part of Europe, and in particular in the alpine region, the precipitations are negatively correlated with the NAOI. On the contrary, there is a positive correlation between the NAOI and the Alpine temperature [see e.g. Beniston, 2012b]. The Mediterranean Oscillation Index is defined as the nor-

malized pressure difference between the Gibraltar's Northern Frontier and Lod Airport in Israel [Palutikof, 2003]. The influence of the NAOI and of the MOI fluctuations on the snow dynamic is less obvious, since they do not have an instantaneous and direct effect, but they influence both precipitation and temperature [Beniston, 2012b].

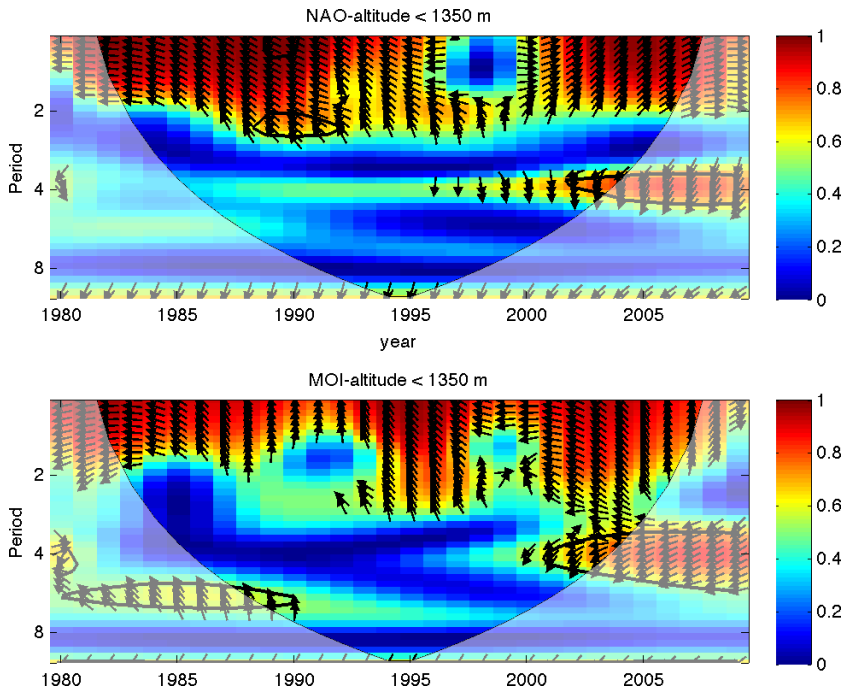


Figure 4.11: Wavelet coherence analysis between the average mean seasonal snow depth of the stations below 1350 m a.s.l. and the NAOI (plot above) and the MOI (plot below).

Figures 4.11-4.14 show the wavelet coherence analysis between the mean seasonal snow depth at different elevations and the two climatic indexes described above. The color indicates the coherence of the two signals for different times and periods (warm colors represent coherence close to 1, while cold colors represent coherence close to 0). The direction of the arrows indicate the relative phase between the signals [Grinsted et al., 2004]. For example, in case the signals are in phase the arrows point to the right, if the signals are in anti-phase the arrows point to the left. The white shaded area shows the cone of influence.

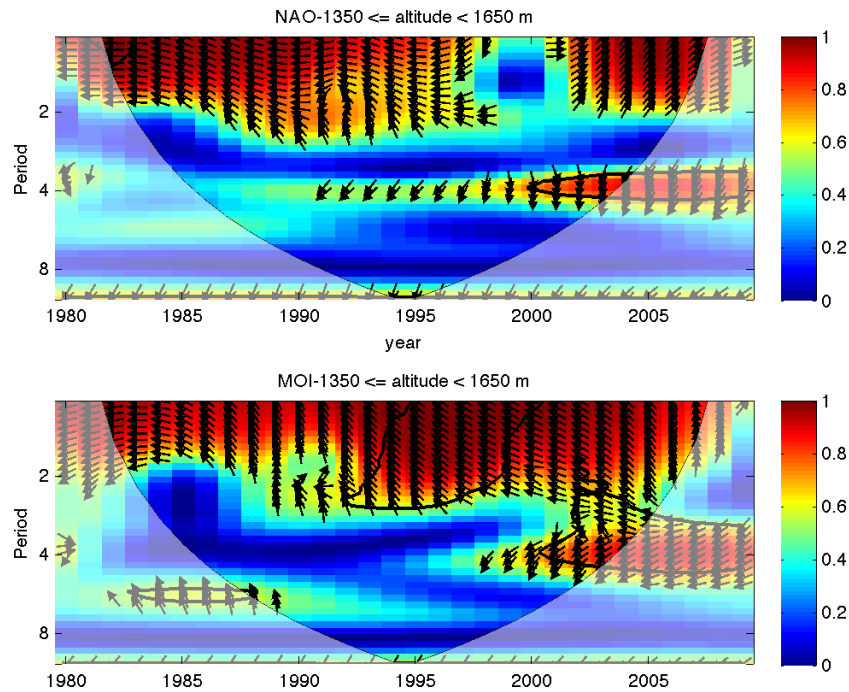


Figure 4.12: Wavelet coherence analysis between the average mean seasonal snow depth of the stations between 1350 m and 1650 m a.s.l. and the NAOI (plot above) and the MOI (plot below).

We observe that the two lowest classes have a similar behavior which differs from the one of the two highest classes. For all elevation classes, the coherence with the NAOI is strong up to 2 year period, but for the stations below 1650 m a.s.l we observe a period at the end of the '90s where there was no coherence between the NAOI and the mean seasonal snow depth at these scales (see Figures 4.11 and 4.12). Except for the 2 years period, no strong correlation with the NAOI is observed for the stations between 1650 m and 2000 m a.s.l., as shown in Figure 4.13. On the contrary, Figure 4.14 shows a period with strong coherence at about 6 years period for about 10 years starting from the middle of the '80s. Finally, Figures 4.11 and 4.12 show a strong coherence at 4 years period starting from 2000 for the lowest stations and starting from the '90s for the stations between 1350 and 1650 m a.s.l..

The relationship between the NAOI and the mean seasonal snow depth in the Adige River Basin does not change in accor-



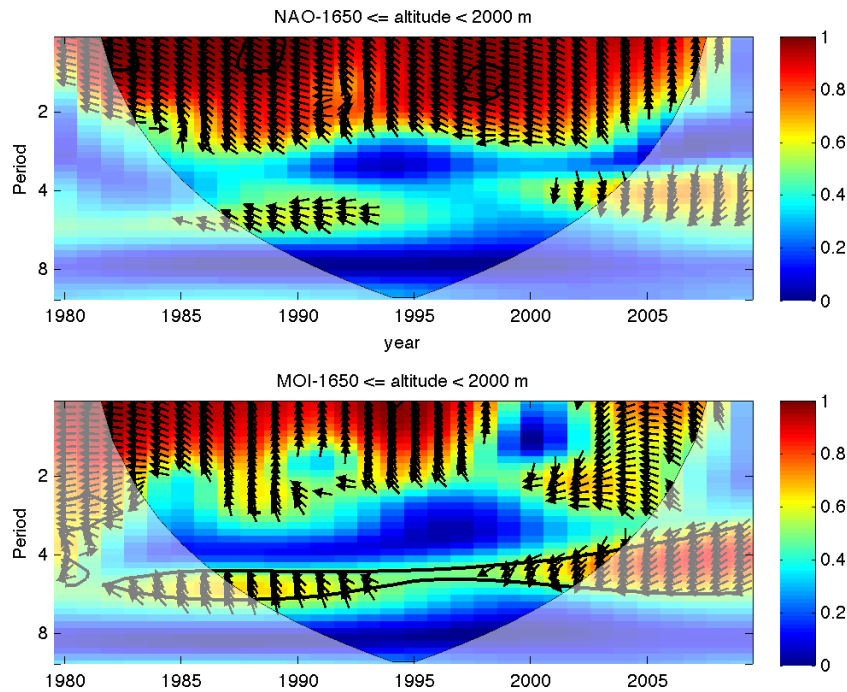


Figure 4.13: Wavelet coherence analysis between the average mean seasonal snow depth of the stations between 1650 m and 2000 m a.s.l. and the NAOI (plot above) and the MOI (plot below).

dance with the changes observed in the behavior of the snow in the last decades, as observed in section 4.3.1. The coherence between snow depth and NAOI signals performed in this study shows that the two time series are generally poorly correlated. Therefore, the correlation between snow signals, such as snow cover duration and snow depth, with the NAOI observed in the Swiss Alps [e.g., Beniston, 1997] is not necessarily representative for the entire Alpine region, as also evidenced for example by Durand et al. [2009] and Schnier et al. [2009]. Therefore, while the correlation between the snow signal and the NAOI is strongly heterogeneous throughout the Alps and in general in the Mediterranean Region [López-Moreno et al., 2011, Luterbacher et al., 2006], the snow-scarce period observed in the '90s is present both in the Northern as well as in the Southern European Alps. We conclude therefore that this behavior has to be mainly attributed to a forcing factor which is common to both sides of the mountain chain, such as

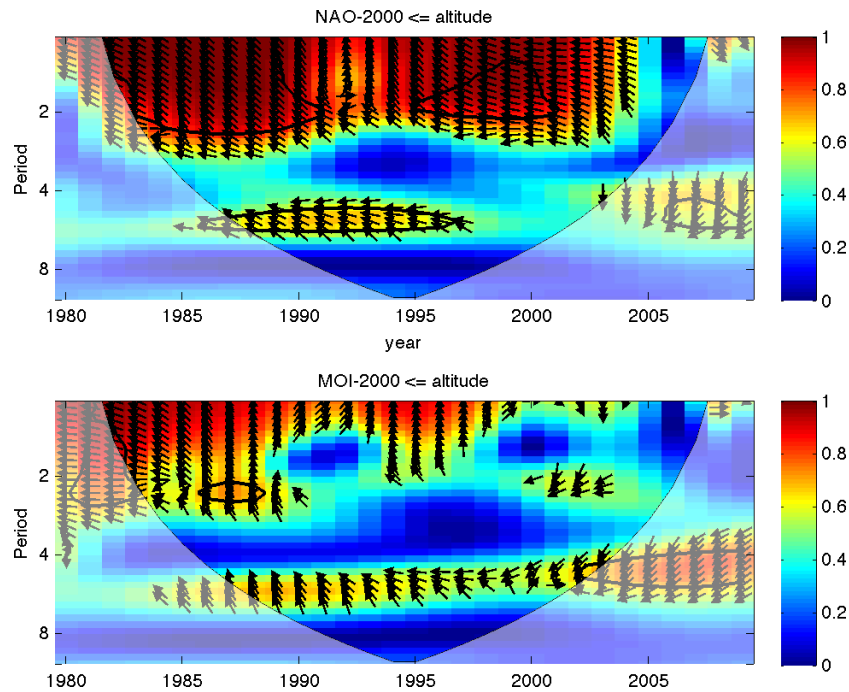


Figure 4.14: Wavelet coherence analysis between the average mean seasonal snow depth of the stations below 2000 m a.s.l. and the NAOI (plot above) and the MOI (plot below).

the observed increase in temperature, that has an affect both on snow cover duration and on phase change in precipitation regime, particularly relevant for sites at lower altitudes [Serquet et al., 2011].

If we consider instead the coherence between the mean seasonal snow depth and the MOI, we find that the highest stations (above 1650 m a.s.l.) show a significant coherence between 6 and 4 year periods for all the time span we are observing, even if it is slightly less strong during the '90s (see Figures 4.13 and 4.14). The lower stations (Figures 4.11 and 4.12) show instead a strong coherence before the '90s at a 6 year period and after 2000 at 4 year period. The arrows pointing to the left indicate that the 6-4 year scale component of the MOI signal is generally decreasing when the 6-4 year scale component of the HS signal is increasing. This anti-phase correlation is interrupted for low elevation sites and weakened for high elevation sites from the late '80s to the end of

the '90s.

This behavior can be linked with the observations performed in section 4.3.1, where we observed a general decrease in HS in the same time frame. We consider remarkable in particular the elevation dependent reduction in the coherence signal at the scale of 6-4 years and the corresponding elevation dependent decrease in HS in the same period. From this study, we can hypothesize a correlation between periods where HS and the MOI display a low coherence at the scale of 6 years and the reduction in HS. Of course, the temporal extension of the time series is limited and therefore this conclusion should be supported by the analysis of longer time series. As we discussed in the introduction to this chapter, the influence of MOI on atmospheric circulation patterns in the Mediterranean area are well known [see e.g. Piervitali et al., 1999, Brunetti et al., 2002, Harding et al., 2009] although their impact on snowfall occurrence in the Alps is not yet completely understood [see e.g. Palutikof, 2003]. Moreover, this analysis confirms again the different behavior between snow depth time series measured at low and high elevation sites. In particular, the low frequency signals composing the time series are differently correlated with climatic indexes. We can therefore expect a different response of low and high elevation sites towards ongoing climatic changes, being the low elevation sites less resilient.

## 4.4 Application to discharge data

In this section we show the application of the wavelet analysis for the study of the discharge time series of the Inn and Adige river basins presented in section 2.4. This analysis was part of the Master's thesis of Teresa Pérez Ciria [Pérez Ciria, 2016], which I cosupervised as Ph.D. student together with Prof. Dr. Chiogna. The goal of the application of the wavelet analysis to these discharge time series was the identification of long term patterns in these alpine catchments. In particular, we aimed at investigating the link between the discharge time series and climate indexes such as the MOI and the NAOI. We applied the wavelet analysis to the yearly average time series (1st January - 31st December).

#### 4.4.1 Continuous wavelet transform

Figure 4.15 compares the global wavelet spectrum and Figure 4.16 compares the 2-8 years scale-average computed for the discharge data of the analyzed gauging stations. The gauging stations are divided according to the catchments (left column Adige, right column Inn) and the drainage area of the stations decreases from the top to the bottom of the figures.

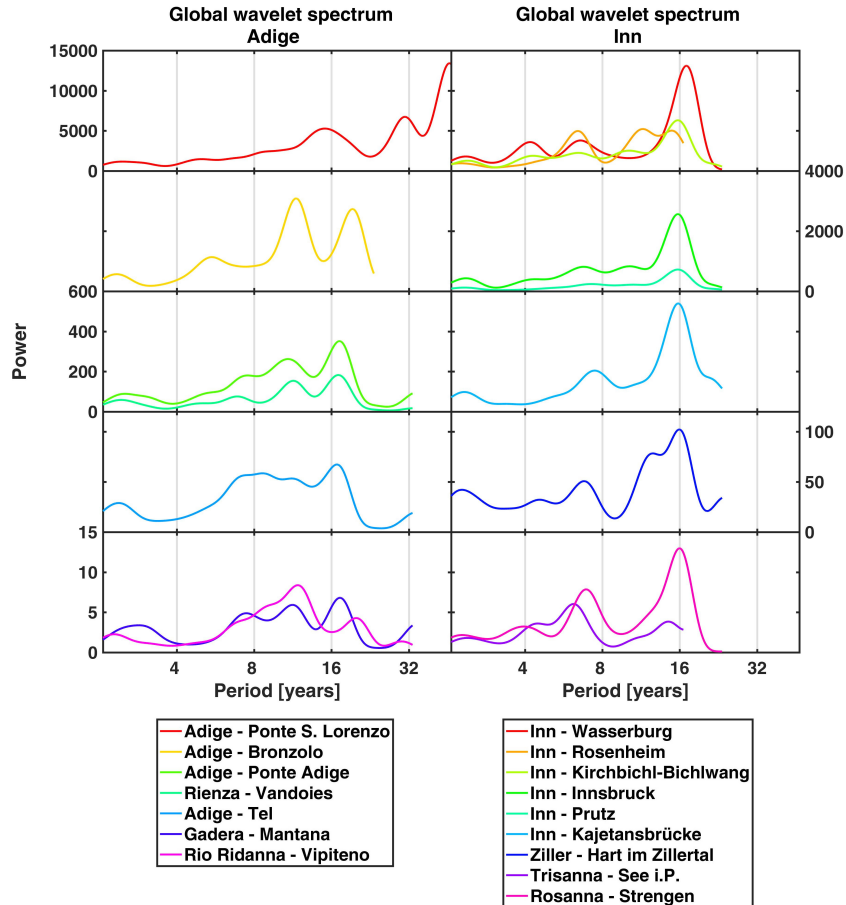


Figure 4.15: Comparison of the global wavelet spectrum for the gauging stations of the Adige (left column) and for the Inn (right column) catchment.

The application of the continuous wavelet transform to the time series relative to the Adige catchment highlighted the pres-

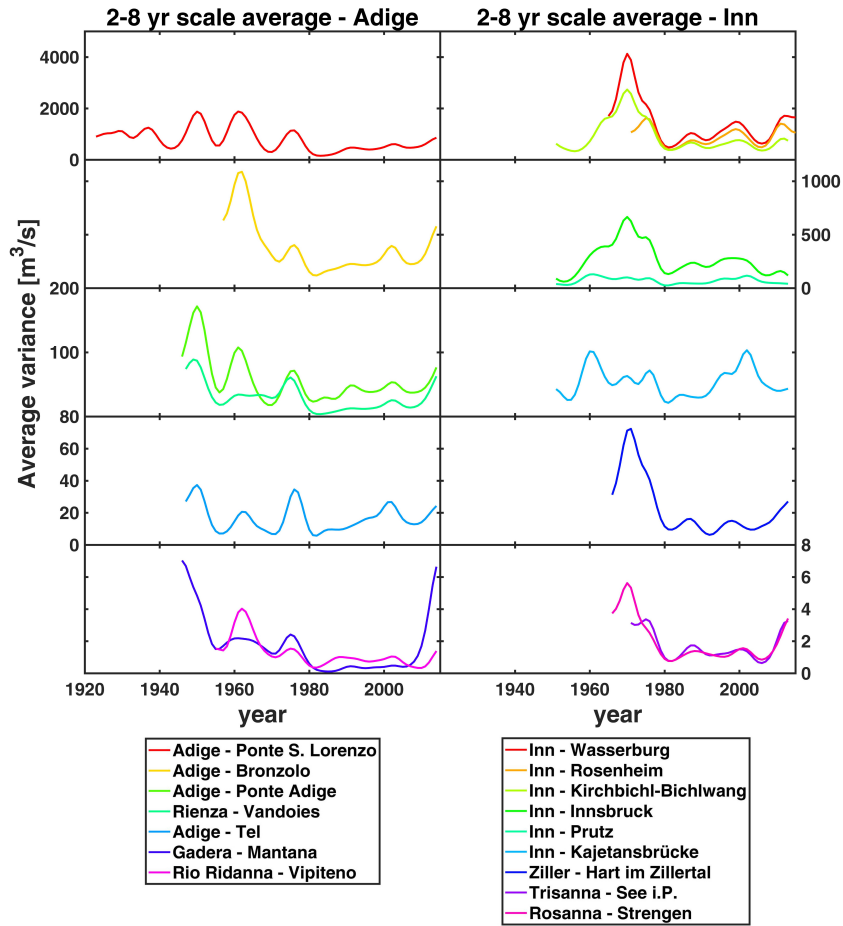


Figure 4.16: Comparison of the 2-8 years scale-average for the stations of the Adige (left column) and for the Inn (right column) catchment.

ence in most of the stations of a 12-years variability signal (Figure 4.15). This peak can not be observed for the station of Adige - Ponte S. Lorenzo and Adite - Tel. Another strong signal can be observed around a longer period of 16 years for the stations located in the northern part of the catchment Adige-Tel, Gadera-Mantana, Rienza-Vandoies and Ponte Adige. The stations of Adige-Bronzolo and Rio Ridanna-Vipiteno, show a peak at a period of about 19 years. Finally, the stations of Adige - Ponte S. Lorenzo show a peak at 15 years period. Unfortunately, the reduced length of the

time series makes the observations at these larger periods less reliable than for shorter periods. An exception is represented by the station of Adige - Ponte S. Lorenzo, because a large part of the interested area is outside the cone of influence.

Also for the gauging stations of the Inn river catchment it is possible to identify a 16-years variability, but, as already observed for the stations of the Adige river catchment, at this period the impact of the cone of influence is such that the conclusions are affected by a high degree of uncertainty. For the stations of the Inn river catchment we observe a peak also at periods of 6-8 years. These are generally smaller than the ones for which we noticed the first peak in the Adige catchment (12 years).

Our results are consistent with those presented in Labat [2008] for the largest European rivers. In particular, it is worth observing that the Inn catchment and the northern gauging stations of the Adige catchment (Adige-Tel, Gadera-Mantana, Rienza-Vandoies and Ponte Adige) display a closer similarity, while the southern stations of the Adige catchment present different main frequencies. This is indicative of the difference in the driving forces controlling streamflow generation in the two catchments.

In Figure 4.16 the 2-8 year scale-average of the stations of the Adige and of the Inn catchments are compared. We decided to analyze the scale average in this interval because it excludes the yearly periodicity and it is not deeply influenced by the cone of influence. Moreover, as outlined by Labat [2008], the main variability in term of variance remain the 2-8 year fluctuations while 8-15 year and 20-30 year variability only account for a less percentage of the global variability but can be explicative for multidecadal trends. We notice two main peaks, common to most of the stations in the Adige river basin around 1962 and 1975. A smaller peak can be identified around 2002 for all stations except Gadera - Mantana. Finally, a peak can be observed around 1950 for all the stations with data available for this year, even if for the station of Gadera - Mantana it is slightly anticipated.

Observing the global wavelet spectrum of the time series of the Inn river basin, we notice a peak common to all Inn stations around 1970. The stations of Inn - Rosenheim, Inn - Prutz, Inn - Kahetansbrücke and Trisanna - See i.P. also show a peak around 1975, as already observed for most of the stations of the Adige river basin. Most of the stations also show a peak around 1988 and around 2000. This last peak can again be linked to the one

observed in for the Adige river basin around 2002.

For all stations of the Adige catchment we observe a period with a low 2-8 years scale average during the '80s and the '90s. This behavior can also be observed for most of the stations in the Inn catchment. The two bigger exceptions are the stations of Inn - Prutz and Inn - Kajetansbruecke, which show a strong increase in the 2-8 years scale average during the '90s, after a lower period during the '80s. Due to the importance of hydrological snow processes, this period of low variability in the variance of the signal at the 2-8 year scales can be qualitatively associated with the snow scarce period discussed in the previous section. However, further analysis are necessary to quantitatively address the correlation between the two observations. The analysis of the 2-8 year scale-average, representing the variability of the time series in terms of variance, shows similar results for gauging stations belonging to the same catchment, while the Adige and the Inn catchment discharge time series can be differentiated. An exception are the time series located in the northern part of the Adige catchment, since they display a pattern more similar to the stations of the Inn than of the Adige catchment.

Concluding, the wavelet analysis of discharge time series, does not only allow to identify statistically significant bands of intermittent fluctuations at different scales [Labat, 2008] but it can be also used as a measure of similarity between measured discharge at different gauging stations. We suggest that the comparison between the global wavelet spectrum and the wavelet scale average of discharge time series can be applied to identify the transition zone between different climatic patterns controlling hydrological processes in the Alpine region.

#### 4.4.2 Wavelet coherence analysis

River discharge is influenced by a series of factors such as precipitation distribution and intensity, temperature, land use and management of water resources for human uses. Climate changes therefore are not the only driver influencing streamflow. However, it is important to distinguish their specific influence, since they are difficult to control and predict. Indeed, climate change can influence factors such as temperature, precipitation and snow melting not only changing the amount, but also the temporal and spatial distribution of these variables Beniston [2006]. The co-

herence analysis, introduced in section 4.2.1, investigates the correlation of two different signals. In this section we explore the discharge in relation with two climate indexes, which we already introduced in section 4.3.3: the North Atlantic Oscillation Index and the Mediterranean Oscillation Index.

### North Atlantic Oscillation Index

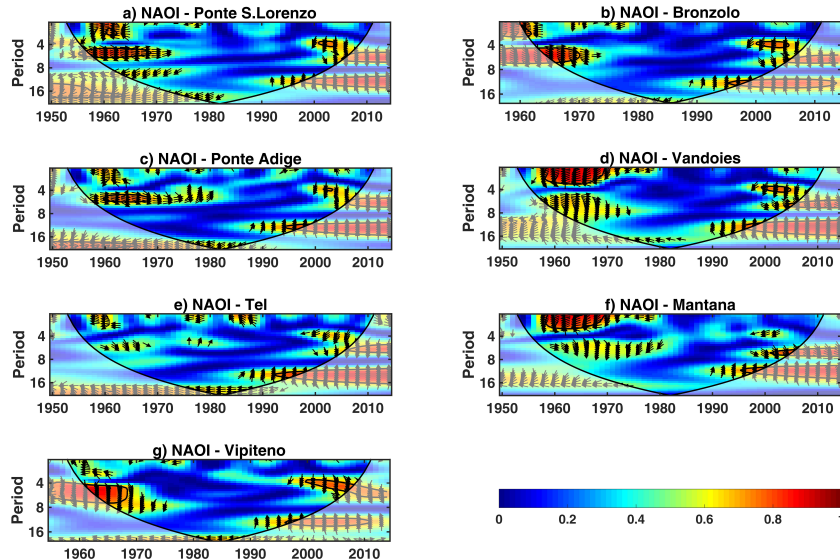


Figure 4.17: Wavelet coherence analysis with the NAOI for the gauging stations of the Adige river catchment. The abscissa axis indicates the year.

The NAOI is correlated positively with the temperatures and negatively with the precipitation in the Alpine region, especially during winter [see e.g. Quadrelli et al., 2001, Schmidli et al., 2002, Bogataj, 2007]. This climatic index describes how changes in pressure systems in the Atlantic affects winter conditions in many parts of the Northern Hemisphere [Hurrell and Van Loon, 1997, Marshall et al., 2001]. As described by Labat [2005], when a strong low pressure is centered near Iceland and a strong high pressure is located over the middle of the North Atlantic Ocean, then in Europe the winter is warmer and wetter. When the area of high and low pressure is weaker, winter in the Mediterranean is rainy while in northern Europe is cold and dry.



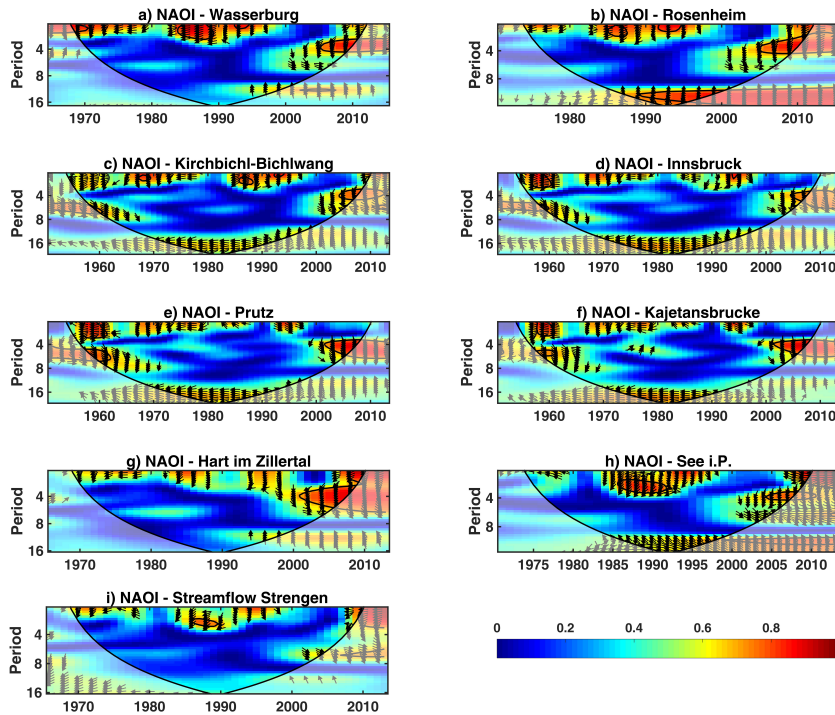


Figure 4.18: Wavelet coherence analysis with the NAOI for the gauging stations of the Inn river catchment. The abscissa axis indicates the year.

The coherence analysis between the yearly discharge in the Adige river basin and the NAOI does not generally show a strong coherence (see Figure 4.17). For the gauging stations of Adige - Ponte S. Lorenzo, Adige - Bronzolo, Adige - Ponte Adige and Rio Ridanna - Vipiteno we observe a higher coherence at a period of about 4-6 years before 1970. All stations of the Adige river basin show a high coherence at the same scale after 2000. Rienza - Vandoies and Gadera - Mantana show high coherence at period shorter than 4 years before 1970. Moreover, strong correlation can also be observed for all stations at a period between 12 and 16 years after 1990, and more significantly after 1995. Unfortunately these observations are based on results which are located in the cone of influence. Nevertheless, since these results are common to all time series in the Adige river basin, it was anyway worth noting it.

The coherence analysis for the Inn river gauging stations gave similar results to what observed in the Adige river basin (see Figure 4.18). Again, we observe some correlation between 4 and 6 years period starting from 2000 and at a period smaller than 4 years until 1995. The stations Inn - Kirchbichl-Bichlwang, Inn-Innsbruck, Inn - Prutz and Inn - Kajetansbrücke also show high coherence at a period of about 4-6 years before 1970. For the stations of Inn - Rosenheim, Inn - Kirchbichl-Bichlwang, Inn-Innsbruck, Inn - Prutz and Inn - Kajetansbrücke and Trisanna - See i.P. we observe a correlation at 16-years period all along the time span considered. Unfortunately, also in this case part of the results are in the cone of influence and can be considered relevant only in relation to their persistence for all the considered stations inside the river basin.

Besides the observations for each basin, it is interesting noting similar patterns between the two catchments in the correlation with the NAOI. The results presented in this section has highlighted for both river basins high coherence of the discharge of most of the stations with the NAOI at a period of 4-6 years before 1970 and after 2000. Moreover, it is worth noting the high coherence at a period of about 16 years noticeable for both river basins, even if in a different interval, since for the Adige river basin this is started after 1990. Unfortunately, this last observation was influenced by the cone of influence. We can notice that the arrows pointing to the left indicate that when the NAOI and the discharge signals are correlated the correlation is negative, in agreement with the behavior of the NAOI and the precipitations over Europe. The results also outline the high non-stationary and complexity in the correlation between this climatic index and river discharge, indicating that the strength of the correlation between the two signals varies over time. The available gauging stations exhibit a temporal correlation with the NAOI sometimes on the entire interval (e.g., 16 year scale) but more often on restricted intervals.

### **Mediterranean Oscillation Index**

The MOI has been indicated by several authors [see e.g. Piervitali et al., 1999, Palutikof, 2003] as one of the main drivers of precipitation patterns in the Mediterranean basin and therefore we aim at testing how changes in this precipitation driver may in-

fluence river discharge. Pressure differences between the western and eastern ends of the basin may influence climate variability at inter-annual scales [Harding et al., 2009]. Favorable circulation for high temperatures in the western part of the basin is associated with unfavorable circulations in the eastern part of the basin.

As shown in Figure 4.19 and in Figure 4.20, the coherence of the discharge with the MOI is more differentiated between the two basins than what observed for the correlation with the NAOI.

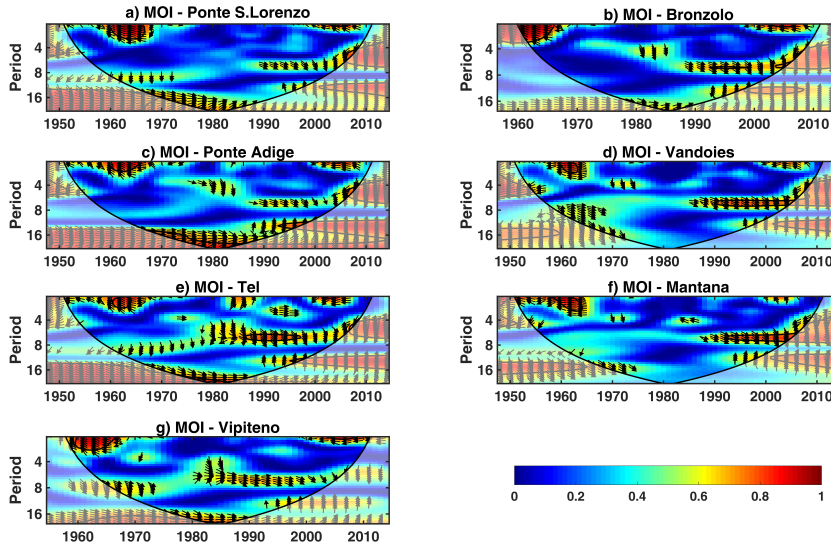


Figure 4.19: Wavelet coherence analysis with the MOI for the gauging stations of the Adige river catchment. The abscissa axis indicates the year.

The MOI shows a correlation with the discharge in the Adige river basin mainly at periods of 6-8 years starting from 1990 (see Figure 4.19). Even if this pattern develops also in the cone of influence, it starts inside the reliable area and it is relevant for this reason. It is particularly evident for the stations of Adige - Bronzolo, Rienza - Vandoies, Adige - Tel and Gadera - Mantana. Before 1970 at a period of 8-10 years we can also observe some correlation between the discharge and the MOI for the stations of Adige - Ponte S. Lorenzo, Rienza - Vandoies, Adige - Tel and Rio Ridanna - Vipiteno, even if it is not as strong as the one mentioned before. An interesting aspect is that for these gauging stations in the interval between 1970 and the end of the '80s only

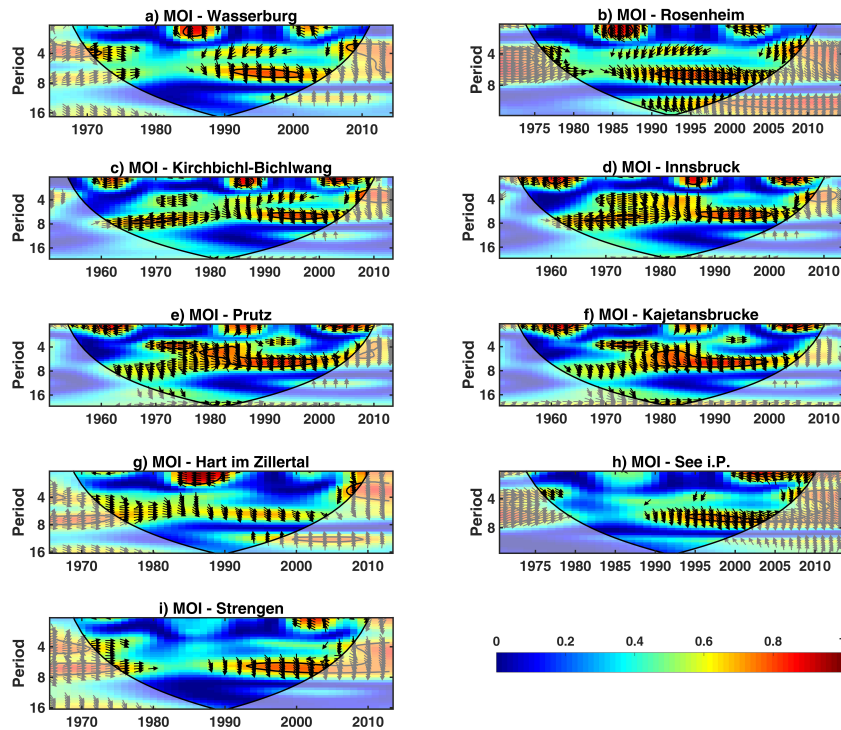


Figure 4.20: Wavelet coherence analysis with the MOI for the gauging stations of the Inn river catchment. The abscissa axis indicates the year.

a weak correlation or even no correlation is observed between the discharge and the MOI. It is also worth noting the time lag between 1960 and 1970 at a period smaller than 4 years where a strong correlation between the two signals can be observed for all gauging stations.

The most evident correlation between the discharge in the Inn river basin and the MOI is found between 1960 and 2000 (which are also the limits of the cone of influence) at a 6-8 year periods, for the stations Inn - Kirchbichl-Bichlwang, Inn - Innsbruck, Inn - Prutz and Inn - Kajetansbrücke (see Figure 4.20). The station of Ziller - Hart im Zillertal shows high coherence at 6-8 years period between 1970 and 2000. Inn - Wasserburg, Inn - Rosenheim, Trisanna - See i.P. and Rosanna - Strengen show high coherence at the same period, but starting from the second half of the '80s or from 1990. High coherence can also be observed for Inn - Wasserburg, Inn -

Rosenheim and Ziller - Hart im Zillertal at period smaller than 4 years between year 1985 and 1990.

A common pattern which can be individuated in the correlation of the discharge in the examined catchments and the MOI is the high coherence at 6-8 periods. It is interesting to notice, that these are the same periods at which we observed high coherence of the snow depth in the Adige river basin with the MOI during the '80s. For most of the analyzed gauging stations this pattern starts at the end of the '80s and beginning of the '90s, while for four stations in the Inn river basin this is present in the whole interval outside the cone of influence (between 1960 and 2000). Also for this climatic index the periods in which the signals are coherent are localized in very specific time windows. The coherence wavelet analysis in this case is useful to identify specific periods in which two signals becomes correlated. This output can be used to distinguish between the behavior of different catchments, while a physical interpretation of the processes leading to this localized coherence is more complex and further research is necessary as outlined also by Labat [2010], since river discharge is not only influenced by climatic changes, but also by anthropogenic activities (e.g., land use change and water management).

## 4.5 Conclusions

In this chapter we performed the wavelet analysis of the datasets introduced in sections 2.3 and 2.4. In particular, the snow depth dataset of the Adige river basin was also analyzed with other statistical techniques, as shown in section 4.3.1.

The analysis of the snow depth dataset of the Adige river basin has shown a pronounced difference in the behavior of the stations below and above 1650 m a.s.l., as well as a different response to other climate forcing, such as temperature, or climate indexes, such as the MOI. In particular we observed for all stations a decrease in the mean seasonal snow depth as well as in the snow cover duration during the '90s, which was more important for the stations below 1650 m a.s.l.. A possible recovery seems to be started from the first years after 2000, but this findings will need to be further confirmed with future analysis with longer and more updated time series. Also the correlation of the mean seasonal snow depth and snow cover duration with the temperature is different for the

stations below and above 1650 m a.s.l.. In fact, especially for the snow cover duration, the values at high elevations do not show a strong correlation with the temperature. The coherence wavelet analysis of the mean seasonal snow depth with the NAOI and with the MOI revealed different patterns of coherence with the two climate indexes. In particular, it is important to notice that the high coherence observed between the mean seasonal snow depth and the MOI at 6-8 year periods, its interruption or decrease during the '90s and its recovery after 2000 at shorter period of 4 years is in accordance with the period of low mean seasonal snow depth observed during the '90s, which is also in accordance with other findings in similar analysis about the alpine area [see e.g. Marty, 2008]. The coherence wavelet analysis in this case provides a useful tool to identify at which scale we can detect changes occurring in a time series and to differentiate between the behavior of stations at low and high elevation in correlation with climatic indexes. Notice that changes in low frequency components of a signal would be very difficult to be identified with other time series analysis methods, yet they play an important role for the understanding of the climate dynamics [Torrence and Compo, 1998].

The wavelet analysis of the discharge in the Adige and Inn river basins allows to identify similarities and differences in the hydrological regime of the two catchments. The two study areas are close to each other and have similar characteristics (since in this analysis we considered only the upper part of the Inn river basin). We observed an interval with low 2-8 year scales average for almost all considered stations of both catchments during the '80s and the '90s. Nonetheless, the stations belonging to the Inn catchment and the northern part of the Adige catchment have a similar variability in the variance at the 2-8 year scales and in the dominant frequencies which differ from the one of the southern part of the Adige catchment. Similar patterns between the stations of the Adige and the Inn river catchments were identified in the coherence analysis with climatic indexes. We noticed an interval of high coherence with the NAOI before 1970 and after 2000 at 4-6 year periods. The interval with the strongest coherence with the MOI was instead the one after the end of the '80s at 6-8 years periods. The wavelet in this case allow to clearly identify the non-stationary behavior in the coherence between climatic indexes and river discharge. Wavelet analysis constitute a useful tool to investigate highly non-stationary river annual runoffs and

---

allows in general to connect long-term hydrological variability to climatic forcing factors [Labat, 2008]. It allows to distinguish between the response of different catchments to the same climatic drivers (i.e., different catchments display different coherence with the MOI or the NAOI). As an important drawback, it does not allow an easy interpretation of the causal effects leading to a change in the coherence between streamflow and climatic signals. This can be explained by the fact that river discharge is highly disturbed by anthropogenic activities which may blur the effects of climatic drivers.





# Chapter 5

## Copula

### 5.1 Introduction

A Copula is a statistical tool used to describe the dependence structure between two or more variables. In this section we briefly introduce this concept. We refer to works like Joe [1997] and Nelsen [1999] for a detailed mathematical description, while the works of Genest and Favre [2007] and Yan et al. [2007] provide a more practical introduction to copulas. Copulas have been widely used in finance [see e.g. Genest et al., 2009] and since a few years they have found a wider and wider application also in hydrology (see e.g. Favre et al. [2004], Salvadori and De Michele [2004], Grimaldi and Serinaldi [2006], Salvadori and De Michele [2007], Serinaldi and Grimaldi [2007], Bárdossy and Li [2008], Grimaldi et al. [2016]).

A copula  $C$  is defined as a multivariate distribution with univariate marginals being uniform in the interval  $[0,1]$ :

$$C : [0, 1]^n \rightarrow [0, 1] \quad (5.1)$$

$$C(\mathbf{u}^{(i)}) = u_i \quad \text{with } \mathbf{u}^{(i)} = (1, \dots, u_i, \dots, 1) \quad (5.2)$$

$$C(\mathbf{u}) = 0 \quad \text{if } \mathbf{u} = (u_1, \dots, u_n) \quad \exists u_i = 0 \quad (5.3)$$

We also have

$$\sum_{j=0}^{2^n-1} (-1)^{n-\sum_{i=1}^n j_i} C(u_1 + j_1 \Delta_1, \dots, u_n + j_n \Delta_n) \geq 0 \quad (5.4)$$

where  $0 \leq u_i \leq u_i + j_i \Delta_i \leq 1$ ,  $i = \sum_{k=0}^{n-1} j_k 2^k$  [Nelsen, 1999, 2003, Bárdossy and Li, 2008].

**Example** The bivariate copula is defined as a bivariate distribution with univariate marginals being uniform on  $[0,1]$ . For the bivariate copula we have

$$C : [0, 1]^2 \rightarrow [0, 1]$$

$$C((u_1, 1)) = u_1 \quad C((1, u_2)) = u_2$$

$$C((u_1, 0)) = 0 \quad C((0, u_2)) = 0$$

$$C(u_1 + \Delta_1, u_2 + \Delta_2) - C(u_1, u_2 + \Delta_2) - C(u_1 + \Delta_1, u_2) + C(u_1, u_2) \geq 0$$

where  $0 \leq u_k \leq u_k + \Delta_k \leq 1$  [Nelsen, 1999].

The Sklar's theorem [Sklar, 1959] is a fundamental conception in copula's theory and states that given an  $n$ -dimensional distribution function  $F$  with marginals  $F_1, \dots, F_n$ , there exists a copula  $C$  such that

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)) \quad (5.5)$$

If the marginals are continuous, then  $C$  is unique [Nelsen, 1999, Yan et al., 2007]. A corollary of the Sklar's theorem states that given an  $n$ -dimensional distribution function  $F$  with continuous marginals  $F_1, \dots, F_n$  and the copula  $C$  that satisfies 5.5, then for any  $\mathbf{u} \in [0, 1]^n$

$$C(u_1, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)). \quad (5.6)$$

The Sklar's theorem allows us to understand one of the main advantages of the copulas, i.e. the possibility to independently model the marginals of the single variables and the dependence structure between them. In this way, copulas can be seen as a pure expression of dependency [Bárdossy, 2006].

Given a copula  $C$ , then

$$W(\mathbf{u}) \leq C(\mathbf{u}) \leq M(\mathbf{u}) \quad (5.7)$$

where  $W(\mathbf{u}) = \max(0, u_1 + \dots + u_n - (n - 1))$ ,  $M(\mathbf{u}) = \min_i(u_i)$  and  $\mathbf{u} = (u_1, \dots, u_n)$ .  $W(\mathbf{u})$  and  $M(\mathbf{u})$  are the so called Fréchet-Hoeffding lower and upper bounds, respectively [Joe, 1997, Genest and Favre, 2007].  $W(\mathbf{u})$  is always a copula, while  $M(\mathbf{u})$  is a copula only for  $n = 2$  [Joe, 1997, Nelsen, 2003].

A nice interpretation can be given for the Fréchet-Hoeffding bounds in the case  $n = 2$ . In this case the inequality 5.7 can be written as

$$W(u_1, u_2) \leq C(u_1, u_2) \leq M(u_1, u_2), \quad (5.8)$$

where  $W(u_1, u_2) = \max(0, u_1 + u_2 - 1)$  and  $M(u_1, u_2) = \min(u_1, u_2)$ . The copula  $M(u_1, u_2)$  indicates positive dependence while  $W(u_1, u_2)$  indicates negative dependence. All the other bivariate copulas are between  $W$  and  $M$  [Nelsen, 2003]. Among them we also find the independence copula

$$\Pi(u_1, u_2) = u_1 u_2 \quad (5.9)$$

### 5.1.1 Measures of dependence

As already done in chapter 4, also in this chapter we use as a measure of dependence between two random variables two rank-based measures of dependence: the Kendall's tau  $\tau_k$  in eq. 4.9 and the Spearman's rho  $\rho_S$  in eq. 4.10 [see e.g. Embrechts et al., 2001, Genest and Favre, 2007]. These measures always exist and are based on the ranks of the data. Moreover, they do not depend from the marginals and they are scale-invariant, i.e. they do not change under strictly increasing transformations of the analyzed variables. In some cases, there is an explicit function that expresses the relationship between the copula parameter and  $\tau_k$  and  $\rho_S$ , and hence these properties can be used to fit these copulas to the data. Both  $\tau_k$  and  $\rho_S$  vary between -1 (perfect negative dependence) and 1 (perfect positive dependence). In case of independence, both  $\tau_k$  and  $\rho_S$  are equal to 0, but, on the contrary, being  $\tau_k = 0$  or  $\rho_S = 0$  we can not conclude that the two variables are independent.

### 5.1.2 Tail dependence

In many studies it is important to investigate the amount of dependence in the upper-quadrant and in the lower-quadrant of a bivariate distribution, i.e. the amount of dependence in case of extremely large or extremely small events. The tail dependence is a property of a copula. Upper tail dependence is defined as it follows:

$$\lambda_U = \lim_{u \nearrow 1} P \{Y > F_Y^{-1}(u) | X > F_X^{-1}(u)\}, \quad (5.10)$$

c.d.f.s of  $X$  and  $Y$ , respectively [Joe, 1997, Embrechts et al., 2001]. Analogously, the lower tail dependence is defined as

$$\lambda_L = \lim_{u \searrow 0} P \{Y \leq F_Y^{-1}(u) | X \leq F_X^{-1}(u)\}, \quad (5.11)$$

where again  $F_X$  and  $F_Y$  are the marginals of  $X$  and  $Y$ , respectively [Joe et al., 2010].

If  $F_X$  and  $F_Y$  are continuous c.d.f.s, then the equations of the upper and lower tail dependences can be reformulated as it follows:

$$\lambda_U = \lim_{u \nearrow 1} \frac{\bar{C}(u, u)}{(1-u)}, \quad (5.12)$$

$$\lambda_L = \lim_{u \searrow 0} \frac{C(u, u)}{u}, \quad (5.13)$$

where  $\bar{C}$  is the survival copula of  $C$  [Joe, 1997].

Notice that both  $\lambda_U$  and  $\lambda_L$  are in the interval  $[0, 1]$  [Embrechts et al., 2001]. Intuitively, the upper tail dependence expresses the probability that a random variable  $Y$  exceeds a given threshold, given that the other random variable  $X$  has exceeded that threshold. Analogously, the lower tail dependence expresses the probability that a random variable  $Y$  does not exceed a given threshold, given that the other random variable  $X$  has not exceeded that threshold.

### 5.1.3 Bivariate copula families

**Elliptical copulas** Elliptical copulas are those corresponding to elliptical distributions [Embrechts et al., 2001]. Examples of elliptical copulas are the normal copula (see for example Figure 5.1) and the t copula (see for example Figure 5.2).

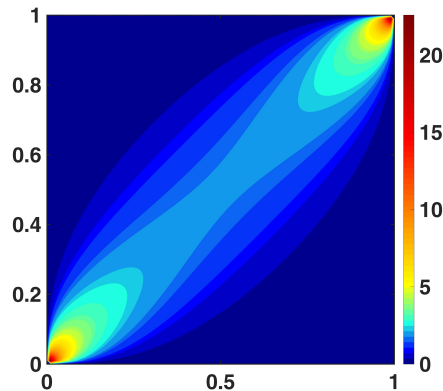


Figure 5.1: Example of density distribution of a normal copula

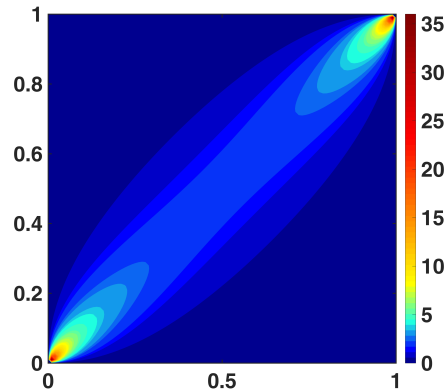


Figure 5.2: Example of density distribution of a t-copula

Elliptical copulas are radially symmetric [Embrechts et al., 2001]. Hence they can not model a different level of dependence for lower or for higher values. A Gaussian copula is used in section 5.2.2 to model the dependence structure between mean seasonal snow depth and temperature for the stations between 1350 m and 1650 m a.s.l. and for the stations between 1650 m and 2000 m a.s.l.

**Archimedean copulas** Archimedean copulas allow the description of a wide range of dependencies and most of the archimedean copulas can be expressed in closed form. A drawback of the archimedean copulas is that it is not straightforward to expand them to more than 2 dimensions [see e.g. Nelsen, 1997, McNeil and Nešlehová, 2009].

Suppose we have a continuous, convex, strictly decreasing function  $\gamma : [0, 1] \rightarrow [0, \infty]$ , with  $\gamma(1) = 0$  and be  $\gamma^{[-1]}$  the pseudoinverse of  $\gamma$  defined as

$$\gamma^{[-1]} = \begin{cases} \gamma^{-1} & \text{if } t \in [0, \gamma(0)] \\ 0 & \text{if } t > \gamma(0) \end{cases} \quad (5.14)$$

then

$$C(u_1, \dots, u_n) = \gamma^{[-1]}(\gamma(u_1) + \dots + \gamma(u_n)) \quad (5.15)$$

is a copula and is called an Archimedean copula [Nelsen, 2003]. We call  $\gamma$  a generator of  $C$ . Some examples of Archimedean cop-

ulas are given in Table 5.1. The only Archimedean copula that is radially symmetric is the Frank copula [Embrechts et al., 2001].

Table 5.1: Examples of Archimedean copulas [Embrechts et al., 2001].

Copula	Generator $\gamma(x)$	$\theta$ range
Clayton	$(x^{-\theta} - 1)/\theta$	$[-1, \infty] \setminus \{0\}$
Gumbel	$(-\ln x)^\theta$	$[1, \infty]$
Frank	$-\ln \left( \frac{\exp(-\theta x) - 1}{\exp(-\theta) - 1} \right)$	$[-\infty, \infty] \setminus \{0\}$

In this work we find an example of application of archimedean copulas in section 5.2.1, where a Clayton copula is used to describe the dependence structure between mean seasonal snow depth and snow cover duration of the stations between 1650 m and 2000 m a.s.l.

#### 5.1.4 Multivariate copula families

Elliptical bivariate copulas have a natural extension to multivariate copulas. This is not the case for the archimedean copulas. Another approach for the creation of multivariate copulas is the use of the vine copulas [see e.g. Bedford and Cooke, 2001, 2002, Aas and Berg, 2009, Aas et al., 2009, Gräler et al., 2013]. The idea of the vine copulas is to create a multivariate copula as a tree structure of bivariate copulas. This idea is based on the observation that we can represent a multivariate density  $f(x_1, \dots, x_d)$  as a product of conditional pair copula densities and marginal densities [see e.g. Czado, 2010]:

$$f(x_1, \dots, x_d) = \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,(i+j)|(i+1),\dots,(i+j-1)} \cdot \prod_{k=1}^d f_k(x_k) \quad (5.16)$$

The structure of a vine copula is not fixed. Two important kind of structures of a vine copula can be the D-vines (Figure 5.4) and C-vines (Figure 5.5). In the D-vine structure the copula structure at each step is a path, i.e. there is no central node connected to all others [Czado, 2010]. In the C-vine structure, instead, at each step there is a unique node that is connected to all other nodes.

Let us suppose we want to construct a vine copula describing the dependence structure between three random variables  $V_1, V_2$

and  $V3$ . In the 3-dimensional case, the structure of the copula is unique, as also shown in Figure 5.3 [Aas et al., 2009].

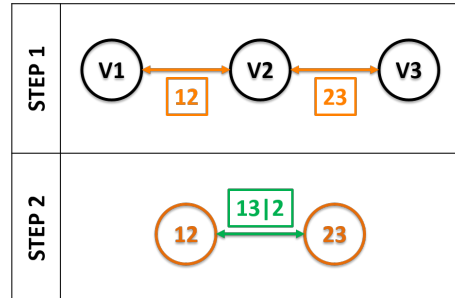


Figure 5.3: Structure of a 3-dimensional vine copula

As illustrated in Figure 5.3, a bivariate copula  $C_{12}$  is built between  $V1$  and  $V2$  and a bivariate copula  $C_{23}$  is built between  $V2$  and  $V3$ . A third copula  $C_{13|2}$  is then built at the last level. Given three variables, between which we want to build a vine copula, their order is a free choice. In Figure 5.3 is  $V1$ ,  $V2$  and  $V3$ , but could have also been another one, such as  $V2$ ,  $V1$  and  $V3$ . The order in which the variables are taken is going to influence the result. In Gräler et al. [2013] it is suggested to model at first the couples of variables, which are best correlated.

In a 4-dimensional case, the structure of a C-vine and of a D-vine do not coincide no more. We can than have the two structures reported in Figure 5.4 and 5.5.

The main advantage of vine copulas is the possibility to combine a wide number of copulas, including also those, such as for example Archimedean copulas, which can hardly be generalized to more than 2 dimensions. This gives the possibility to model a wide variety of dependence structures and asymmetries [Czado, 2010].

Vine copulas are applied in this work in sections 5.3.2 and 5.3.3 for the description of flood events.

### 5.1.5 Empirical copula

Let us suppose we want to investigate the copula underlying some multivariate observations  $\mathbf{X}_1, \dots, \mathbf{X}_n$  from a random vector  $\mathbf{X} = (X^1, \dots, X^d)$  with continuous marginals. The first step is to test that  $\mathbf{X}_1, \dots, \mathbf{X}_n$  are mutually independent, i.e. that they can be

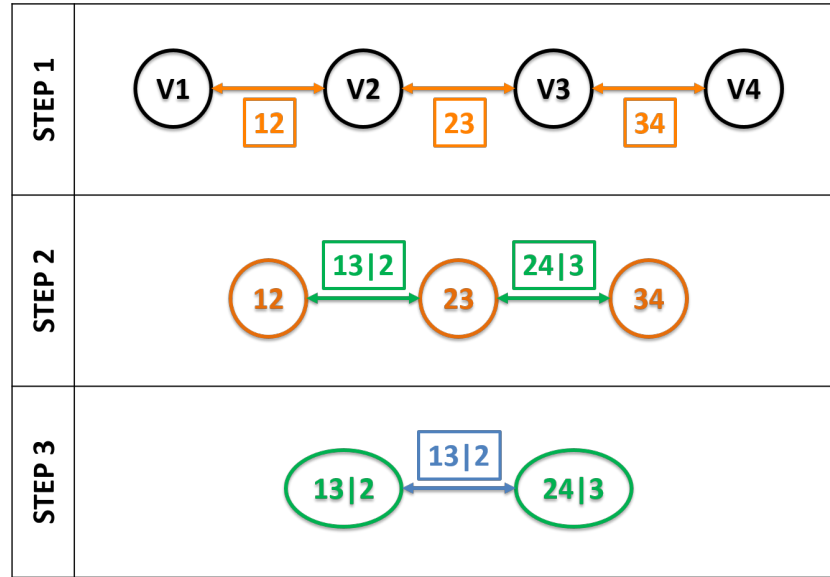


Figure 5.4: Example of possible structures of a D-vine copula.

regarded as a random sample from  $F(\mathbf{x}) = C[F_1(x_1), \dots, F_d(x_d)]$ , where  $C$  is the copula underlying the multivariate data and  $F_1, \dots, F_d$  are the marginals of  $\mathbf{X}^1, \dots, \mathbf{X}^d$ . The following step is to test the data for the presence of ties. If ties are present there are two possible solutions. The first option is to exclude some of the data, in order to eliminate the ties. This is however not recommended if the time series is short. Another possibility is to add a uniform noise, that makes the data slightly varying inside the measurement uncertainty [Salvadori et al., 2014].

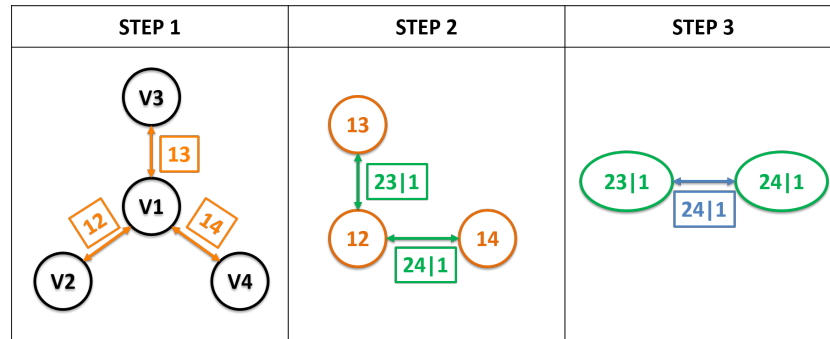


Figure 5.5: Examples of possible structures of a C-vine copula.



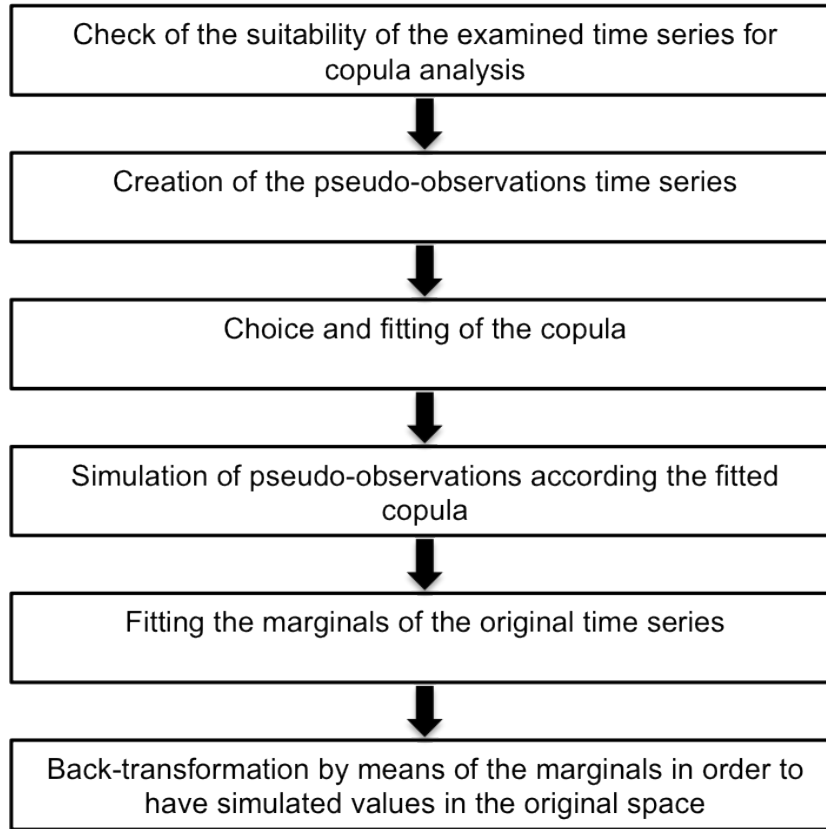


Figure 5.6: Scheme of the procedure applied for the application of the copula procedure and the generation of simulated data.

Once that these two preliminary step have been performed we can construct the so called empirical copula, following the procedure described in Bücher et al. [2016]. Let  $\hat{U}_{ij} = R_{ij}/n$  for  $i = 1, \dots, n$  and  $j = 1, \dots, d$ , where  $R_{ij}$  is the rank of  $X_{ij}$ . Then the pseudoobservations of the underlying copula  $C$  and corresponding to the observations  $X$  are defined as  $\hat{\mathbf{U}}$ , where  $\hat{\mathbf{U}}_i = (\hat{U}_{i1}, \dots, \hat{U}_{id})$ . We then define the empirical cumulative distribution function  $\hat{F}$  of the empirical copula  $\hat{C}$  as

$$\hat{F}_j(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}(\mathbf{U}_i \leq \mathbf{u}) \quad (5.17)$$

where  $\mathbf{u} \in [0, 1]^d$ .

The following step is testing the goodness of fit of several cop-

ulas to the pseudoobservations. The copula families that are not rejected are then fitted to the data, mainly with the method of moments or maximizing the pseudo-likelihood [see e.g. Genest and Rivest, 1993, Genest et al., 1995, Joe, 1997].

The question of which is the best copula fitting to the data rarely has a unique answer. Several measures can be taken into account in order to chose the best copula, such as the maximum likelihood and the Akaike information criterion, defined as  $AIC = 2k - 2\ln(L)$ , where  $k$  is the number of parameters and  $L$  is the maximized value of the likelihood function of the model. Another check that can be performed is the comparison of the isolines of the cumulative distribution function of the fitted and of the empirical copula.

Once a copula has been chosen and fitted, we can generate random values from it. The marginals of the generated data will be the uniform distribution. In order to obtain simulations in the original space, we first have to identify the marginals of the original observations and then we have to apply a back-transformation based on the fitted marginals.

The procedure described in this section is summarized in Figure 5.6.

### 5.1.6 Software

The analysis presented in this chapter were performed mainly in Matlab and R [R Core Team, 2015] with the aid of the package *copula* [Yan et al., 2007, Kojadinovic and Yan, 2010, Hofert and Mächler, 2011, Hofert et al., 2017] and the package *VineCopula* [Schepsmeier et al., 2016]. Some of the graphical representations were done with the aid of the package *latticeExtra* [Sarkar and Andrews, 2016].

## 5.2 Analyzing snow dynamics with copula

In chapter 4 we showed the statistical analysis of mean seasonal snow depth and snow cover duration data in the Adige river basin. We observed the variations of these variables in the last decades and analyzed how mean seasonal snow depth and snow cover duration are linked to climatic indexes, such as the MOI and the NAOI, and to temperature. In this section we analyze the relationship between these variables with the aid of the copula analysis. It

is indeed interesting to show how this kind of analysis can reflect the physical characteristics we observed with other statistical techniques.

### 5.2.1 Mean seasonal snow depth and snow cover duration

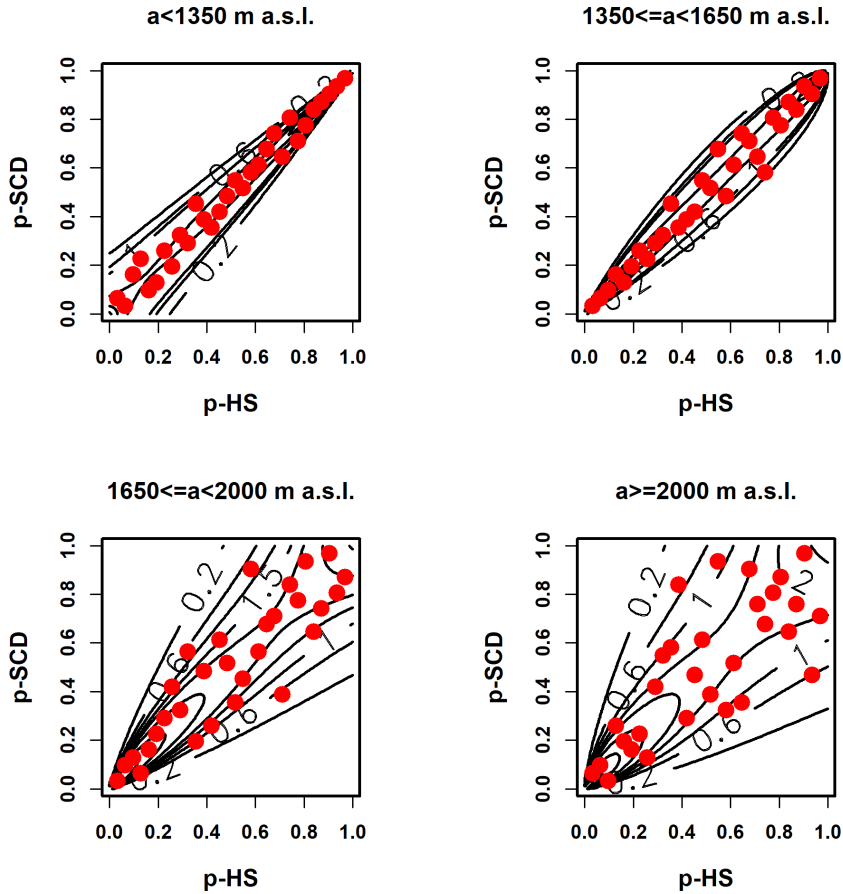


Figure 5.7: Comparison of the pseudoobservations of mean seasonal snow depth and snow cover duration (red dots) with the isolines of the fitted copula (black lines).

We first analyze the correlation between the mean seasonal snow depth and the snow cover duration as defined in section 4.3.1. The copula analysis shows that HS and SCD are very well correlated below 1650 m a.s.l. (first row of Figure 5.7), with  $\tau_k =$

0.92 and  $\rho_S = 0.99$  for the lowest class and  $\tau_k = 0.90$  and  $\rho_S = 0.98$  for the stations between 1350 m and 1650 m a.s.l.. For the higher stations (above 1650 m a.s.l., see second row of Figure 5.7), the correlation, still present, loses strength, in particular for the largest values and for stations above 2000 m a.s.l.. The correlation coefficients are in fact  $\tau_k = 0.73$  and  $\rho_S = 0.88$  for the stations between 1650 m and 2000 m a.s.l. and  $\tau_k = 0.57$  and  $\rho_S = 0.76$  for the stations of the highest class. This can be explained by the fact that at high elevations the snow cover often lasts after the 30th April. The classical definition of SCD is bounded to a maximum value of 182 days, therefore SCD is not very sensitive to snow rich winters at high elevations. Moreover, at higher elevations the lower temperature allows for a longer snow pack duration than at lower warmer sites, even if the mean seasonal snow depth may not be not particularly high.

The copula in this section were chosen on the basis of the Akaike criterion. For the lower stations the Joe Copula with a parameter of 16.47 was chosen. The Joe copula displays an upper tail dependence, which means that for this class of sites the probability of observing a winter with high HS and high SCD is high. On the contrary, if we have a low HS value the SCD can display a large degree of variability, although it will be characterized by small values. This means that it is more challenging to model snow processes at low elevation sites for snow scarce winters. For the second class of stations (between 1350 m and 1650 m a.s.l.) the survival Gumbel with a parameter of 9.99 was chosen. The survival Gumbel copula displays a lower tail dependence, which means that for this class of sites low values of HS are strongly correlated with low values of SCD. Starting from this elevation, therefore, it is possible to observe the transition from the behavior characteristic of the lowest elevation sites and the one of the highest elevation sites that will be described in the following. The high values of the parameters of these copulas are in accordance with the high correlation values reported before, since the Joe Copula and the Gumbel copula express always increasing correlation for increasing values of the parameters. The values of the stations between 1650 m and 2000 m a.s.l. are approximated with a Clayton copula with a parameter of 4.33. The same copula was used for the values of the highest stations, but with a lower parameter of 2.61. The copulas used in these two last cases are also indicative of the properties qualitatively observed in Figure 5.7. In fact the

Clayton copula does not have upper tail dependence. Moreover, the Clayton copula tends to the independence copula for values of the parameter tending to 0, while it tends to the perfect dependence for values of the parameter going to infinite. The parameter of the Clayton copula was already small for the stations between 1650 m and 2000 m a.s.l.. The fact that it reduces for the stations above 2000 m a.s.l. reflects the loss of dependency observed before. Therefore for high elevation sites, we have a high probability of observing a reduced SCD if also HS is small, while for high values of HS, the SCD is highly variable. These results are consistent with the fact that SCD is sensitive to snow rich winters however it is also sensitive to temperature as it will be shown in the following section.

### 5.2.2 Mean seasonal snow depth and temperature

Let us now observe the correlation between temperature and mean seasonal snow depth at different altitudes (Figure 5.8). We always observe a negative correlation, as already observed in chapter 4 and in particular in Table 4.3. We find a higher correlation below 2000 m a.s.l. In fact, as shown in Table 4.3, the Kendall's correlation coefficients for the three lower classes are -0.52, -0.46 and -0.46, respectively. For the three lowest classes, the dependence structure between the mean seasonal snow depth is always approximated with radially symmetrical copula, indicating the same amount of dependence in the upper as well as in the lower quadrant: Frank copula with a parameter of -6.46 for the lowest class, a Gaussian copula with a parameter of -0.7 and of -0.69 for the two remaining classes respectively. The negative values of the parameters describe the negative dependence, while their relatively large absolute values describe the observed dependency strength. Above 2000 m a.s.l., the relationship between the mean seasonal snow depth and temperature is approximated by a rotated ( $270^\circ$ ) Joe copula with a negative parameter (indicating negative correlation), but with an absolute value quite small (-1.76). The Joe copula tends to the independence copula for values of the parameter close to 0. This is an other indicator of the weak correlation of the mean seasonal snow depth and the temperature at high elevations. Therefore, above 2000 m a.s.l. temperature is generally low enough to guarantee snow accumulation during the winter season, even during warm winters.

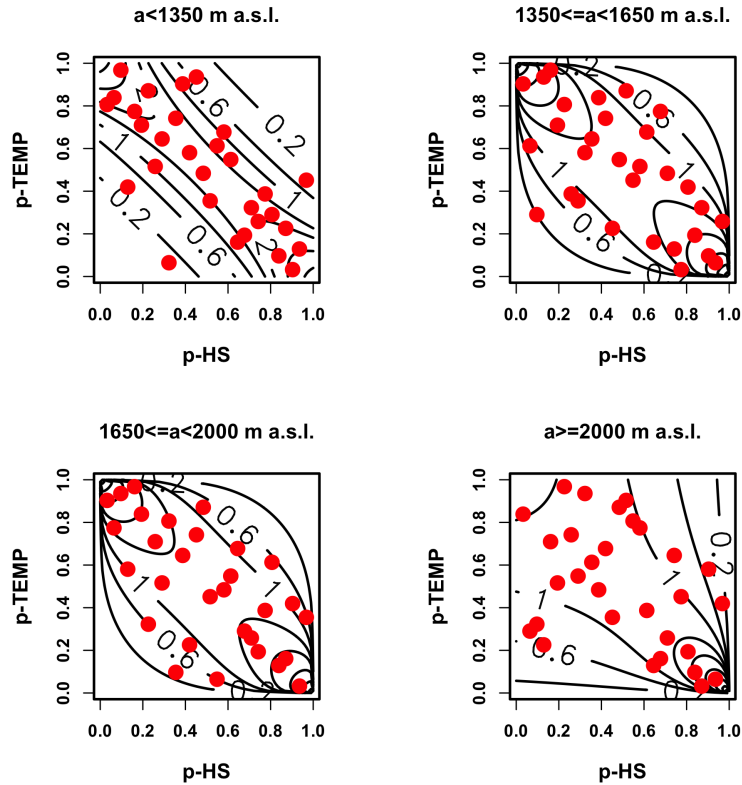


Figure 5.8: Comparison of the pseudoobservations of mean seasonal snow depth and temperature (red dots) with the isolines of the fitted copula (black lines).

### 5.2.3 Snow cover duration and temperature

The correlation between temperature and snow cover duration show a remarkable change below and above 1650 m a.s.l. (see Figure 5.9 and Table 4.3). For the lower stations we still find a good negative correlation, as observed for the correlation between mean seasonal snow depth and temperature, with a Kendall's correlation coefficient of  $-0.54$  and  $-0.46$  for the stations below 1350 m a.s.l. and between 1350 m and 1650 m a.s.l., respectively. The data of the stations above 1650 m a.s.l. indicate a clear loss of dependency. The values of the Kendall's correlation coefficient are smaller than those observed for the lower stations, varying between  $-0.24$  (stations between 1650 m and 2000 m a.s.l.) and  $-0.15$

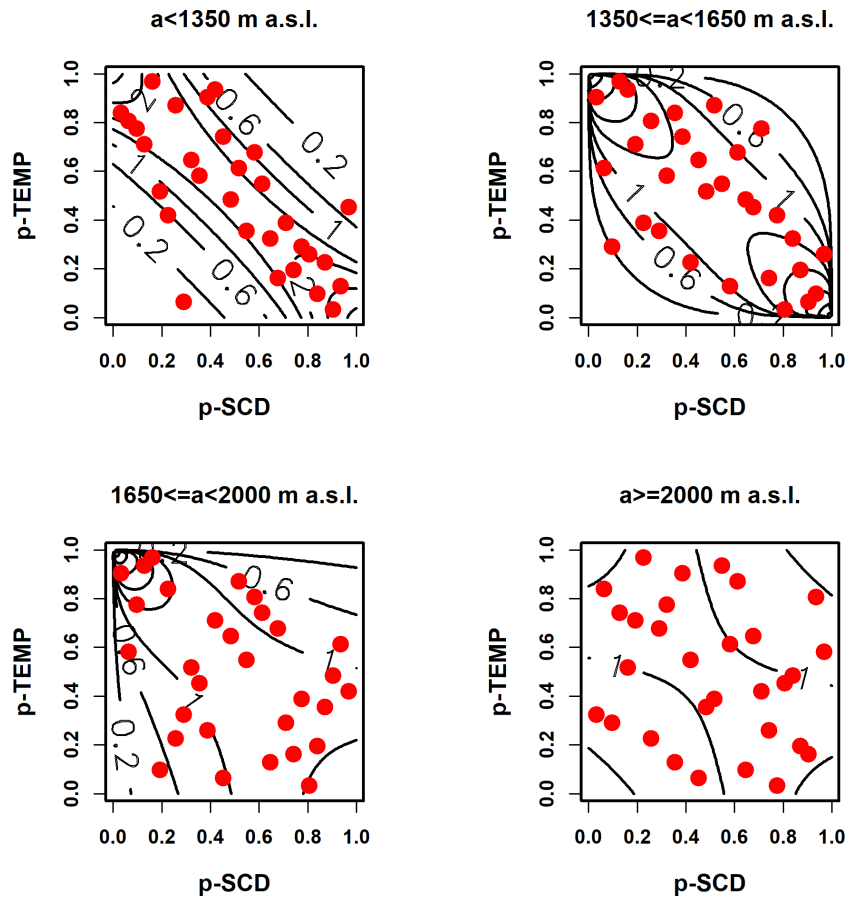


Figure 5.9: Comparison of the pseudoobservations of snow cover duration and temperature (red dots) with the isolines of the fitted copula (black lines).

(stations above 2000 m a.s.l.). The dependence structure between SCD and temperature below 1650 m a.s.l. is described, similarly to what observed for mean seasonal snow depth and temperature, with two radially symmetric copulas: a Frank copula with a parameter of  $-6.38$  and a Gaussian copula with a parameter of  $-0.69$  for the lowest class and the stations between 1350 m and 1650 m a.s.l. respectively. For the two highest classes a rotated ( $270^\circ$ ) Clayton Copula with a parameter of  $-0.65$  and a Frank Copula with a parameter of  $-1.38$  were chosen, respectively. The small parameters make both copula tending to the independence copula, which means that in general at high elevation sites the tempera-

ture is low enough to sustain the persistence of the snow pack also during warmer winters.

#### 5.2.4 Mean seasonal snow depth and MOI

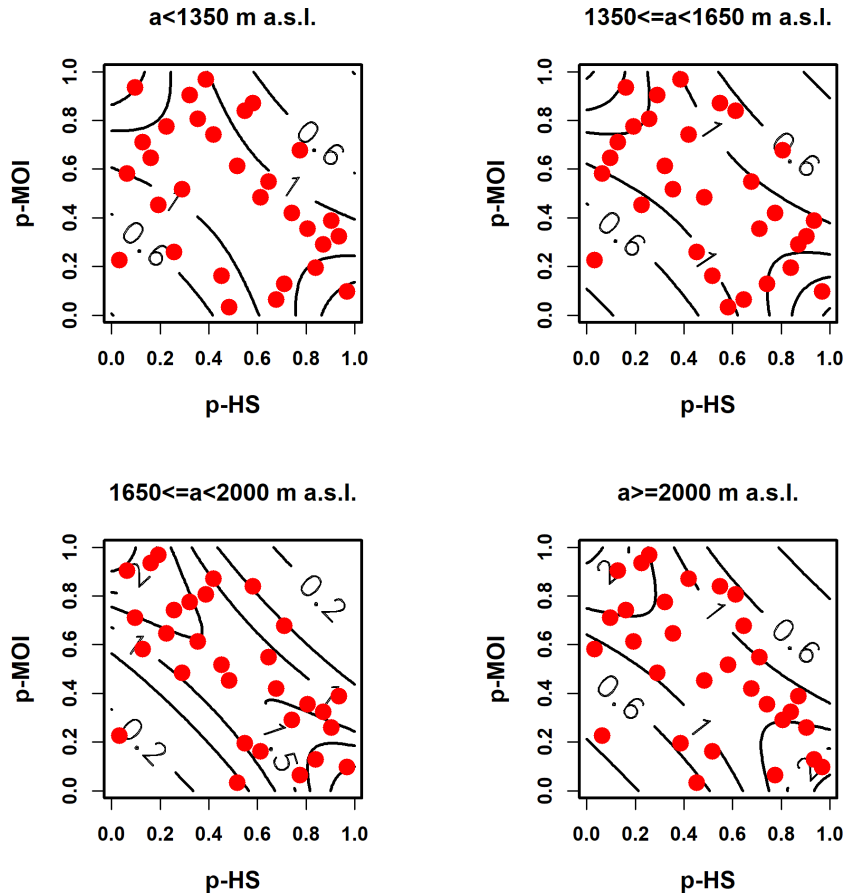


Figure 5.10: Comparison of the pseudoobservations of mean seasonal snow depth and MOI (red dots) with the isolines of the fitted copula (black lines).

In chapter 4 we observed a different coherence between the MOI and the mean seasonal snow depth at different altitudes. In particular, we observed that for stations below 1650 m a.s.l. there was no correlation with the MOI during the '90s. Observing the copulas used to describe the relationship between these data in Figure 5.10, we observe a negative correlation, which is slightly



stronger for the stations of the two highest classes. Here the  $\tau_k$  is equal -0.43 and to -0.39 for the stations between 1650 m and 2000 m a.s.l. and for the stations above 2000 m a.s.l., respectively. For the stations below 1350 m a.s.l., instead,  $\tau_k$  is -0.28 and it is equal to -0.31 for the stations between 1350 m and 1650 m a.s.l. The results shown in Figure 5.10 were obtained always with a Frank copula, whose parameter has taken the value -2.69 for the stations below 1350 m a.s.l., the value -3.14 for the stations between 1350 m and 1650 m a.s.l., the value -4.78 for the stations between 1650 m and 2000 m a.s.l. and -3.75 for the highest stations above 2000 m a.s.l. We remind that a Frank copula with negative parameter describes a negative correlation and that values of the parameter tending to 0 let the Frank Copula tend to the independence copula. Hence, the variations of these parameters as well as their negative values reflects again the properties observed in Figure 5.10 and described by the values of  $\tau_k$  mentioned before. In fact, the lowest classe, where  $\tau_k$  reached the smallest value, is also described by the Frank copula with the smallest parameter. Further, the Frank copula is radially symmetric, which means that it has the dependence between large values of the MOI and small values of the mean seasonal snow depth and the dependence between large values of the mean seasonal snow depth and small value of the MOI is the same. Moreover, this copula has no tail dependence, which means that the dependence between these two variables is not dependent on the magnitude of the variables. In other words, we do not expect more correlation between large values of mean seasonal snow depth and small values of the MOI, or, viceversa, large values of MOI and small values of the mean seasonal snow depth.

Notice that copulas can not highlight the changes in the correlation between the mean seasonal snow depth and the MOI in the examined period, as we have observed for the wavelet analysis in chapter 4. The different results given by the two methods of analysis is indicative of their complementarity. In fact, the coherence wavelet analysis decomposes the time series in several signals with different frequencies, while copula analysis considers the time series as a whole. Hence, copula and wavelet coherence analysis can highlight different aspects of the relation between two variables. In our case, the correlation between the MOI and the mean seasonal snow depth was mainly due to the signal with large periodicity (between 4 and 6 years). For this reason, the copula

analysis could not highlight the same correlation aspects between these two variables.

It could have been interesting to analyze the dependence structure between the mean seasonal snow depth and the MOI removing the years in which we observed a reduction in the coherence between the two signals. Unfortunately, this would have significantly reduced the number of the data, and hence negatively influenced the reliability of the results.

The analysis of the mean seasonal snow depth in correlation with the NAOI does not show any kind of correlation.

### 5.3 Modeling flood events with copula

In this chapter we present an application of the copula for modeling the flood events registered at two different gauging stations in the German part of the Inn River: Wasserburg and Passau. This modeling activity was part of the project *Retentionspotential-Studie am Inn*, which was founded by the *Bayerische Landesamt für Umwelt* (Bavarian Environment Agency - Germany) and by the *Ministerium für ein lebenswertes Österreich* (Austria). Besides the similarities of the discharge at the two measuring points, it is interesting to compare the results obtained with the same methodology at these two different gauging stations for their different characteristics. In fact Wasserburg is one of the first gauging stations on the German part of the Inn river. Passau is the last gauging station of the Inn river, almost at its confluence with the Danube river. Depending on the goal of the analysis, it can be interesting modeling different variables related to the flood events. In this section, when we talk about year, we mean the hydrological year starting on the 1st November and ending on the 31st October. In section 5.3.1 we show the modeling of the direct peak discharge - direct volume dependence structure, taking the *MHQ* of the gauging stations, defined as the average value of the yearly maximum discharge time series, as threshold value. In sections 5.3.2 and 5.3.3 we fit 4-dimensional vine copulas for the description of the dependence structure of the direct peak discharge, direct volume, rising time and a base line, which in section 5.3.2 is the base flow and in section 5.3.3 is an optimized threshold value. Goal of this analysis was to investigate the performance of the copula for the description of the dependence structure between variables

describing the flood events. We also wanted to investigate how the definition of the variables can influence their correlation. This is in fact important in order to evaluate which variables should be modeled with the copula analysis and how they could be used in further studies. In particular, in this project, the results of the simulations of the analysis involving only the volume and the discharge of the event respect to MHQ will be used for the comparison of the performances of our model with a physically based forecasting flood model. The results of the analysis involving also the rising time and either baseflow or the moving threshold will be important for the evaluations of the volume of water that could be saved in retention basins in case of flood events.

### 5.3.1 Bivariate application: direct peak discharge - direct volume modeling

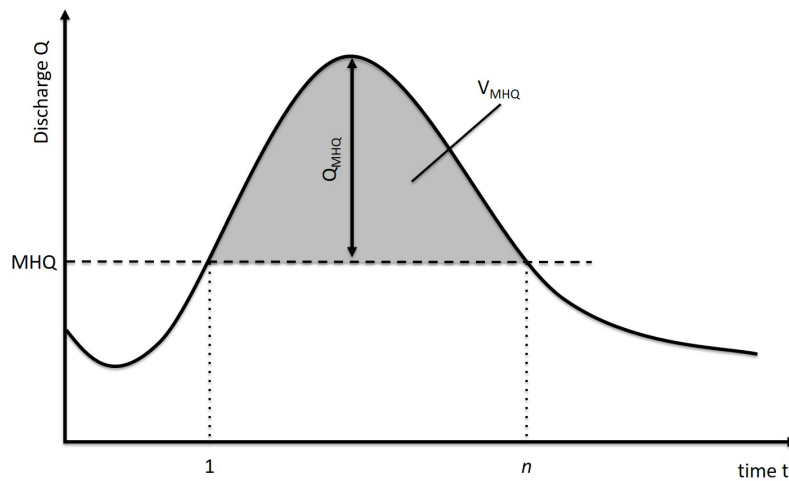


Figure 5.11: Example of the determination of  $Q_{MHQ}$  and  $V_{MHQ}$ .

For each gauging station, the time series of the largest flood event of each year was taken. In this section, the two analyzed variables represented in Figure 5.11 are defined as it follows :

$Q_{MHQ}$  is defined as the difference between the event peak discharge  $Q_{max}$  and  $MHQ$ , also

$$Q_{MHQ} = Q_{max} - MHQ \quad (5.18)$$

$V_{MHQ}$  is defined as the volume of the discharge wave above  $MHQ$ , also

$$V_{MHQ} = \sum_{i=0}^n (Q_i - MHQ) \cdot \Delta t \quad (5.19)$$

where  $Q_i$  is the discharge at timestep  $i$ ,  $n$  is the length of the wave above  $MHQ$  and  $\Delta t$  is the measurement time interval.

The flood events with a peak discharge smaller than  $MHQ$  were not considered for this analysis. Hence, the time series contains less than one value per year.

Unfortunately, there is not a unique answer to the question of which is the best copula fitting a given dataset. As a basis for our choice we took the BIC Index, defined as  $BIC = \ln(n)k - 2\ln(\hat{L})$ , where  $n$  is the number of observations,  $L$  is the maximized value of the likelihood function of the model and  $k$  is the number of free parameters. As shown in Figures 5.12 and 5.13, the time series of the direct peak discharge  $Q$  and of the direct volume  $V$  are not autocorrelated.

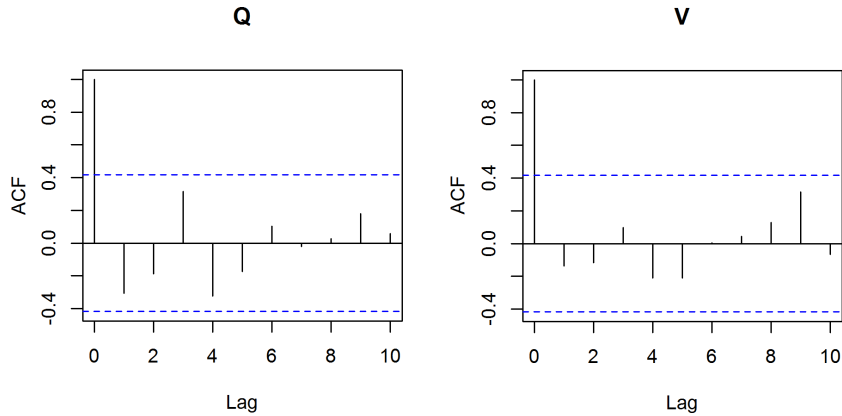


Figure 5.12: Autocorrelation analysis for the Q-V time series of Wasserburg.

In the case of Wasserburg a rotated Tawn copula type 1 was chosen, while the pseudo-observations of Passau have been fitted with a survival Gumbel copula. In Figures 5.14a and 5.14b the observed pseudo-observations (red dots) and the simulated ones (black circles) are shown for Wasserburg and Passau Ingling, respectively. In particular, it is interesting to observe in Figure 5.14a that there is a point characterized by a large direct volume in comparison to the direct peak. This is due to a long event occurred

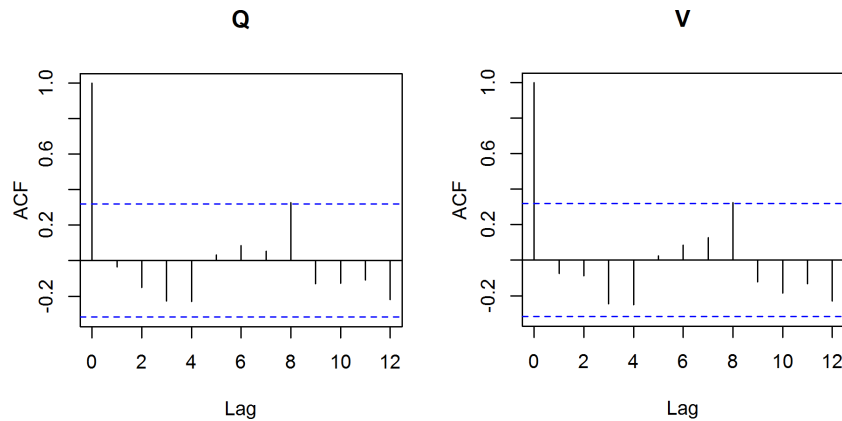


Figure 5.13: Autocorrelation analysis for the Q-V time series of Passau.

between Juni and July 1965, whose peak discharge was on the contrary not so large in comparison to the other events registered at this gauging station. This event was also recorded by the close Inn gauging stations of Mühldorf, and therefore we exclude a measurement error. This event affected the selection of the copula to fit the data and brought in particular to the choice of the Tawn copula, which is an extension of the Gumbel Copula that allows for asymmetries.

Another useful tool to compare the observed and simulated pseudo-observations is the comparison of the cdf of the fitted and of the empirical copula. This can be done observing the corresponding isolines. In Figures 5.15a and 5.15b we can see that in both cases, but especially for Passau, the isolines are matching very well, which means that the fitted copulas well represent the relationship between the direct peak discharge  $Q_{MHQ}$  and the direct volume  $V_{MHQ}$  for these gauging stations. In Figure 5.15a we observe more oscillations of the black solid line, representing the isolines of the empirical copula of the observed values, than the ones observed for Passau. This is due to a smaller amount of available data.

The next step was to report the generated pseudo-observations in the real space. This was done applying the inverse of the cdf of the marginals of  $Q_{MHQ}$  and  $V_{MHQ}$ . The choice of the marginals was based on the evaluation of the Akaike Index. In order to get physically reasonable values, the research was restricted to the

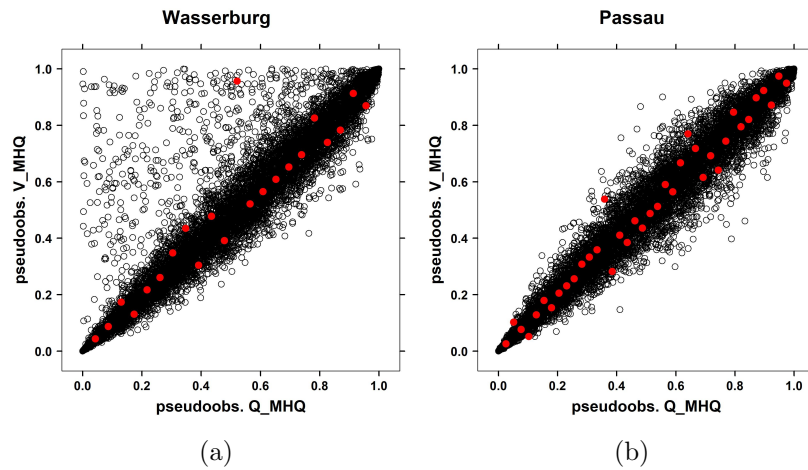


Figure 5.14: Comparison between the observed pseudo-observations (red dots) and the simulated ones (black circles) of the gauging stations of Wasserburg (a) and Passau (b).

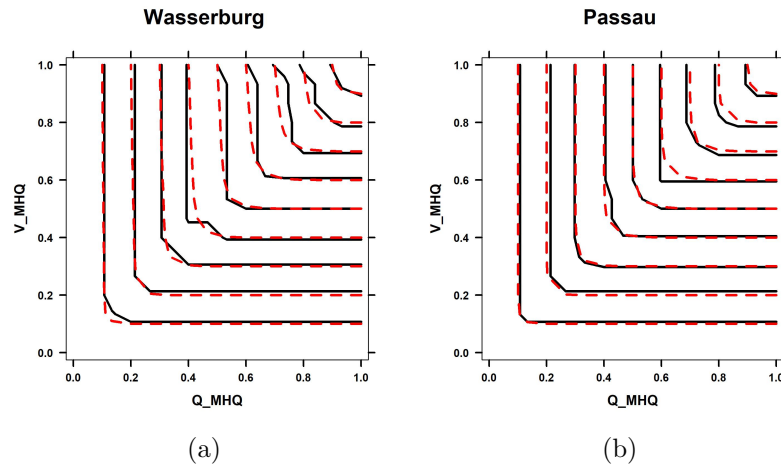


Figure 5.15: Comparison between the isolines of the cdf of the empirical copula of the observed (black solid line) and of the simulated (red dashed line) pseudo-observations of the gauging stations of Wasserburg (a) and Passau (b).

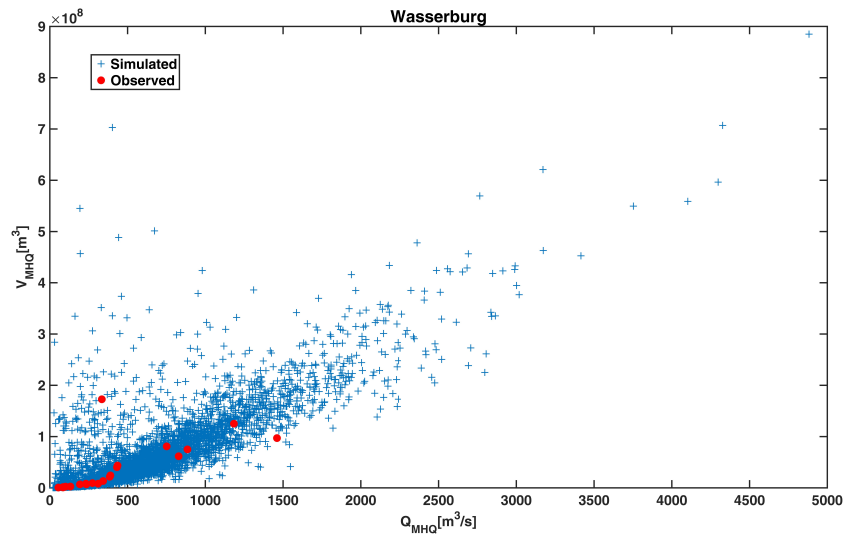


Figure 5.16: Comparison between the observed and simulated  $Q_{MHQ}$ - $V_{MHQ}$  data for the gauging station of Wasserburg.

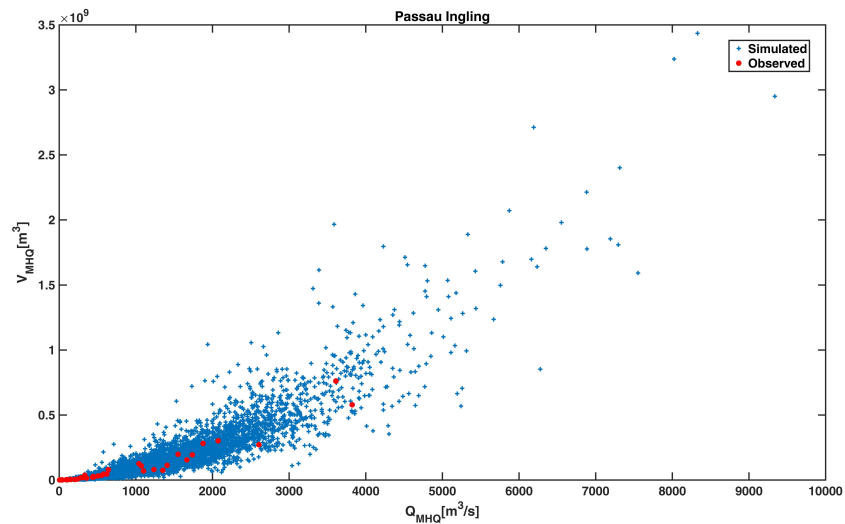


Figure 5.17: Comparison between the observed and simulated  $Q_{MHQ}$ - $V_{MHQ}$  data for the gauging station of Passau Ingling.

distributions allowing only non-negative values.

It is also important that the dependence structure between the observed  $Q_{MHQ}$  and  $V_{MHQ}$  is preserved by the simulated values (see Table 5.2).

Table 5.2: In this table the Kendall's tau  $\tau_k$  and the Spearman's rho  $\rho_S$  correlation coefficients of the observed and of the simulated  $Q_{MHQ}$  and  $V_{MHQ}$  are shown.

	Wasserburg		Passau Ingling	
	$\tau_k$	$\rho_S$	$\tau_k$	$\rho_S$
observed	0.84	0.92	0.91	0.98
simulated	0.82	0.93	0.88	0.98

### 5.3.2 Multivariate application: direct flood peaks - direct volume - rising time - base flow modeling

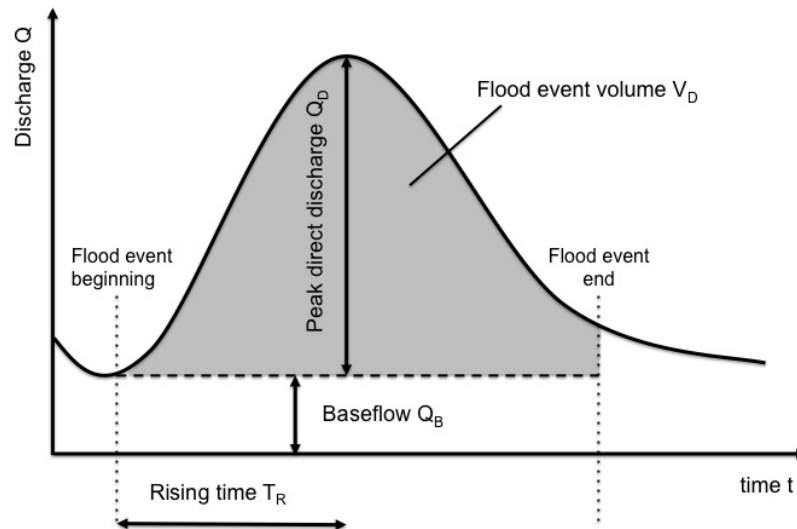


Figure 5.18: Example of the determination of the direct flood peak  $Q_D$ , the direct volume  $V_D$ , the rising time  $T_R$  and the base flow  $Q_B$ .

For each gauging station, the largest flood event of the year was taken in order to create the time series to be analyzed. For each event the direct flood peak  $Q_D$ , the direct volume  $V_D$ , the



rising time  $T_R$  and the base flow  $Q_B$  were computed (see Figure 5.18). The event beginning and end were computed with the aid of the R-packages `lfstat` [Koffler et al., 2016] and `EcoHyDrology` [DR et al., 2014]. Unfortunately, these definitions were not always accurate and for this reason sometimes it was necessary to adjust the results with a visual analysis. As illustrated in Figure 5.18, the baseflow variable  $Q_B$  was determined as the minimum discharge of the event. The direct peak discharge  $Q_D$  was the difference of the maximum discharge of the event and  $Q_B$ . The direct volume  $V_D$  was the volume of the wave above the baseflow  $Q_B$ . The rising time  $T_R$  was defined as the time occurring between the start of the event and the peak discharge. The disaggregation of the daily data maintained the daily volume, but the peak discharge was not always well approximated. For this reason, when dealing with disaggregated data, the direct peak discharge  $Q_D$  was calculated as the difference between the official monthly peak discharge and the baseflow  $Q_B$ .

Table 5.3: Kendall’s tau  $\tau_k$  and the Spearman’s rho  $\rho_S$  correlation coefficients of the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Q_B$  for Wasserburg and Passau.

	Wasserburg		Passau Ingling	
	$\tau_k$	$\rho_S$	$\tau_k$	$\rho_S$
$Q_D-V_D$	0.36	0.50	0.54	0.73
$Q_D-T_R$	-0.04	-0.06	0.20	0.29
$Q_D-Q_B$	0.20	0.33	-0.03	-0.03
$V_D-T_R$	0.52	0.72	0.45	0.63
$V_D-Q_B$	0.03	0.03	-0.09	-0.15
$T_R-Q_B$	-0.14	-0.21	-0.07	-0.12

As reported in Table 5.3 and shown in Figures 5.19 - 5.22, the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Q_B$  extracted in this way are not always very well correlated. The highest correlation can be recognized between the direct volume and the direct peak discharge, and between the direct volume and the rising time. The baseflow appears to be almost uncorrelated with all other three variables.

A 4-dimensional vine copula was fitted for each examined gauging station (see Appendix A). The correlation coefficients of the simulated values are reported in Table 5.4. Comparing them with those reported in Table 5.3 we notice that the correlation coef-

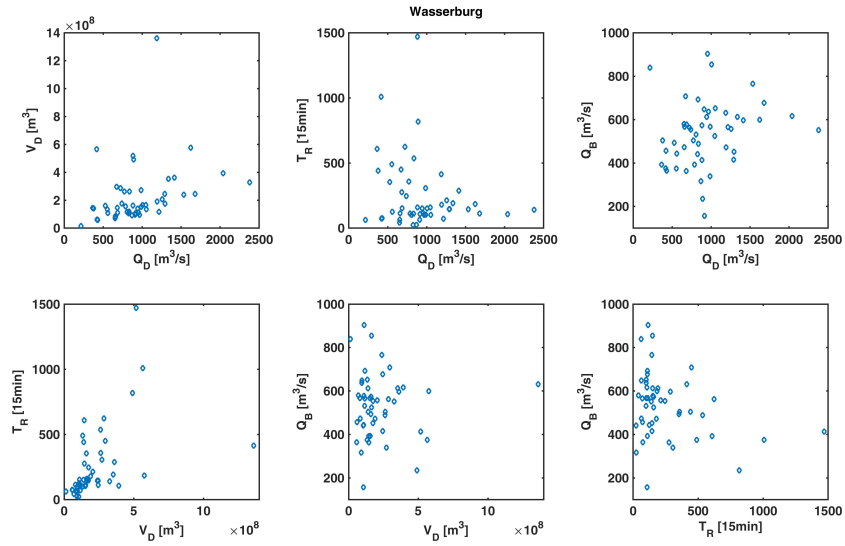


Figure 5.19: Coupled scatter plots for the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Q_B$  for Wasserburg.

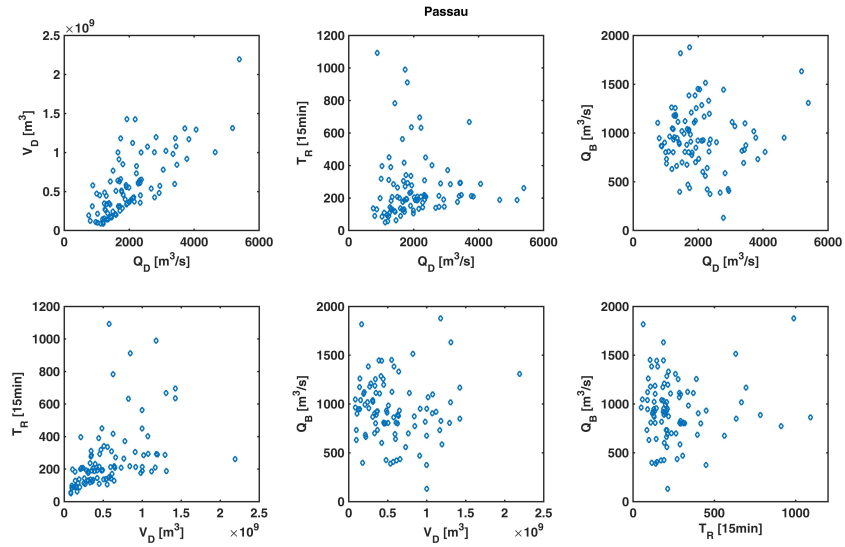


Figure 5.20: Coupled scatter plots for the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Q_B$  for Passau.

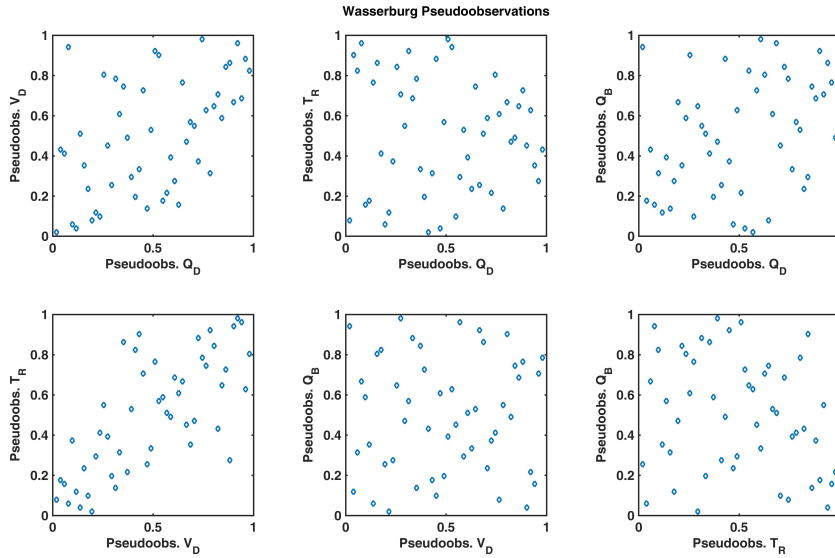


Figure 5.21: Coupled scatter plots for the pseudo-observations of the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Q_B$  for Wasserburg.

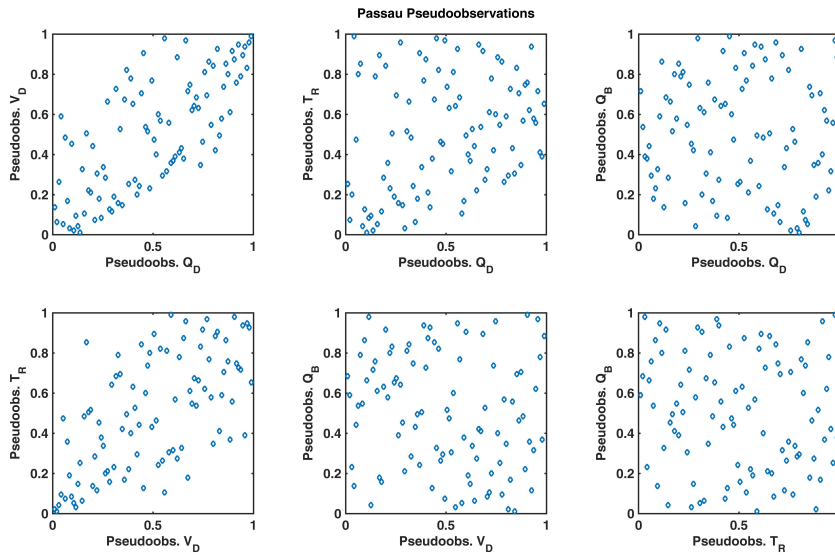


Figure 5.22: Coupled scatter plots for the pseudo-observations of the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Q_B$  for Passau.

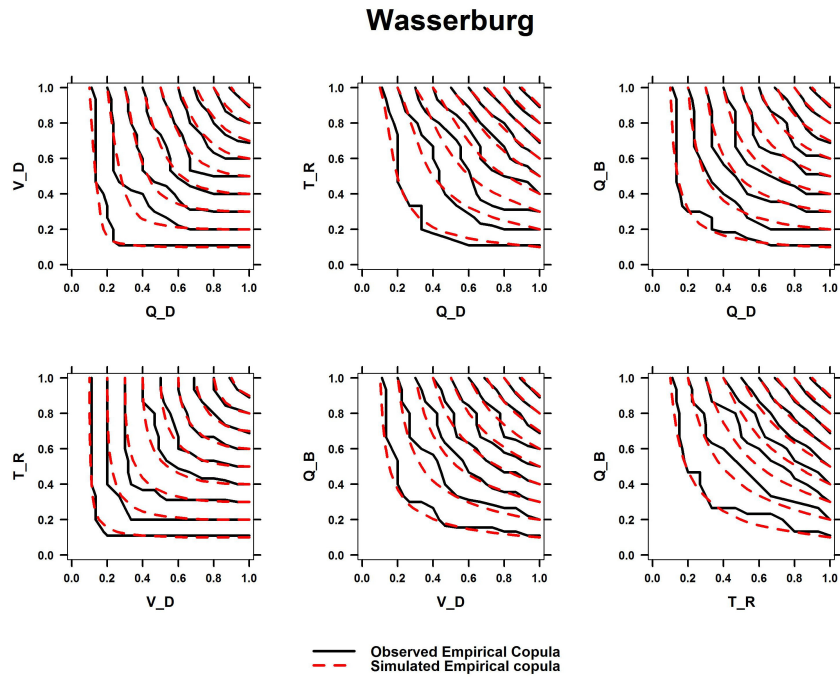


Figure 5.23: Isolines of the cdf of the bivariate empirical copula between each couple of the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Q_B$  of Wasserburg.

Table 5.4: In this table the Kendall's tau  $\tau_k$  and the Spearman's rho  $\rho_S$  correlation coefficients of the simulated variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Q_B$  for Wasserburg and Passau are shown.

	Wasserburg		Passau Ingling	
	$\tau_k$	$\rho_S$	$\tau_k$	$\rho_S$
$Q_D-V_D$	0.40	0.51	0.46	0.61
$Q_D-T_R$	-0.02	-0.03	0.10	0.14
$Q_D-Q_B$	0.22	0.32	-0.0024	-0.003
$V_D-T_R$	0.53	0.72	0.43	0.58
$V_D-Q_B$	0.11	0.17	-0.0014	-0.002
$T_R-Q_B$	-0.0008	-0.001	-0.0023	-0.003

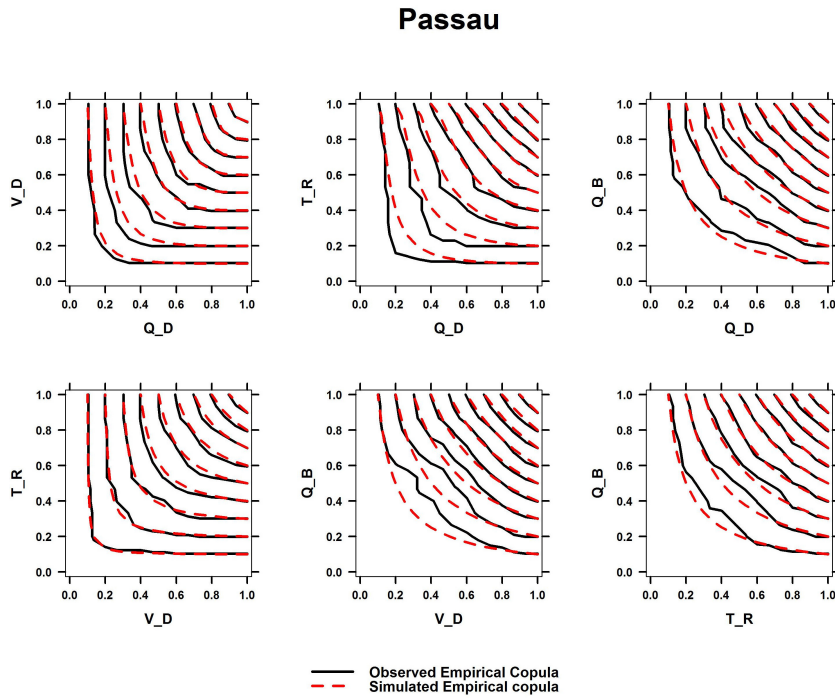


Figure 5.24: Isolines of the cdf of the bivariate empirical copula between each couple of the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Q_B$  of Passau.

ficients of the simulated values are quite similar to those of the observed ones, especially for Wasserburg.

Unfortunately, it is not possible to visualize the isolines of a four dimensional copula. For this reasons we decided to compute the isolines of the cdf of the bivariate empirical copulas for the observed and simulated values (Figures 5.23 and 5.24). Comparing the isolines we can evaluate how the simulated data reproduce the dependence structure present between each couple of variables. We observe a good agreement for the isolines of Wasserburg (Figure 5.23), especially for the couples  $Q_D - V_D$  and  $V_D - T_R$ . A less satisfying result can be observed for the couple  $T_R - Q_B$  for the isolines referring to the values 0.2 and 0.3. Also for Passau the results are quite good (Figure 5.24), especially for the couples  $Q_D - V_D$ ,  $Q_D - Q_B$  and  $V_D - T_R$ . For the couples  $Q_D - T_R$  and  $V_D - Q_B$  the isolines do not agree very well for quantile smaller than 0.3.

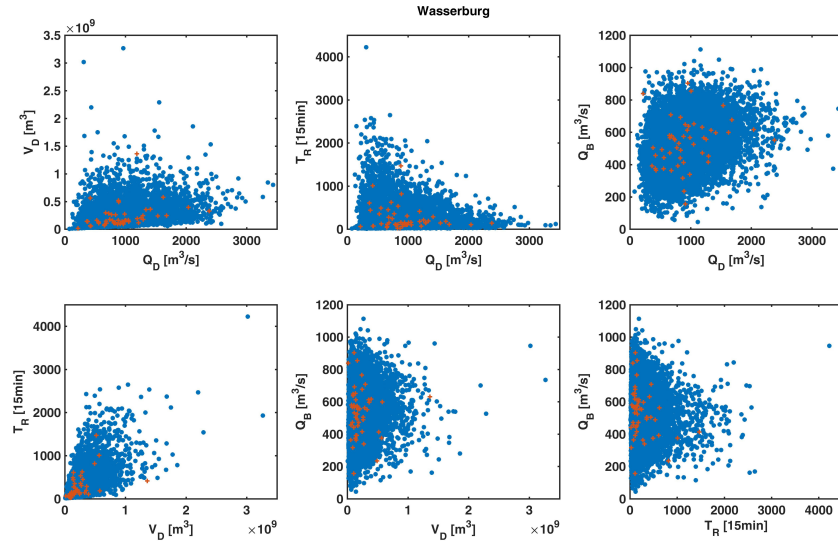


Figure 5.25: Coupled scatter plots of the simulated (blue stars) and of the observed (red crosses) values of the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Q_B$  for Wasserburg.

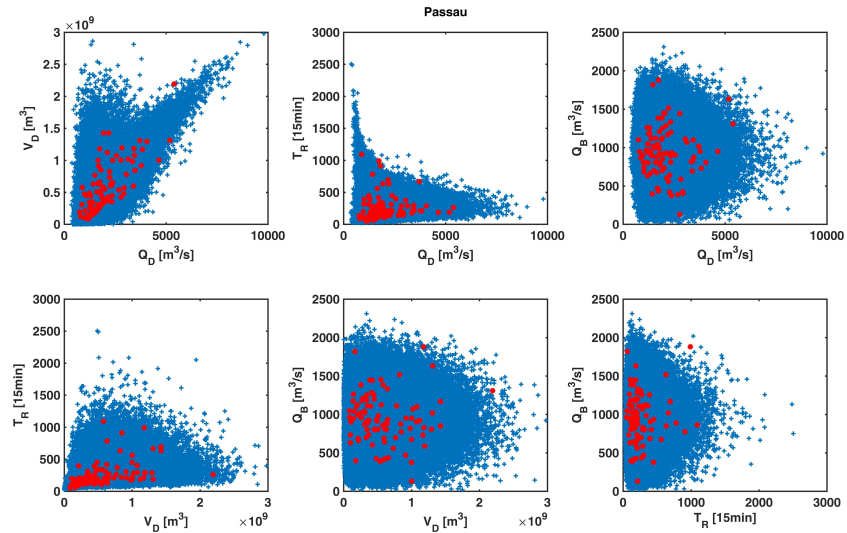


Figure 5.26: Coupled scatter plots of the simulated (blue stars) and of the observed (red crosses) values of the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Q_B$  for Passau.

In Figures 5.25 and 5.26 the simulated (blue stars) and observed (red crosses) values of the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Q_B$  for Wasserburg and Passau are shown, respectively. We can observe that the simulated values well reproduce the behavior of the observed ones.

### 5.3.3 Multivariate application: direct flood peaks - direct volume - rising time - threshold modeling

In this section, in order to gain better correlation between the variables, an optimized threshold  $Th$  was chosen, as illustrated in Figure 5.27. The variables  $Q_D$ ,  $V_D$  and  $T_R$  were then computed as described in section 5.3.2, but using the threshold  $Th$  as a reference instead of the baseflow  $Q_B$ . Similarly to the methodology described in section 5.3.2, when dealing with disaggregated data, the direct peak discharge is calculated as the difference between the official monthly peak discharge and the chosen threshold  $Th$ .

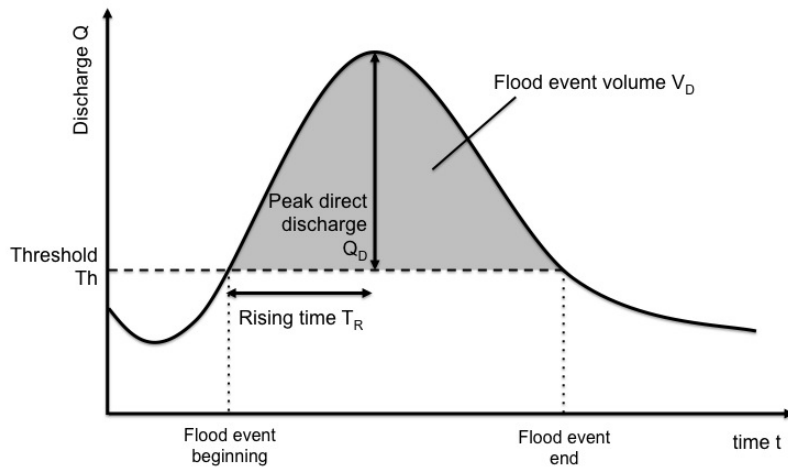


Figure 5.27: Example of the determination of the direct flood peak  $Q_D$ , the direct volume  $V_D$ , the rising time  $T_R$  and the threshold  $Th$ .

For the choice of the threshold  $Th$  the fitted Kozeny-Function [Dyck and Grünwald, 1980, Sackl, 1994] and the corresponding line of retention effect were calculated. In particular, the Kozeny-Function is a function describing a direct wave and is defined as

$$Q_d(t) = Q_D \left( \frac{t}{t_R} \right)^m \exp \left( m \left( 1 - \frac{t}{t_R} \right) \right), \quad (5.20)$$

where  $m$  is a shape parameter.

In order to better describe the variety of direct waves, we optimized two different shape parameters  $m_1$  and  $m_2$  for the rising part and for the descending part of the wave respectively, as illustrated in equation 5.21.

$$Q_d(t) = \begin{cases} Q_D \left( \frac{t}{t_R} \right)^{m_1} \exp \left( m_1 \left( 1 - \frac{t}{t_R} \right) \right) & \text{when } t \leq T_R \\ Q_D \left( \frac{t}{t_R} \right)^{m_2} \exp \left( m_2 \left( 1 - \frac{t}{t_R} \right) \right) & \text{when } t > T_R \end{cases} \quad (5.21)$$

We let  $m_1$  and  $m_2$  varying between 0.1 and 10. This interval for the shape parameters was chosen empirically.

To compute the line of retention effect, the volume of the direct wave between a certain retention level  $R_L$  and  $Q_D$  is computed, with  $R_L$  varying between  $Q_D$  and  $Th$ .

We chose as threshold  $Th$  the smallest discharge value such that the mean absolute percentage error between the line of retention effect of the observed direct wave over the threshold  $Th$  and the line of retention effect of the corresponding fitted Kozeny-Function according to equation 5.21 is less than 0.2. This value was empirically defined. Figures 5.28 and 5.29 show the coupled scatterplot of the variables obtained in this way.

Table 5.5: Kendall's tau  $\tau_k$  and the Spearman's rho  $\rho_S$  correlation coefficients of the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Th$  for Wasserburg and Passau.

	Wasserburg		Passau Ingling	
	$\tau_k$	$\rho_S$	$\tau_k$	$\rho_S$
$Q_D-V_D$	0.81	0.95	0.75	0.92
$Q_D-T_R$	0.62	0.81	0.43	0.59
$Q_D-Th$	-0.47	-0.67	-0.20	-0.28
$V_D-T_R$	0.77	0.93	0.58	0.77
$V_D-Th$	-0.50	-0.70	-0.23	-0.33
$T_R-Th$	-0.50	-0.67	-0.14	-0.20

As shown in Figures 5.30 and 5.31 and as reported in Table 5.5, this definition of the four variables brings a strong improvement



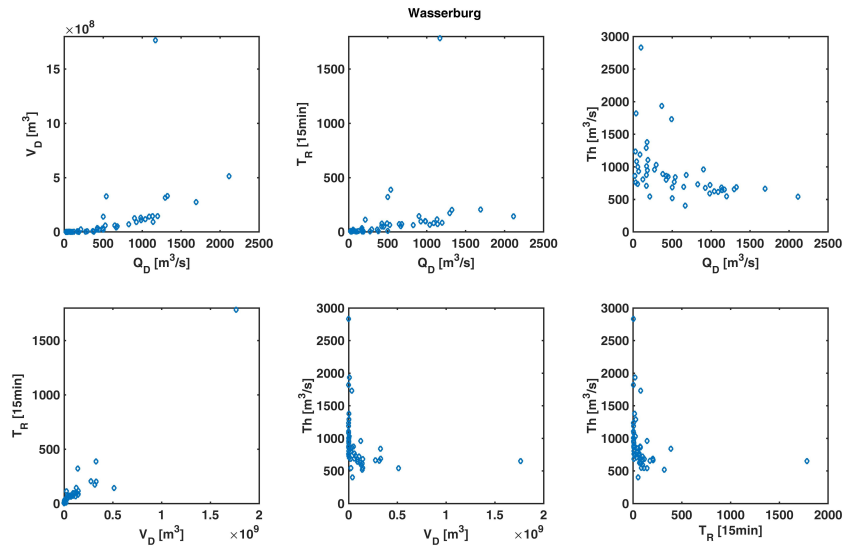


Figure 5.28: Coupled scatter plots for the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Th$  for Wasserburg.

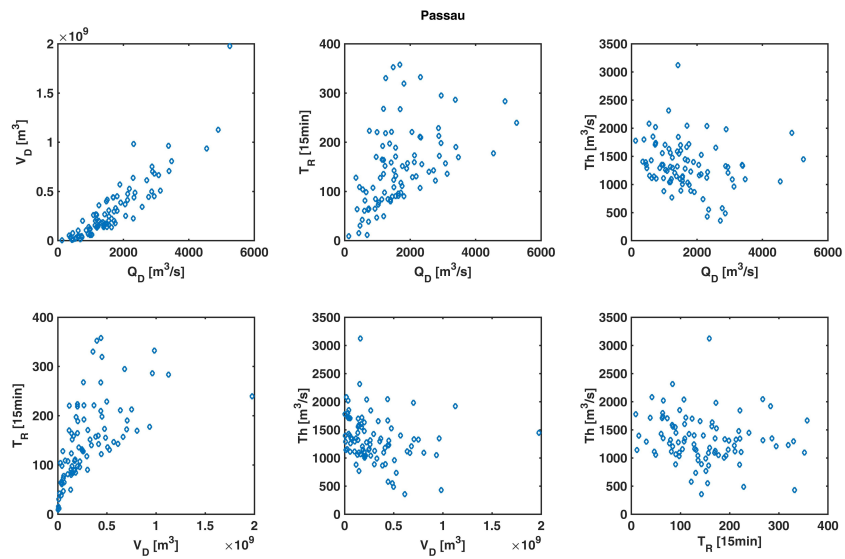


Figure 5.29: Coupled scatter plots for the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Th$  for Passau.

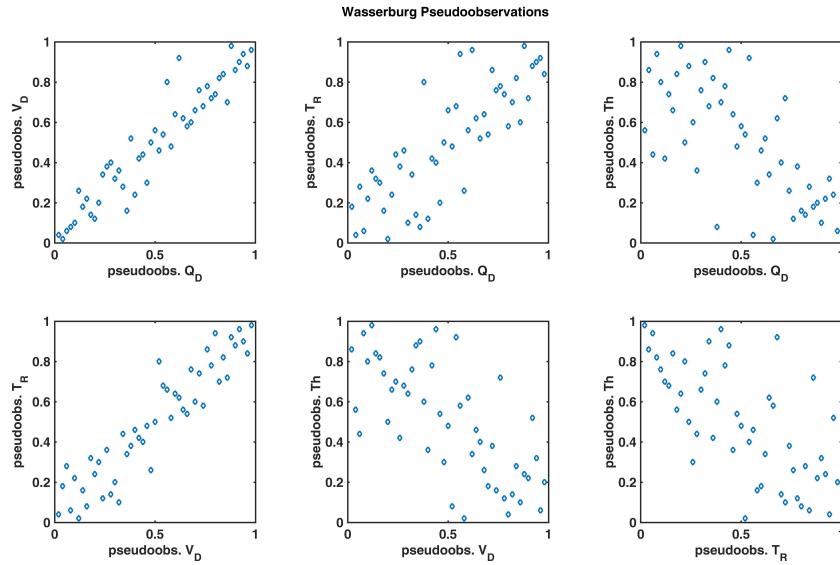


Figure 5.30: Coupled scatter plots for the pseudo-observations of the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Th$  for Wasserburg.

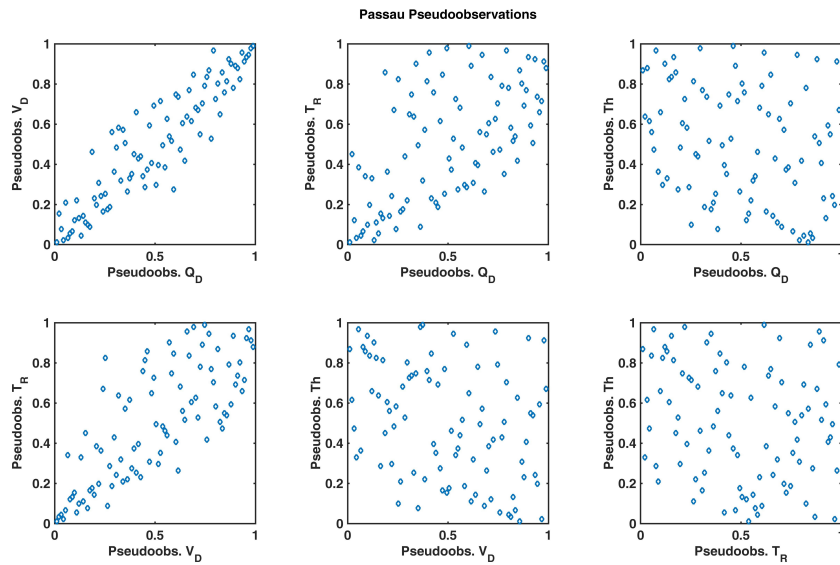


Figure 5.31: Coupled scatter plots for the pseudo-observations of the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Th$  for Passau.

in the correlation of the variables, especially for Wasserburg. In particular, the correlation between the direct peak discharge and the direct volume becomes very strong ( $\tau_k = 0.81$  instead of  $\tau_k = 0.36$  for Wasserburg and  $\tau_k = 0.75$  instead of  $\tau_k = 0.54$  for Passau). It is also important to notice that the moving threshold introduced in this section has a much higher correlation with  $Q_D, V_D$  and  $T_R$  than what observed for  $Q_B$  in section 5.3.2.

We then fitted a 4-dimensional vine copula to the computed variables (see Appendix A). As already done for the other sections, we used the BIC index as a criterion for the choice of the vine copula and then we adjusted the parameter with the loglikelihood function.

Table 5.6: Kendall's tau  $\tau_k$  and the Spearman's rho  $\rho_S$  correlation coefficients of the simulated variables  $Q_D, V_D, T_R$  and  $Th$  for Wasserburg and Passau.

	Wasserburg		Passau Ingling	
	$\tau_k$	$\rho_S$	$\tau_k$	$\rho_S$
$Q_D-V_D$	0.79	0.94	0.76	0.92
$Q_D-T_R$	0.59	0.79	0.42	0.59
$Q_D-Th$	-0.46	-0.64	-0.17	-0.25
$V_D-T_R$	0.76	0.93	0.57	0.75
$V_D-Th$	-0.53	-0.73	-0.19	-0.27
$T_R-Th$	-0.55	-0.75	-0.14	-0.21

Comparing Tables 5.5 and 5.6 we observe that also in this case the correlation coefficients between each couple of the simulated variables are quite close to those of the observed ones. The consistence of the dependence structure between each pair of variables is further analyzed, as already done in section 5.3.2, showing the difference of the isolines of the cdf of the empirical copula of the simulated and observed values (see Figures 5.32 and 5.33). The isolines are very close to each other which shows that the bivariate dependence structure is quite well reproduced.

After fitting the marginals we could then reproduce and compare the simulated values in a real space. The results are given in Figures 5.34 and 5.35. Also in this case, the simulated values seem to reproduce quite well the behavior of the observed ones.

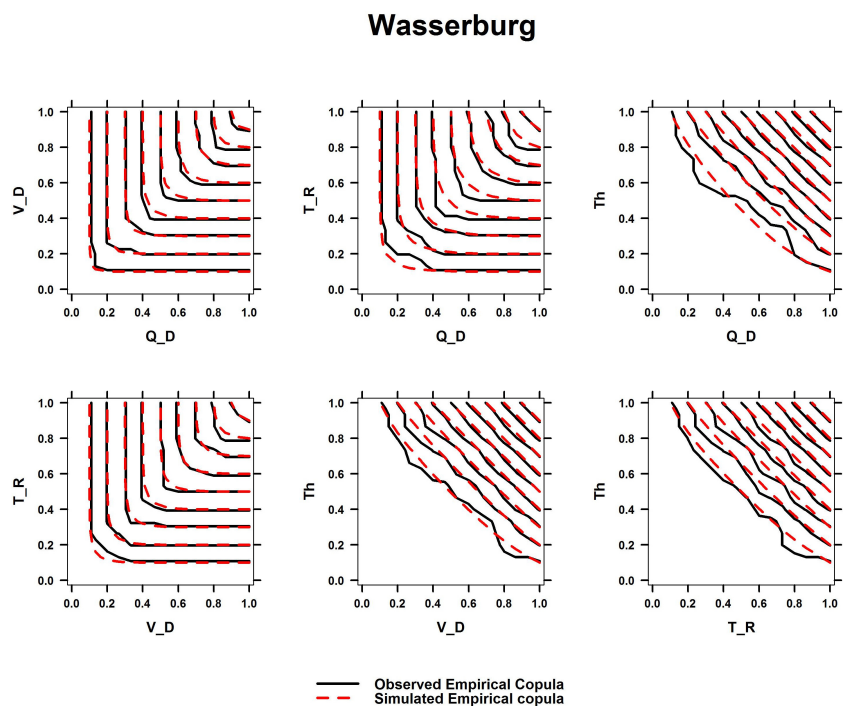


Figure 5.32: Isolines of the cdf of the bivariate empirical copula between each couple of the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Th$  of Wasserburg.

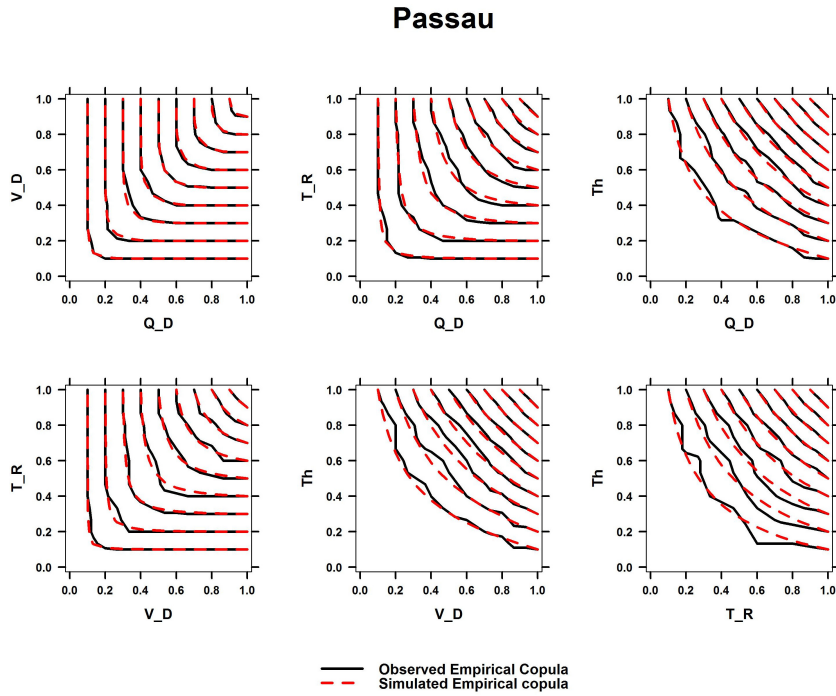


Figure 5.33: Isolines of the cdf of the bivariate empirical copula between each couple of the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Th$  of Passau.

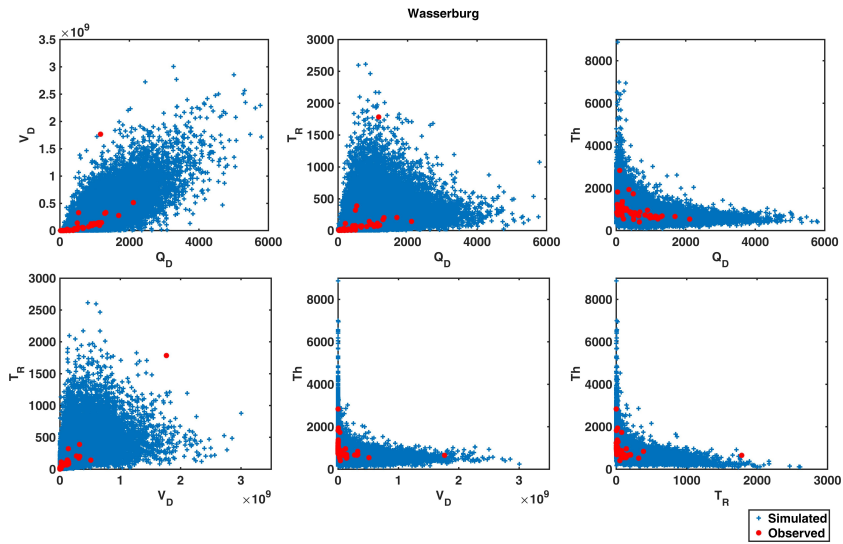


Figure 5.34: Coupled scatter plots of the simulated values of the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Th$  for Wasserburg.

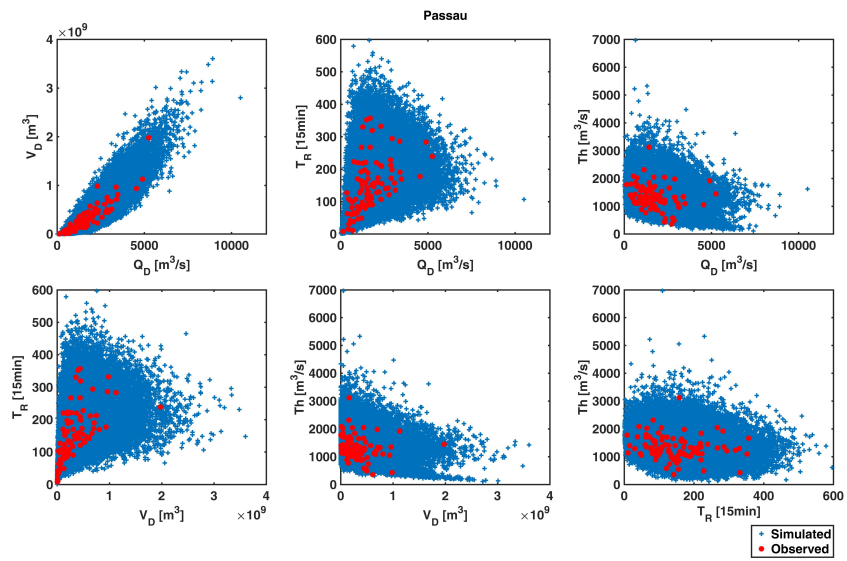


Figure 5.35: Coupled scatter plots of the simulated values of the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Th$  for Passau.

## 5.4 Conclusions

In this chapter we introduced and applied copula analysis to two different types of case studies. In section 5.2 we investigated the dependence structure between the mean seasonal snow depth and the snow cover duration, and between these variables and temperature, MOI and NAOI. The representation of the data as pseudo-observations as well as the characteristics of the copulas fitting the empirical ones confirmed the findings highlighted in chapter 4. Moreover, the copula analysis highlighted the differences for the four examined classes of stations in the dependence for large and small values.

The analysis shown in section 5.3 showed three different ways of describing flood events and how the copula performed in describing them. We observed in section 5.3.1 a strong correlation for the two examined variables (direct peak discharge - direct volume) and a good matching between the c.d.f.s of the observed and simulated empirical copula. The obtained simulations will be used in the project "Retentionspotential-Studie am Inn" for the comparison to other models based on rainfall simulations. The primary goal of the project "Retentionspotential-Studie am Inn" is the evaluation of the retention potential along the Inn river. Therefore it was necessary to simulate the whole waves. This could be done with the aid of the Kozeny-function in eq. 5.20, which required also to recreate the rising time and the baseflow of the wave. For this reason two other combinations of variables were introduced in sections 5.3.2 and 5.3.3. We observed that introducing the moving threshold improved the correlation between the variables. In both cases, the matching between the c.d.f.s of the observed and simulated empirical copula was good. The analysis presented in this chapter contains a novel attempt to reconstruct flood waves using parameters derived by a 4-dimensional copula approach.





## Chapter 6

# Conclusions

The statistical analysis of hydrological time series is a challenging, yet very important task. Hydrological studies suffer for the difficulties in collecting reliable data. Anyone dealing with the analysis of hydrological variables should be well aware of this problem and should handle the data consequently. When analyzing an hydrological variable, it is important to understand the different causes of its variability. We have shown in this work three levels of this analysis, which where, in particular for the snow related variables, not yet explored in details.

The first step was the identification of variabilities that have an anthropogenic origin. This was done with the homogenization analysis. We observed that the homogenization of mean seasonal snow depth data was still an open question, while for other variables, such as temperature and precipitation, several algorithms had already been developed, tested and compared [see e.g. Caussinus and Mestre, 2004, Easterling and Peterson, 1995, Peterson et al., 1998]. Snow dynamics is very important in mountainous regions and can influence many aspects of the ecosystem, such as water availability, or economical activities, such as winter tourism or hydro power production [see e.g. Beniston et al., 2003, Majone et al., 2015]. In chapter 3 we presented an algorithm based on the Standard Normal Homogeneity Test for the homogenization of mean seasonal snow depth time series and we applied it to a dataset collected in the Province of Trento. This algorithm gave good results also in an intercomparison experiment that was performed with the ZAMG of Vienna, where the performance of the SNHT are compared to those obtained with the algorithm HO-

MOP. This research field is still in its infancy and much research is still needed. In fact it would be important to test the performance of other homogenization techniques in order to find which is the most appropriate one for the mean seasonal snow depth as well as for other snow related variables, such as snow cover duration. Moreover, as indicated by Marcolini et al. [2017], a quantitative approach to estimate the uncertainty in the temporal location of a detected breakpoint is still missing and should be developed to improve homogenization tests.

In chapter 4 we reported the statistical analysis of the mean seasonal snow depth and of the snow cover duration between 1980 and 2010 in the Adige river basin. We observed a decrease in the mean seasonal snow depth and snow cover duration, especially during the '90s and we observed how these changes are more important for stations located below 1650 m a.s.l.. We then analyzed the time series by means of the wavelet transform analysis. This analysis is particularly valuable since it highlights the local variability of the various components of a signal in time. This analysis further highlighted the different behavior of the stations below and above 1650 m a.s.l. and the decrease of the mean seasonal snow depth during the '90s. Then, by means of the wavelet coherence analysis, we investigated the relationship between mean seasonal snow depth and snow cover duration and climatic indexes such as the MOI and the NAOI. The analysis showed a correlation with both signals. It was particularly interesting to observe how the changes in the correlation of the mean seasonal snow depth and the MOI reflect the changes observed in the statistical analysis of the mean seasonal snow depth in the last decades. The reduction in mean seasonal snow depth and snow cover duration during the '90s, which was reported in literature also for other parts of the Alpine ridge [see e.g. Beniston et al., 2003], shows a common pattern for the behavior of these variables in the Alps, which is stronger than the local variability affecting snow measurements. The changes in the coherence of the mean seasonal snow depth with the MOI for signals with large periodicity (6-4 years) shows that the behavior of this variable is linked to climate changes. The modification of this coherence during the '90s, in correspondence with the period of the decrease of mean seasonal snow depth and snow cover duration, let us hypothesize a correlation between the MOI and the reduction of the snow.

We also applied the wavelet analysis to the discharge data in

the Adige and Inn river catchments. The wavelet transform analysis highlighted some peaks in the global wavelet spectrum characteristic of stations belonging to the same catchment. Analyzing the variability of the 2-8 years scale average it was possible to identify common patterns for the stations inside the same catchment. Both these analysis indicate that the wavelet spectrum of streamflow is a characteristic for a river basin. The coherence analysis between the discharge data in the Adige an Inn river catchments with the MOI and the NAOI was also performed. The river discharge signal shows a weaker correlation with climate indexes than the one observed for snow related variables, most probably because the discharge is less directly linked to the precipitation and in general to the atmospheric forcing than snow, while it is influenced by human activities such as water management and land use change.

As indicated by Labat [2010] results obtained through wavelet transform and wavelet coherence could allow a first order prediction of the future evolutions of water resources from a climate change point of view in relationship with climate scenarios of evolution of atmospheric and sea surface temperature conditions. However, the high non-stationarity between the coherence of streamflow time series and climatic indexes greatly complicates the achievement of this goal, at least for the Alpine catchments considered in this study. Nonetheless, we have shown that the application of wavelet for hydrological studies is useful to identify areas with similar hydrological behaviors and with similar correlation with climatic indexes. Therefore, wavelet can be used not only for the analysis of hydrological variable in order to identify their main modes of variability and where they are localized in time, but their application can be envisioned also for classification and clustering purposes (e.g., to define similarities between catchment responses towards climatic drivers).

Copulas were applied in chapter 5. They are multivariate distributions that provide wide opportunities to model the interdependence between variables, giving the possibility to model separately the marginals of each variable and their dependence structure. First we used copulas to model and investigate the correlation between mean seasonal snow depth, snow cover duration, temperature and MOI. We observed how the relationship between these variables changes with elevation and also how the properties of the copulas describing the dependence structures between them reflect the changes we observed in chapter 4. While copulas

provide very interesting insights about the dependency between snow variables and temperature, they were not able to describe the changes in the correlation between the mean seasonal snow depth and the MOI occurred in the last decades. This is due to the fact that they provide a different kind of information than wavelet. In fact, wavelet analysis allows us to determine the different frequency content of a signal in different time intervals. For example, the correlation between the MOI and the mean seasonal snow depth observed in chapter 4 was detected at a period of 6-4 years. Furthermore, this coherence changed during the examined period (especially in the '90s, when it vanishes for the stations below 1650 m a.s.l. and decreases for the stations above 1650 m a.s.l.). Copula analysis consider the time series as a whole and, as such, it was not able to highlight the different kinds of correlation in the different periods. This analysis therefore highlights some of the merits and limits of copulas.

We then applied copulas for three different modeling schemes of the yearly flood events of the Inn river registered at the gauging stations of Wasserburg and Passau. We observed good modeling performance of the copulas, especially in the bivariate case. Also the two schemes with 4 variables were well modeled, but the increase in the number of variables made it more difficult to identify the optimal copula, as well as to evaluate its performance. Anyway, despite of the weaker correlation between the variables characterizing the 4-dimensional scheme and the difficulties linked to the increasing number of dimensions, the copula model still showed good performances in modeling the dependence structure between the direct peak discharge, direct volume, rising time and base flow or optimized threshold. This study highlighted the importance of the definition of the variables to model flood events. The three different combinations we have applied can in fact lead to results which can be suitable for different objectives. Modeling the bivariate copula between direct peak and direct volume provides simulations that can be used, for example, for the comparison with physically based methods. Describing flood events by mean of direct peak, direct volume, rising time and a base threshold, such as the base flow or the moving threshold, allows to recreate waves that can be used, for example, for flood preventions studies. For the modeling of 4-dimensional copula, anyway, our experience brings us to recommend the use of a moving thresholds, which leads to an higher correlation among the selected variables.

The analysis shown in chapters 4 and 5 highlighted the advantages and disadvantages of the investigation of snow and discharge related variables with wavelet and copula analysis. In particular, on one hand, the wavelet analysis highlights the changes in the frequency content of a signal in different time steps and give also the possibility to investigate the different correlation between two variables in the time frequency domain. On the other hand, the copula analysis permits to isolate the dependence structure between two variables, independently on their marginals. It also permits to describe the simultaneous correlation of several variables. Copula analysis has also shown to be a valuable tool for modeling hydrological variables. As a drawback, we pointed out the necessity of having long reliable time series, which is rare in case of hydrological variables. We think that wavelet and copula analysis are two valuable tools, whose possible contribution for the hydrological time series analysis has not been fully explored yet. It would be very interesting, for example, to couple the wavelet coherence and copula analysis in order to deepen the dependence structure of the frequency content of two signals. This would also permit to investigate how the dependence structure changes in time. In order to take advantage of both copulas and wavelet analysis, therefore, we could imagine to analyze with the copulas the dependence structure of the scale components obtained after decomposing the signals of two time series with the wavelet transform. Another possible way to couple both analysis would be to identify the non-stationary correlation between two variables using the wavelet transform. Then, separate copulas analysis for the different part of the time series can be performed to identify how the dependence structure between two variables changes in the different parts of the time series. Both proposed approaches could find wide application in hydrological studies for example for the generation of non-stationary time series and for developing novel bivariate analysis methods.

The findings of this work show the importance of the statistical analysis of the hydrological time series. We have shown how it is necessary and useful to combine different approaches, well aware of the strengths and weakness of each method. Hydrological time series analysis allows us to achieve a comprehensive picture of the quality of the data, of the variability contained in the time series and of its relation to other variables.



# Appendix A

## 4-dimensional vine copulas

In this appendix we show the structure and the parameters of the 4-dimensional Vine Copulas fitted in sections 5.3.2 and 5.3.3.

**Multivariate application:**  $Q_D$ - $V_D$ - $T_R$ - $Q_B$

Step 1	Step 2	Step 3
$V_D$ - $T_R$ : Gaussian (0.74)	Rot. Gumbel 90° (-2.04)	Independence
$Q_B$ - $V_D$ : Rot. Tawn t1 180°(2.77,0.5)		
$Q_B$ - $Q_D$ : Frank (2.1)	Independence	

Table A.1: Vine Copula for the modeling of the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Q_B$  of Wasserburg.

Step 1	Step 2	Step 3
$V_D-T_R$ : Clayton (1.43)	Rot. Tawn t2 90° (-3.41,0.34)	Independence
$V_D-Q_D$ : Tawn t1 (2.8,0.63)		
$Q_B-V_D$ : Clayton (1.43)	Independence	

Table A.2: Vine Copula for the modeling of the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Q_B$  of Passau.

**Multivariate application:** $Q_D-V_D-T_R-Th$

Step 1	Step 2	Step 3
$V_D-Q_D$ : Surv. Gumbel (4.66)	Frank (-5)	Independence
$V_D-T_R$ : Frank (14.51)		
$Th-V_D$ : Frank (-6.29)	Rot. Joe 90° (-1.4)	

Table A.3: Vine Copula for the modeling of the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Th$  of Wasserburg.

Step 1	Step 2	Step 3
$V_D-Q_D$ : Gaussian (0.93)	Gaussian (-0.44)	Independence
$V_D-T_R$ : Clayton (2.58)		
$Th-V_D$ : Rot. Tawn t1 90° (-2.67,0.23)	Independence	

Table A.4: Vine Copula for the modeling of the variables  $Q_D$ ,  $V_D$ ,  $T_R$  and  $Th$  of Passau.



# Acknowledgments

First, I would like to thank my advisors Prof. Dr. Alberto Bellin and Prof. Dr.-Ing. Markus Disse for their guidance and support. I would also like to thank Prof. Dr. Salvatore Grimaldi and Prof. Dr. Ralf Ludwig for their interesting questions and comments.

During these years I had the opportunity to collaborate also with institutions and companies outside the universities of Trento and Munich, whome I am very thankful. I would like in particular to mention the Meteorological Survey of the Province of Trento, Roland Koch, Barbara Chimani and the Prof. Dr. Schoener of the Central Institute of Meteorology and Geodynamic of Vienna and Dr. Winfried Willems of IAWG.

My PhD was a wonderful journey that gave me the chance to grow as person and as scientist. I would like to thank all my PhD students and Post-Docs at the Chair of Prof. Dr. Bellin for supporting me in these years, even when I was in Munich. Thanks for the nice coffee breaks, lunch breaks and interesting discussions.

Going back to Germany after so many years was a new and challenging experience. I am very thankful for the wonderful people I met at the Chair of Hydrology and River Basin Management of Prof. Dr.-Ing. Disse. It was very precious to share these years with you.

I am also grateful to the staff of the University of Trento and of the Technical of University of Munich for supporting me in these years of infinite bureaucratic challenges. In particular I would like to thank, in alphabetical order, Mrs. Freizingher, Mrs. Martuscelli, Mrs. Menz, Mrs. Rogato, Dr.-Ing. Spengler, Mrs. Stinzel and Mrs. Wallmeroth.

I would also like to thank my parents and my brother who always supported me in these years, just as they did in they entire life.

In this journey the most important travel companions have

been Gabriele and Giovanni. Their love has supported me and their smile made me feel good and at home, wherever we have been. I am particularly grateful to Gabriele also for his guidance and support. To Giovanni and the new baby, who's coming, goes my last grateful and loving thought.

# Bibliography

- Kjersti Aas and Daniel Berg. Models for construction of multivariate dependence—a comparison study. *The European Journal of Finance*, 15(7-8):639–659, 2009.
- Kjersti Aas, Claudia Czado, Arnaldo Frigessi, and Henrik Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, 44(2):182–198, 2009.
- Enric Aguilar, Inge Auer, Manola Brunet, Thomas C Peterson, and Jon Wieringa. Guidance on metadata and homogenization. *WMO TD*, 1186:53, 2003.
- Hans Alexandersson. A homogeneity test applied to precipitation data. *Journal of climatology*, 6(6):661–675, 1986.
- Hans Alexandersson and Anders Moberg. Homogenization of swedish temperature data. part i: Homogeneity test for linear trends. *International Journal of Climatology*, 17(1):25–34, 1997.
- Theodore W Anderson. *The statistical analysis of time series*, volume 19. John Wiley & Sons, 2011.
- Ingeborg Auer, Reinhard Bohm, Anita Jurkovic, Wolfgang Lipa, Alexander Orlik, Roland Potzmann, Wolfgang Schoner, Markus Ungersbock, Christoph Matulla, Keith Briffa, et al. Histalpre-historical instrumental climatological surface time series of the greater alpine region. *International Journal of Climatology*, 27(1):17–46, 2007.
- András Bárdossy. Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research*, 42(11), 2006.

- András Bárdossy and Jing Li. Geostatistical interpolation using copulas. *Water Resources Research*, 44(7), 2008.
- Tim P Barnett, Jennifer C Adam, and Dennis P Lettenmaier. Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, 438(7066):303–309, 2005.
- Marshall G Bartlett, David S Chapman, and Robert N Harris. Snow and the ground temperature record of climate change. *Journal of Geophysical Research: Earth Surface (2003–2012)*, 109(F4), 2004.
- Tim Bedford and Roger M Cooke. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence*, 32(1): 245–268, 2001.
- Tim Bedford and Roger M Cooke. Vines: A new graphical model for dependent random variables. *Annals of Statistics*, pages 1031–1068, 2002.
- Michael Begert, Evelyn Zenklusen, Christian Häberli, Christof Appenzeller, and Lisette Klok. An automated procedure to detect discontinuities; performance assessment and application to a large european climate data set. *Meteorologische Zeitschrift*, 17(5):663–672, 2008.
- Martin Beniston. Variations of snow depth and duration in the swiss alps over the last 50 years: links to changes in large-scale climatic forcings. *Climatic Change*, 36(3-4):281–300, 1997.
- Martin Beniston. Mountain weather and climate: a general overview and a focus on climatic change in the alps. *Hydrobiologia*, 562(1):3–16, 2006.
- Martin Beniston. Impacts of climatic change on water and associated economic activities in the swiss alps. *Journal of Hydrology*, 412:291–296, 2012a.
- Martin Beniston. Is snow in the alps receding or disappearing? *Wiley Interdisciplinary Reviews: Climate Change*, 3(4):349–358, 2012b.

- Martin Beniston, Franziska Keller, and Stéphane Goyette. Snow pack in the swiss alps under changing climatic conditions: an empirical approach for climate impacts studies. *Theoretical and Applied Climatology*, 74(1-2):19–31, 2003.
- L Bogataj. How will the alps respond to climate change. *Alpine space–man & environment*, 3:43–51, 2007.
- Michele Brunetti, Maurizio Maugeri, and Teresa Nanni. Atmospheric circulation and precipitation in italy for the last 50 years. *International Journal of Climatology*, 22(12):1455–1471, 2002.
- Michele Brunetti, Maurizio Maugeri, Fabio Monti, and Teresa Nanni. Temperature and precipitation variability in italy in the last two centuries from homogenised instrumental time series. *International Journal of Climatology*, 26(3):345–381, 2006.
- Michele Brunetti, Gianluca Lentini, Maurizio Maugeri, Teresa Nanni, Ingeborg Auer, Reinhard Bhm, and Wolfgang Schnier. Climate variability and change in the Greater Alpine Region over the last two centuries based on multi-variable analysis. *International Journal of Climatology*, 29(15):2197–2225, 2009. ISSN 1097-0088. doi: 10.1002/joc.1857. URL <http://dx.doi.org/10.1002/joc.1857>.
- Axel Bücher, Ivan Kojadinovic, et al. A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing. *Bernoulli*, 22(2):927–968, 2016.
- A Buzzi and S Tibaldi. Cyclogenesis in the lee of the alps: A case study. *Quarterly Journal of the Royal Meteorological Society*, 104(440):271–287, 1978.
- Sean K Carey, Doerthe Tetzlaff, Jim Buttle, Hjalmar Laudon, Jeff McDonnell, Kevin McGuire, Jan Seibert, Chris Soulsby, and Jamie Shanley. Use of color maps and wavelet coherence to discern seasonal and interannual climate influences on streamflow variability in northern catchments. *Water Resources Research*, 49(10):6194–6207, 2013.
- Henri Caussinus and Faouzi Lyazrhi. Choosing a linear model with a random number of change-points and outliers. *Annals of the Institute of Statistical Mathematics*, 49(4):761–775, 1997.

- Henri Caussinus and Olivier Mestre. Detection and correction of artificial shifts in climate series. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53(3):405–425, 2004.
- Chris Chatfield. *The analysis of time series: an introduction*. CRC press, 2016.
- Gabriele Chiogna, Emilio Santoni, Federica Camin, Agostino Tonon, Bruno Majone, Alberto Trenti, and Alberto Bellin. Stable isotope characterization of the vermigliana catchment. *Journal of Hydrology*, 509:295–305, 2014.
- Gabriele Chiogna, Bruno Majone, Karina Cano Paoli, Elena Diamantini, Elisa Stella, Stefano Mallucci, Valeria Lencioni, Fabiana Zandonai, and Alberto Bellin. A review of hydrological and chemical stressors in the adige catchment and its ecological status. *Science of The Total Environment*, 540:429 – 443, 2016. ISSN 0048-9697. doi: <http://dx.doi.org/10.1016/j.scitotenv.2015.06.149>. URL <http://www.sciencedirect.com/science/article/pii/S0048969715303430>. 5th Special Issue SCARCE: River Conservation under Multiple stressors: Integration of ecological status, pollution and hydrological variability.
- P Coulibaly and D H Burn. Wavelet analysis of variability in annual Canadian streamflows. *Water Resour. Res.*, 40(3), 2004. doi: 10.1029/2003WR002667.
- Claudia Czado. Pair-copula constructions of multivariate copulas. *Copula theory and its applications*, pages 93–109, 2010.
- Peter Domonkos. Measuring performances of homogenization methods. *Időjárás*, 117:91–112, 2013.
- Peter Domonkos and Petr Štěpánek. Statistical characteristics of detectable inhomogeneities in observed meteorological time series. *Studia Geophysica et Geodaetica*, 53(2):239–260, 2009.
- Fuka DR, Walter MT, Archibald JA, Steenhuis TS, and Easton ZM. *EcoHydrology: A community modeling foundation for Eco-Hydrology.*, 2014. URL <http://CRAN.R-project.org/package=EcoHydrology>. R package version 0.4.12.
- Jean-François Ducre-Robitaille, Lucie A Vincent, and Gilles Boulet. Comparison of techniques for detection of discontinuities

- in temperature series. *International Journal of Climatology*, 23(9):1087–1101, 2003.
- Yves Durand, Gérald Giraud, Martin Laternser, Pierre Etchevers, Laurent Mérindol, and Bernard Lesaffre. Reanalysis of 47 years of climate in the french alps (1958-2005): Climatology and trends for snow cover. *Journal of applied meteorology and climatology*, 48(12):2487–2512, 2009.
- Siegfried Dyck and Uwe Grünewald. *Berechnung und Regelung des Durchflusses der Flüsse*. VEB Verlag für Bauwesen, 1980.
- David R Easterling and Thomas C Peterson. A new method for detecting undocumented discontinuities in climatological time series. *International journal of climatology*, 15(4):369–377, 1995.
- David R Easterling, Thomas C Peterson, and Thomas R Karl. On the development and use of homogenized climate datasets. *Journal of climate*, 9(6):1429–1434, 1996.
- Paul Embrechts, Filip Lindskog, and Alexander McNeil. Modelling dependence with copulas. *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich*, 2001.
- Paul Embrechts, Alexander McNeil, and Daniel Straumann. Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, 176223, 2002.
- Anne-Catherine Favre, Salaheddine El Adlouni, Luc Perreault, Nathalie Thiémonge, and Bernard Bobée. Multivariate hydrological frequency analysis using copulas. *Water resources research*, 40(1), 2004.
- David Gampe, Grigory Nikulin, and Ralf Ludwig. Using an ensemble of regional climate models to assess climate change impacts on water scarcity in european river basins. *Science of the Total Environment*, 573:1503–1518, 2016.
- Christian Genest and Anne-Catherine Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering*, 12(4):347–368, 2007.

- Christian Genest and Louis-Paul Rivest. Statistical inference procedures for bivariate archimedean copulas. *Journal of the American statistical Association*, 88(423):1034–1043, 1993.
- Christian Genest, Kilani Ghoudi, and L-P Rivest. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, pages 543–552, 1995.
- Christian Genest, Michel Gendron, and Michaël Bourdeau-Brien. The advent of copulas in finance. *The European Journal of Finance*, 15(7-8):609–618, 2009.
- Benedikt Gräler, Martinus van den Berg, Sander Vandenberghe, Andrea Petroselli, Salvatore Grimaldi, Bernard De Baets, and Niko Verhoest. Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrology and Earth System Sciences*, 17(4):1281–1296, 2013.
- S Grimaldi, A Petroselli, G Salvadori, and C De Michele. Catchment compatibility via copulas: A non-parametric study of the dependence structures of hydrological responses. *Advances in Water Resources*, 90:116–133, 2016.
- Salvatore Grimaldi and Francesco Serinaldi. Asymmetric copula in multivariate flood frequency analysis. *Advances in Water Resources*, 29(8):1155–1167, 2006.
- Aslak Grinsted, John C Moore, and Svetlana Jevrejeva. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear processes in geophysics*, 11(5/6):561–566, 2004.
- K. Guan, S. E. Thompson, C. J. Harman, N. B. Basu, P. S. C. Rao, M. Sivapalan, A. I. Packman, and P. K. Kalita. Spatiotemporal scaling of hydrological and agrochemical export dynamics in a tile-drained midwestern watershed. *Water Resources Research*, 47(10):n/a–n/a, 2011. ISSN 1944-7973. doi: 10.1029/2010WR009997. URL <http://dx.doi.org/10.1029/2010WR009997>.



- Inger Hanssen-Bauer and Eirik J Fjørland. Homogenizing long norwegian precipitation series. *Journal of Climate*, 7(6):1001–1013, 1994.
- Michael Hantel, Martin Ehrendorfer, Annemarie Haslinger, et al. Climate sensitivity of snow cover duration in austria. *International Journal of Climatology*, 20(6):615–640, 2000.
- Andrew Harding, Jean Palutikof, and Tom Holt. The climate system. *The Physical Geography of the Mediterranean*. Oxford University Press, Oxford, pages 69–88, 2009.
- Marius Hofert and Martin Mächler. Nested archimedean copulas meet R: The nacopula package. *Journal of Statistical Software*, 39(9):1–20, 2011. URL <http://www.jstatsoft.org/v39/i09/>.
- Marius Hofert, Ivan Kojadinovic, Martin Maechler, and Jun Yan. *copula: Multivariate Dependence with Copulas*, 2017. URL <https://CRAN.R-project.org/package=copula>. R package version 0.999-15.
- James W Hurrell. Decadal trends in the north atlantic oscillation: regional temperatures and precipitation. *Science*, 269(5224):676–679, 1995.
- James W Hurrell and Harry Van Loon. Decadal variations in climate associated with the north atlantic oscillation. *Climatic change*, 36(3):301–326, 1997.
- JW Hurrell. Decadal trends in the north atlantic oscillation: regional temperatures and precipitation. *Oceanographic Literature Review*, 2(43):116, 1996.
- Harry Joe. *Multivariate models and multivariate dependence concepts*. CRC Press, 1997.
- Harry Joe, Haijun Li, and Aristidis K Nikoloulopoulos. Tail dependence functions and vine copulas. *Journal of Multivariate Analysis*, 101(1):252–270, 2010.
- Kees Jong, Elena Marchiori, Aad Van Der Vaart, Bauke Ylstra, Marjan Weiss, and Gerrit Meijer. Chromosomal breakpoint detection in human cancer. In *Applications of Evolutionary Computing*, pages 54–65. Springer, 2003.

- MN Khaliq and TBMJ Ouarda. On the critical values of the standard normal homogeneity test (snht). *International Journal of Climatology*, 27(5):681–687, 2007.
- Roland Koch, Barbara Chimani, and Wolfgang Schöner. Experiences in homogenization of austrian snow depth observations. In *EGU General Assembly Conference Abstracts*, volume 16, page 11665, 2014.
- Urs Koenig and Bruno Abegg. Impacts of climate change on winter tourism in the swiss alps. *Journal of sustainable tourism*, 5(1): 46–58, 1997.
- Daniel Koffler, Tobias Gauster, and Gregor Laaha. *lfstat: Calculation of Low Flow Statistics for Daily Stream Flow Data*, 2016. URL <http://CRAN.R-project.org/package=lfstat>. R package version 0.9.4.
- Ivan Kojadinovic and Jun Yan. Modeling multivariate distributions with continuous margins using the copula R package. *Journal of Statistical Software*, 34(9):1–20, 2010. URL <http://www.jstatsoft.org/v34/i09/>.
- J Korck, J Danneberg, and W Willems. Impacts of climate change on the water regime of the inn river basin—extracting adaptation-relevant information from climate model ensembles and impact modelling. *Advances in Geosciences*, 32:99–107, 2012.
- David Labat. Recent advances in wavelet analyses: Part 1. a review of concepts. *Journal of Hydrology*, 314(1):275–288, 2005.
- David Labat. Wavelet analysis of the annual discharge records of the worlds largest rivers. *Advances in Water Resources*, 31(1): 109–117, 2008.
- David Labat. Cross wavelet analyses of annual continental fresh-water discharge and selected climate indices. *Journal of Hydrology*, 385(1):269–278, 2010.
- Martin Laternser and Martin Schneebeli. Long-term snow climate trends of the swiss alps (1931–99). *international Journal of climatology*, 23(7):733–750, 2003a.

- Martin Laternser and Martin Schneebeli. Long-term snow climate trends of the swiss alps (1931–99). *international Journal of climatology*, 23(7):733–750, 2003b.
- K M Lau and H Weng. Climate signal detection using wavelet transform: How to make a time series sing. *Bull. Amer. Meteor. Soc.*, 76:2391–2402, 1995. doi: 10.1175/1520-0477(1995)076<2391:CSDUWT>2.0.CO;2.
- Émilie Lebarbier. Detecting multiple change-points in the mean of gaussian process by model selection. *Signal processing*, 85(4): 717–736, 2005.
- Ralf Lindau and Victor Venema. The uncertainty of break positions detected by homogenization algorithms in climate records. *International Journal of Climatology*, 36(2):576–589, 2016. ISSN 1097-0088. doi: 10.1002/joc.4366. URL <http://dx.doi.org/10.1002/joc.4366>.
- Jl López-Moreno, SM Vicente-Serrano, E Morán-Tejeda, J Lorenzo-Lacruz, A Kenawy, and Martin Beniston. Effects of the north atlantic oscillation (nao) on combined temperature and precipitation winter modes in the mediterranean mountains: observed relationships and projections for the 21st century. *Global and Planetary Change*, 77(1):62–76, 2011.
- Jürg Luterbacher, Elena Xoplaki, Carlo Casty, Heinz Wanner, Andreas Pauling, Marcel Küttel, Stefan Brönnimann, Erich Fischer, Dominik Fleitmann, Fidel J Gonzalez-Rouco, et al. Mediterranean climate variability over the last centuries: a review. *Developments in Earth and environmental Sciences*, 4: 27–148, 2006.
- Stefanie R Lutz, Stefano Mallucci, Elena Diamantini, Bruno Majone, Alberto Bellin, and Ralf Merz. Hydroclimatic and water quality trends across three mediterranean river basins. *Science of the Total Environment*, 571:1392–1406, 2016.
- Bruno Majone, Francesca Villa, Roberto Deidda, and Alberto Bellin. Impact of climate change and water use policies on hydropower potential in the south-eastern alpine region. *Science of The Total Environment*, 2015.

- Giorgia Marcolini, Alberto Bellin, and Gabriele Chiogna. Performance of the standard normal homogeneity test for the homogenization of mean seasonal snow depth time series. *International Journal of Climatology*, 2017.
- John Marshall, Yochanan Kushnir, David Battisti, Ping Chang, Arnaud Czaja, Robert Dickson, James Hurrell, MICHAEL McCARTNEY, R Saravanan, and Martin Visbeck. North atlantic climate variability: phenomena, impacts and mechanisms. *International Journal of Climatology*, 21(15):1863–1898, 2001.
- Christoph Marty. Regime shift of snow days in switzerland. *Geophysical Research Letters*, 35(12):n/a–n/a, 2008. ISSN 1944-8007. doi: 10.1029/2008GL033998. URL <http://dx.doi.org/10.1029/2008GL033998>. L12501.
- Alexander J McNeil and Johanna Nešlehová. Multivariate archimedean copulas, d-monotone functions and  $\ell_1$ -norm symmetric distributions. *The Annals of Statistics*, pages 3059–3097, 2009.
- Matthew J Menne and Claude N Williams Jr. Detection of undocumented changepoints using multiple test statistics and composite reference series. *Journal of Climate*, 18(20):4271–4286, 2005.
- Matthew J Menne and Claude N Williams Jr. Homogenization of temperature series via pairwise comparisons. *Journal of Climate*, 22(7):1700–1717, 2009.
- Olivier Mestre, Christine Gruber, Clémentine Prieur, Henri Causinus, and Sylvie Jourdain. Splidhom: A method for homogenization of daily temperature observations. *Journal of Applied Meteorology and Climatology*, 50(11):2343–2358, 2011.
- Olivier Mestre, Peter Domonkos, Franck Picard, Ingeborg Auer, Stéphane Robin, Emilie Lebarbier, Reinhard Böhm, Enric Aguilar, Jose Guijarro, Gregor Vertachnik, et al. Homer: a homogenization software—methods and applications. *Idojaras, Quarterly journal of the Hungarian Meteorological Service*, 117(1):47–67, 2013.
- Paul W Mielke Jr. Non-metric statistical analyses: some metric alternatives. *Journal of Statistical Planning and Inference*, 13: 377–387, 1986.

- Paul W Mielke Jr. The application of multivariate permutation methods based on distance functions in the earth sciences. *Earth-Science Reviews*, 31(1):55–71, 1991.
- Anders Moberg and Hans Alexandersson. Homogenization of swedish temperature data. part ii: homogenized gridded air temperature compared with a subset of global gridded air temperature since 1861. *International Journal of Climatology*, 17(1):35–54, 1997.
- Roger B Nelsen. Dependence and order in families of archimedean copulas. *Journal of Multivariate Analysis*, 60(1):111–122, 1997.
- Roger B. Nelsen. *An Introduction to Copulas*. Springer, New York, 1999.
- Roger B Nelsen. Properties and applications of copulas: A brief survey. In *Proceedings of the First Brazilian Conference on Statistical Modeling in Insurance and Finance*, (Dhaene, J., Kolev, N., Morettin, PA (Eds.)), *University Press USP: Sao Paulo*, pages 10–28. Citeseer, 2003.
- Johanna Nemeč, Christine Gruber, Barbara Chimani, and Ingeborg Auer. Trends in extreme temperature indices in austria based on a new homogenised dataset. *International Journal of Climatology*, 33(6):1538–1550, 2013.
- J Palutikof. Analysis of mediterranean climate data: measured and modelled. In *Mediterranean Climate*, pages 125–132. Springer, 2003.
- D Penna, M Engel, L Mao, A Dell’Agnese, G Bertoldi, and F Comiti. Tracer-based analysis of spatial and temporal variations of water sources in a glacierized catchment. *Hydrology and Earth System Sciences*, 18(12):5271–5288, 2014.
- Teresa Pérez Ciria. Wavelet analysis of alpine catchments at multiple spatial and temporal scales. Master’s thesis, Technical University Munich, 2016.
- Thomas C Peterson and David R Easterling. Creation of homogeneous composite climatological reference series. *International journal of climatology*, 14(6):671–679, 1994.

- Thomas C Peterson, David R Easterling, Thomas R Karl, Pavel Groisman, Neville Nicholls, Neil Plummer, Simon Torok, Ingeborg Auer, Reinhard Boehm, Donald Gullett, et al. Homogeneity adjustments of in situ atmospheric climate data: a review. *International Journal of Climatology*, 18(13):1493–1517, 1998.
- E Piervitali, M Colacino, and Conte M. Rainfall over the central-western mediterranean basin in the period 1951-1995. part ii: precipitation scenarios. *Il Nuovo Cimento C*, 22(5):649–662, 1999.
- R Quadrelli, M Lazzeri, C Cacciamani, and S Tibaldi. Observed winter alpine precipitation variability and links with large-scale circulation patterns. *Climate Research*, 17(3):275–284, 2001.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL <http://www.R-project.org/>.
- Jaxk Reeves, Jien Chen, Xiaolan L Wang, Robert Lund, and Qi Qi Lu. A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, 46(6):900–915, 2007.
- Philip C Reid, Renata E Hari, Grégory Beaugrand, David M Livingstone, Christoph Marty, Dietmar Straile, Jonathan Barichivich, Eric Goberville, Rita Adrian, Yasuyuki Aono, et al. Global impacts of the 1980s regime shift. *Global change biology*, 22(2):682–703, 2016.
- Bernhard Sackl. *Ermittlung von Hochwasser-Bemessungsganglinien in beobachteten und unbeobachteten Einzugsgebieten*. Eigenverl. d. Inst. für Hydromechanik, Hydraulik u. Hydrologie, 1994.
- G Salvadori and C De Michele. Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water Resources Research*, 40(12), 2004.
- G Salvadori and C De Michele. On the use of copulas in hydrology: theory and practice. *Journal of Hydrologic Engineering*, 12(4): 369–380, 2007.

- G Salvadori, GR Tomasicchio, and F D'Alessandro. Practical guidelines for multivariate analysis and design in coastal and off-shore engineering. *Coastal Engineering*, 88:1–14, 2014.
- Deepayan Sarkar and Felix Andrews. *latticeExtra: Extra Graphical Utilities Based on Lattice*, 2016. URL <http://CRAN.R-project.org/package=latticeExtra>. R package version 0.6-28.
- Ulf Schepsmeier, Jakob Stoeber, Eike Christian Brechmann, Benedikt Graeler, Thomas Nagler, and Tobias Erhardt. *VineCopula: Statistical Inference of Vine Copulas*, 2016. URL <http://CRAN.R-project.org/package=VineCopula>. R package version 2.0.1.
- Simon C Scherrer, Christian Wüthrich, Mischa Croci-Maspoli, Rolf Weingartner, and Christof Appenzeller. Snow variability in the swiss alps 1864–2009. *International Journal of Climatology*, 33(15):3162–3173, 2013.
- Jürg Schmidli, Christoph Schmutz, Christoph Frei, Heinz Wanner, and Christoph Schär. Mesoscale precipitation variability in the region of the european alps during the 20th century. *International Journal of Climatology*, 22(9):1049–1074, 2002.
- W. Schnier, I. Auer, and R. Bhm. Long term trend of snow depth at sonnblick (austrian alps) and its relation to climate change. *Hydrological Processes*, 23(7):1052–1063, 2009. ISSN 1099-1085. doi: 10.1002/hyp.7209. URL <http://dx.doi.org/10.1002/hyp.7209>.
- Francesco Serinaldi and Salvatore Grimaldi. Fully nested 3-copula: procedure and application on hydrological data. *Journal of Hydrologic Engineering*, 12(4):420–430, 2007.
- Gaëlle Serquet, Christoph Marty, Jean-Pierre Dulex, and Martine Rebetz. Seasonal trends and temperature dependence of the snowfall/precipitation-day ratio in switzerland. *Geophysical Research Letters*, 38(7), 2011.
- M Sifuzzaman, MR Islam, and MZ Ali. Application of wavelet transform and its advantages compared to fourier transform. 2009.

- Abe Sklar. Fonctions de rpartition n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 1959.
- Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.
- Jean-Paul Theurillat and Antoine Guisan. Potential impact of climate change on vegetation in the european alps: a review. *Climatic change*, 50(1-2):77–109, 2001.
- A Toreti, FG Kuglitsch, E Xoplaki, PM Della-Marta, E Aguilar, M Prohom, and J Luterbacher. A note on the use of the standard normal homogeneity test to detect inhomogeneities in climatic time series. *International Journal of Climatology*, 31(4):630–632, 2011.
- Christopher Torrence and Gilbert P Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological society*, 79(1):61–78, 1998.
- Mauro Valt and Paola Cianfarra. Recent snow cover variability in the italian alps. *Cold Regions Science and Technology*, 64(2): 146–157, 2010.
- Victor KC Venema, Olivier Mestre, Enric Aguilar, Ingeborg Auer, Jose A Guijarro, Peter Domonkos, G Vertacnik, Tamas Szentimrey, Petr Stepanek, P Zahradnicek, et al. Benchmarking homogenization algorithms for monthly data. *Climate of the Past*, 8(1):89–115, 2012.
- Lucie A Vincent. A technique for the identification of inhomogeneities in canadian temperature series. *Journal of Climate*, 11(5):1094–1104, 1998.
- Lucie A Vincent, X Zhang, BR Bonsal, and WD Hogg. Homogenization of daily temperatures over canada. *Journal of Climate*, 15(11):1322–1334, 2002.
- M. Wagner. *Regionalisierung von Hochwasserscheiteln auf Basis einer gekoppelten Niederschlag-Abfluss-Statistik mit besonderer Beachtung von Extremereignissen*. PhD thesis, Technische Universität Dresden, 2012.
- W. Willems. *HyStat / NQ HQ . Einzelplatz Version 4.0.8 (Schulversion / Sonderedition)*. IAWG, 2013.



- Claude N Williams, Matthew J Menne, and Peter W Thorne. Benchmarking the performance of pairwise homogenization of surface temperatures in the united states. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 117(D5), 2012.
- E Xoplaki, JF Gonzalez-Rouco, J u Luterbacher, and H Wanner. Wet season mediterranean precipitation variability: influence of large-scale dynamics and trends. *Climate dynamics*, 23(1): 63–78, 2004.
- Jun Yan et al. Enjoy the joy of copulas: with a package copula. *Journal of Statistical Software*, 21(4):1–21, 2007.
- G Zolezzi, A Bellin, MC Bruno, B Maiolini, and A Siviglia. Assessing hydrological alterations at multiple temporal scales: Adige river, italy. *Water resources research*, 45(12), 2009.