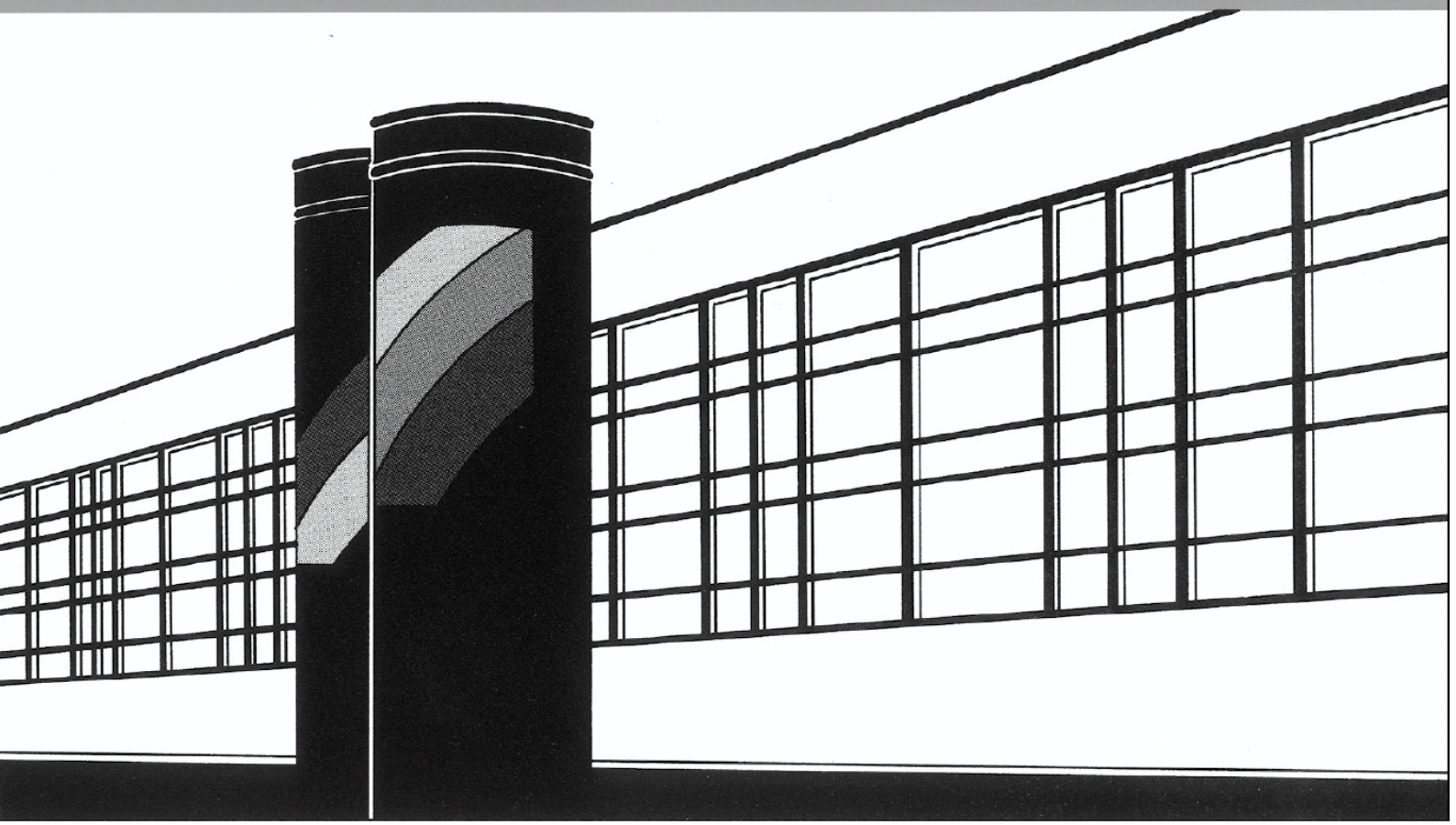
Universität Stuttgart

Institut für Wasser- und Umweltsystemmodellierung

Mitteilungen



Heft 306 Faizan Anwar

Spatial aspects of hydrological extremes: Description and simulation

Spatial aspects of hydrological extremes: Description and simulation

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

vorgelegt von

Faizan Anwar

aus Peshawar, Pakistan

Hauptberichter:	Prof. Dr. rer. nat. Dr.-Ing. András Bárdossy
Mitberichter:	Prof. Dr.-Ing. Erwin Zehe

Tag der mündlichen Prüfung: 01.02.2024

Institut für Wasser- und Umweltsystemmodellierung
der Universität Stuttgart
2024

Heft 306

Spatial aspects of hydrological extremes: Description and simulation

von
Dr.-Ing.
Faizan Anwar

Eigenverlag des Instituts für Wasser- und Umweltsystemmodellierung
der Universität Stuttgart

D93 Spatial aspects of hydrological extremes: Description and simulation

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

Anwar, Faizan:
Spatial aspects of hydrological extremes: Description and simulation, Universität Stuttgart. - Stuttgart: Institut für Wasser- und Umweltsystemmodellierung, 2024

(Mitteilungen Institut für Wasser- und Umweltsystemmodellierung, Universität Stuttgart: H. 306)
Zugl.: Stuttgart, Univ., Diss., 2024
ISBN 978-3-910293-10-6

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

Herausgegeben 2024 vom Eigenverlag des Instituts für Wasser- und Umweltsystem-modellierung

Druck: P+K Solutions GmbH & Co. KG, Stuttgart

Acknowledgements

It was a long journey to reach to this point for me. On the way, I met fellows in science that taught me valuable lessons in one way or another. Every day brought new experiences. Apart from science, I was also taught the German culture and language that made my time easy here in uncountable ways. It was truly a privilege. I cannot put into words how I grateful I actually feel, nevertheless, I will try.

To begin with, I would like to thank Prof. András Bárdossy whose supervision over the years taught me so much that when I look back at the time before I started as a Ph.D. student, it is as I am a different person. His open door and kind behavior made me always feel comfortable to ask him questions and get the answers that I needed and many times he proposed new approaches that yielded results better than what I had hoped for. I was not made to feel like a person who does know anything but rather a one that wants to learn. Even when his schedule did not allow it. He taught me the way to approach problems and develop skills in a generic way so that I could apply them to solve many problems. He gave me precious time, even when he was not healthy. He taught me how to review scientific papers and contribute to science behind the scenes. This is what I aspire to become now for others.

My special thanks to Prof. Erwin Zehe for agreeing to be my second supervisor. His useful comments in the first round of revisions helped to improve the readability of this text. Apart from that I really enjoyed our conversations at other times when we visited each other over the years. It is very interesting to meet people from other fields that are trying to solve the same problems. I hope that this is not the end of our relationship.

Looking back in time, it was Jochen with whom it all started when I came to him bored out of my mind in the Summer semester break of 2014 during my masters. He gave me some ideas to solve using Python back then and now, here I am. He prepared me during my masters to handle much larger problems and always had new ones to solve that helped hone my problem-solving and programming skills. He is someone that I can always go to in time of need. He listened to and encouraged me during difficult times and always did what he could to make my life easy.

Astrid is someone that binds us to the outside world and makes sure that we do not do anything stupid and land in trouble. She was always kind to me and helped me a lot to deal with the all the bureaucracy during my time here. In a way, I think that her life is way more difficult than mine, given the amount of people and paperwork she had to deal with for me, while I sat in my office without ever realizing what she had done for me until much later. I am grateful to her.

Gabriele is responsible for us students in the ENWAT doctoral program. She helped me

during my enrollment and time as a Ph.D. student. Moreover, she was always interested in what I had to say. Which is rare. I must say that the Ph.D. Seminar that she organizes twice a year for us to present our work, prepares us to interact with the outside world later and we get very useful comments on our work and presentation skills.

Wolfgang became my new boss after Prof. Bárdossy retired. What I particularly like about him are the comments that he gives on presentations of students and his ability to translate complicated ideas into understandable chunks. Much like Prof. Bárdossy, he is also very relaxed with his organization structure and does not force us to do pursue ideas that we are not particularly enthusiastic about. I hope to keep getting bosses like these in the future.

Of course, I was not always interacting with my superiors. Most of my social time was spent with my wonderful colleagues. I was not the one to interact much with people but it all changed for good during six months into the corona pandemic where I realized that I am only human and need my fellows to interact with, share ideas, learn new things and become a better person. So, I made friends, without realizing, with Masoud and Dhiraj who came to the office regularly despite the danger outside. One could only stay so long at home. With them, I learned many new things apart from science. We always shared the good and the bad news with each other. I gained valuable information about the outside world from the both of them. I will always miss our tea breaks.

I learned much from Dirk about Python during my time here. His gentle behavior is what sets him apart from many. I will always remember my discussions with him to be very pleasant and informative. One day, he took me and Masoud fishing. Even though the setup was rigged, I caught fish for the first time as he had prepared every thing for us.

Abbas is one of those people whose motivation to do things is what impresses me. He is always kind to everyone and is pleasant to talk to. Ning is another person whose motivation is what everyone should look up to. Next time someone says that Ph.D. is hard, I will always give the example of Ning, who despite the worst circumstances, persevered and finished her studies. I find her very shy but whenever we had discussions in our office, I always admired how honest she was about situations.

Finally, I would like to thank my former colleagues Ehsan, Tobias, Thomas, Sebastian, Mischa and Claudia. It was sad to see people leave one after another. Only a part of the large family remains as of now.

I must mention Ms. Heidi Drauschke from the university's administration. Dealing with Visa officers is never easy here as everyone from outside can confirm. It is true but in time of need, it was her that got me out of a pickle where I did not have any solutions. Due to which I am still here. I am grateful to her. You never know how a random act of kindness by a stranger can change your life in a significant manner.

I was not always at the university, of course. Outside, I got to be with Family Günther (Mutti, Vati, Opa und Kinder), who are my landlords. I cannot express how I felt when they accepted me as their tenant. They welcomed me, taught me the German language and culture, took me on excursions, showed to me the way to live and deal with the bureaucracy here, invited me to coffee, lunches and dinners. Without them, I cannot imagine how my

time would have been. Their generous behavior is what I will always remember. I am grateful to them.

My journey in science might would have come to an end had Markus not offered me a postdoc position. It was one of the few times that I was proud of myself when he casually asked me join his group after finishing my studies in Stuttgart over dinner in a seminar. I felt honored. I appreciate his kindness and relaxed and supportive behavior. Apart from that, Gabriele, Ye, and Christiane helped me transition to their group. I am grateful for their efforts and support and looking forward to our time together.

Finally, I would like to thank my beautiful and supportive wife and daughter. My wife stood by me while, I finished my studies very slowly and always believed in my work and was there when I needed her.

Contents

List of Figures

List of Tables

Abbreviations, Acronyms and Symbols

A

A_2	Theoretical order asymmetry in 2 dimensions as a scalar
A_3	Theoretical directional asymmetry in 2 dimensions as a scalar
A_d	Theoretical directional asymmetry in 2 dimensions as a scalar
AAFT	Amplitude Adjusted Fourier Transform
AD_{raw}	Empirical copula directional asymmetry in 2 dimensions as a scalar
AD_{max}	Maximum copula directional asymmetry in 2 dimensions as a scalar
AD	Directional asymmetry
AD^d	Copula directional asymmetry in d dimensions
AM	Available soil moisture
AO_{raw}	empirical order asymmetry in 2 dimensions as a scalar
AO_{max}	Maximum order asymmetry in 2 dimensions as a scalar
AO	Order asymmetry
AO^d	Copula order asymmetry in d dimensions
AR	Auto-Regressive process
ARIMA	Auto-Regressive-Integrated-Moving-Average process
ARMA	Auto-Regressive-Moving-Average process
AWGEN-2D	An arbitrary waveform generator

B

β	Relative runoff exponent
BW	Baden-Würrttemberg

C

c	Copula density
C	A copula or combinations
CC	Runoff to discharge conversion constant
CDF	Cumulative distribution function
CLIMGEN	A spatial climate scenario generator
CLIGEN	A stochastic weather generator
CLIMA	Climate alliance model
CM	Degree-day factor for snow-melt
$CMCORR$	Cumulative mean correlation periodogram
$CMPOW$	Cumulative power spectrum in 2 dimensions
$CMPOW^d$	Cumulative power spectrum in d dimensions

D

d	Dimensions
DE	Differential evolution
ΔPHS	Phase difference spectrum
DWD	Deutscher Wetter Dienst

E

EDK	External Drift Kriging
ET	Actual evapotranspiration

F

FC	Field capacity
FFA	Flood frequency analysis
FFTMA	Fast Fourier Transform Moving Average algorithm
FT	Fourier Transform

G

Grade	Non-exceedence probability

H

H_{raw}	Empirical Shannon's copula entropy
H_{max}	Maximum Shannon's copula entropy
H	Normalized Shannon's copula entropy
HBV	Hydrologiska Byråns Vattenbalansavdelning, a conceptual rainfall-runoff model

I

IAAFT	Iterative Amplitude Adjusted Fourier transform algorithm
IAAFT-PSC1	Iterative Amplitude Adjusted Fourier Transform algorithm that conserves Pearson's and Spearman's correlation (variant 1)
IAAFT-PSC2	Iterative Amplitude Adjusted Fourier Transform algorithm that conserves Pearson's and Spearman's correlation (variant 2)

K

K_d	Upper to lower reservoir seepage coefficient
K_{ll}	Upper reservoir runoff coefficient
K_{uu}	Upper reservoir upper outlet runoff coefficient
K_{ul}	Upper reservoir lower outlet runoff coefficient

L

LARSWG	A stochastic weather generator
LnNSE	Nash-Sutcliffe efficiency based on the logarithm of values
LP	Liquid precipitation
LRO	Lower reservoir runoff
$LRST$	Lower reservoir storage
LUBW	Landesanstalt für Umwelt Baden-Württemberg

M

MA	Moving-Average process
MAG	Magnitude spectrum
ME	Snow-melt
MI	Mutual information copula entropy
MIAAFT	Multivariate Iterative Amplitude Adjusted Fourier Transform algorithm
$MNPHS$	Mean phase correlation spectrum
MMIAAFT	Modified Multivariate Iterative Amplitude Adjusted Fourier Transform algorithm
MLR	Multiple linear regression
$MNPOW$	Mean power periodogram
$MXCORR$	Maximum correlation in terms of the Fourier transform
$MXPOW$	Maximum power spectrum

N

NSE	Nash-Sutcliffe Efficiency

P

ϕ	Phase
$\prod$	Product
p	Pearson's correlation
P	Probability, Precipitation
PA	Phase Annealing
PE	Potential evapotranspiration
PHS	Phase spectrum
POW	Power spectrum
PR	Phase Randomization
PWP	Permanent wilting point

Q

QQ	Quantile-Quantile
QS	Simulated discharge

R

ρ_p	Pearson's correlation in 2 dimensions
ρ_s	Spearman's or rank correlation in 2 dimensions
ρ_d	Spearman's correlation in d dimensions
$\tilde{\rho}_d$	Spearman's correlation in d dimensions
RN	Runoff

S

s	Spearman's or rank correlation
σ	Standard deviation
SA	Simulated Annealing
SARIMA	Seasonal-Auto-Regressive-Integrated-Moving-Average process
SM	Soil moisture
SN	Snow depth

T

T	Temperature, timestep
T_0	Initial temperature
TT	Threshold temperature

U

UT	Upper reservoir upper outlet threshold depth
$URLO$	Upper reservoir lower outlet runoff
$URLR$	Upper to lower reservoir seepage
$URST$	Upper reservoir storage
$URUO$	Upper reservoir upper outlet runoff

W

WeaGETS	A stochastic weather generator
WGEN	A stochastic weather generator

Summary

Hydrological extremes such as large floods happen rarely but whenever they do, they cause extensive damage to life and property, and major disruption of activities. In order to defend ourselves against the possible dangers, structures such as dams, flood retention reservoirs, floodwalls and levees are built. As per the definition of engineering, science used to solve problems that humans (may) face in an economic manner, the size of such defense structures and evacuation plans need to be developed in an economical manner as well. To do so, it is important to understand how large the magnitude of these floods may be in near or far future.

Understanding the phenomenon occurring in the atmosphere and the catchments underneath them is not trivial and has been a subject of research for more than a century. Even with state-of-the-art knowledge, it is not uncommon to encounter new cases where disaster took the concerned authorities and individuals in an area by surprise. This leads to the question, what is still missing in the techniques that are used to understand and predict large scale precipitation that leads to large floods?

This thesis tries to address one of the important and less investigated aspects of hydrological extremes, namely the description of the structure of dependence of multiple points in space-time of variables such as precipitation before and during a large flood event. And additionally, the possibility of generating this structure synthetically via numerical simulation.

Consider the following problem. The catchment area of a certain river is monitored for precipitation, temperature, land-cover, soil type, geology, evapotranspiration and finally, river discharge. The matter of interest is the amount of the resulting flow in the river due to liquid precipitation or snow-melt. A rainfall-runoff model (physically-based or conceptual) is chosen to represent the interactions between the various variables to model river discharge. Two main problems, among many, exist. Firstly, the model is highly unlikely to represent the catchment in a manner that is complete enough to obtain usable. Secondly, the catchment is not monitored for precipitation/temperature at each point. For all the remaining area except for the monitoring points, these variables have to be interpolated. To obtain correct estimates at locations where no measurement were made, geostatistical techniques such as Kriging and its various variants are used. And, this is where the problem starts. For instance temperature varies very smoothly over large areas and decreases, mostly, as the altitude increases at a near constant rate. Such a behavior can be easily interpolated by Kriging that takes the effect of elevation into account. On the other hand, precipitation follows no such behavior, especially at the shorter temporal scales e.g., minutely or hourly scales. If the aim is to model large scale precipitation then it is imperative that such behavior be understood first and then modeled. This would be the ideal case that may not be possible in practice.

The next possible research avenue could be to investigate the ground-truth i.e., precipitation measurements at the monitoring locations. Mainly, the question that needs answering is that how does the precipitation at multiple points behave before and during a large river discharge event? Is precipitation occurring at all gauges, a subset or none? It can come as a surprise that river flows show increase without any precipitation being observed at the monitoring locations liquid- or snow-melt- wise. This may very well be the case because, the entire catchment is not monitored. Possibility of such a scenario grows as the density of monitoring locations becomes smaller.

It is not trivial to study the problem of precipitation retrospectively or even in real-time due to sparse coverage of the catchment. In some cases, this problem could be studied by analyzing discharge time series recorded at multiple locations. Consider the following problem of stream flows in multiple sub-catchments inside a catchment. Flow in the stream is the cumulative response of the catchment to precipitation while accounting for all the possible interactions. Each sub-catchment can be considered as one large precipitation measurement gauge. If discharge of different headwater sub-catchments of the entire catchment can be monitored, then these can also be used to understand the dynamics of large floods. For example, the timings of peaks and their relative magnitudes can be analyzed. It can also be studied whether all the large floods at the mouth of the catchment are due to all of them yielding large flows or only a subset of them. Such information can be used to design worst-case scenarios.

To describe the multivariate dependence in space-time, copulas were utilized. These provide the advantage that the modeling process is translated from the marginal to rank space where the relative order rather than the magnitudes of values are of importance. Such a construction helps with mitigating the effects of erroneous values or outliers to a large extent as large river flow values are not physically measured but are indirectly derived from a rating curve.

Strength of dependence between two variables is quantified, traditionally, by calculating their Pearson's correlation. In the rank space, this correlation is called the Spearman's or rank correlation. Furthermore, copulas also offer the possibility to differentiate between the dependence of high and low values relative to each other. These are described by the order and directional asymmetries. The order asymmetry describes the relative mismatch between the relative occurrence of low values appearing together with respect to high values appearing together. Such a behavior can be analyzed for the time series with itself or for two time series recorded at different locations.

As one major goal is to understand the dependence in space-time by analyzing variables at multiple locations, it would be more optimal to describe dependence of more than two points i.e., multivariate dependence.

The next step was to develop algorithms that can simulate synthetic time series that have prescribed properties. In that aspect, Fourier transform-based algorithms were utilized. These mainly work by moving time series to the frequency domain where phases, that control timing, are manipulated only while keeping the magnitudes, that control variance, of all frequencies constant. Their main advantage is their speed and the conservation of the

auto- and cross-correlations. However, these suffer from the disadvantage that other properties such as the asymmetric dependence is not present necessarily in the resulting time series. The long existing algorithms Phase Randomization, Amplitude Adjusted Fourier transform, Iterative Amplitude Adjusted Fourier transform were used as benchmarks. The series resulting from these are called *surrogates* or *surrogate time series*.

To address the drawback of the missing asymmetric dependence in the surrogates, the existing *Phase Annealing* algorithm was selected. The main advantage of this algorithm is that it can incorporate arbitrary constraints, e.g., point and/or areal, while simulating. However, a measure that not only describes the overall asymmetry but the variability of it in space-time was still missing. Therefore, a new measure that incorporates such behavior was formulated by taking the values as a time series that describe the order and directional asymmetry in space-time. As an objective function, the power spectrum (the square of the absolute values of the Fourier frequencies) of the asymmetry time series was matched for the observed and simulated series by minimizing their squared differences in the Phase Annealing algorithm. This resulted in series that exhibited correct asymmetric space-time behavior for both precipitation and discharge. As the order and directional asymmetries are defined for pairs only, the problem of considering multivariate asymmetry still remained.

Additionally, two new algorithms were also proposed, IAAFT-PSC1 and IAAFT-PSC2, that can achieve some aspects of the asymmetric behavior of dependence in new time series rather quickly. These were based on the mixing of time series of the observed values and their ranks in a certain proportion. The exact proportions were determined by using the Simulated Annealing algorithm. These resulted in time series having Pearson's and Spearman's correlations being closer to the observed ones in space-time as compared to the benchmark algorithms that can conserve only one of them.

To test, independently, whether various algorithms produce overall high-dimensional dependence in space-time similar to that of the observed, asymmetries in d-dimensions were also developed. These are collectively called d-asymmetries. The approach used was geometric, where each dimension among d was assigned the value of asymmetry based on the mean of the cube of distances of points from a given line that passed through the opposing corners of a unit hypercube. This resulted in vectors for each set of time series which could then be compared to the vectors of the observed, and evaluated for the difference.

The d-asymmetries describe the overall dependence of points in a high-dimensional space while the question of describing the behavior during large flood events still remained. For this issue, measures in the frequency domain were developed because they are capable of representing the temporal behavior in a more complete and transparent manner. Firstly, the maximum correlation ($MXCORR$) was defined which is akin to a worst-case scenario where all the values of similar magnitudes synchronize in a set of time series. As a side effect, it also resulted in a spectrum $CMPOW^d$ that showed at which time scales, e.g., daily, seasonal or annual, the various time series can synchronize with each other. Secondly and more importantly, another measure $CMCORR_p^d$ was developed that described the behavior of existing time series in higher dimensions by incorporating the differences of phases into the former measure. Such a construction helps in evaluating the multivariate dependence in a more complete manner by taking into account the real behavior while using various

algorithms to simulate surrogates.

Yet another method was also used to simulate time series of discharges indirectly by creating synthetic rainfall-runoff model inputs, namely precipitation, temperature and potential evapotranspiration that were then fed to rainfall-runoff models that were calibrated already using observed inputs. The resulting time series of discharges were then compared to observed and the calibrated model discharges. The aim was to test whether it makes more sense to use synthetic model inputs and then obtain discharge time series or by simulating discharge time series directly. The rationale for such an approach was that river flows are a result of a very deterministic process that is difficult to mimic using statistical approaches. On the other hand, precipitation and temperature proved to be much easier to simulate.

Time series of daily discharge, precipitation, temperature and potential evapotranspiration and hourly precipitation were simulated using all the aforementioned techniques and with various objective functions for the case of Phase Annealing.

To evaluate whether the time series have realistic properties, apart from the existing pairwise and the new high-dimensional measures; two more tests were also done. The first test was that of the distributions of summation of time series in space (spatial-sums) while the second test was that of the return periods for precipitation and discharge. For the case of spatial-sums, where a new time series was obtained in which the value at each time step was the sum of all the values of all the series at the same time step, the upper tails were compared to the observed distributions. The second test of return periods was built in such a manner where a group of four time series was taken and the probabilities of only three of them having floods of return periods of two years or more, against all of them having floods of return periods of two or more years was compared for observed and simulated cases.

Finally, the results showed a mixture of behaviors. The new methods did outperform the existing ones when considering the various properties that are unique to a given type of variable. However, the spatial-sum distributions were simulated very well by almost all methods for all variables. This was evidenced by the fact that the observed distributions were enveloped very well by distributions of the simulations. These envelopes are analogues of uncertainty bounds on distributions. This was the case for daily discharge, precipitation and model discharges. For the case of the hourly precipitation, only one method, namely IAAFT-PSC2, performed acceptably. This signifies that simulating at finer temporal scales still needs more research, where the behavior is likely to be very different compared to that of the daily. It was also noted that complete reproduction of the various evaluation metrics was not achieved by any of the simulation methods and the results did show under- or over-estimation of strength for many of them.

By examining simulated time series that perform well for the case of extreme values, in terms of synchronization in time and magnitudes, it was found out that the only property that all those had in common was that their $CMCORR_p^d$ spectra were close to those of the observed, especially in the high frequency ranges. This was not the case for the other measures that focused on the overall behavior.

Zusammenfassung

Hydrologische Extremereignisse wie große Überschwemmungen kommen zwar selten vor, aber wenn sie auftreten, verursachen sie große Schäden an Leben und Eigentum sowie eine erhebliche Beeinträchtigung der Aktivitäten. Um sich gegen die möglichen Gefahren zu schützen, werden Bauwerke wie Dämme, Hochwasserrückhaltebecken, Hochwassermauern und Deiche errichtet. Gemäß der Definition des Begriffs İngenieurwesenäls Wissenschaft, die dazu dient, Probleme, mit denen der Mensch konfrontiert wird, auf sparsame Weise zu lösen, müssen auch die Größe solcher Verteidigungsanlagen und Evakuierungspläne auf sparsame Weise entwickelt werden. Dazu ist es wichtig zu verstehen, wie groß das Ausmaß dieser Überschwemmungen in naher oder ferner Zukunft sein kann.

Das Verständnis der Phänomene, die in der Atmosphäre und den darunter liegenden Einzugsgebieten auftreten, ist nicht trivial und wird seit mehr als einem Jahrhundert erforscht. Selbst mit den neuesten Erkenntnissen ist es nicht ungewöhnlich, dass neue Fälle auftreten, bei denen die Katastrophe die betroffenen Behörden und Personen in einem Gebiet überrascht hat. Dies wirft die Frage auf, was in den Methoden, die zum Verständnis und zur Vorhersage großflächiger Niederschläge, die zu großen Überschwemmungen führen, eingesetzt werden, noch fehlt.

In dieser Arbeit wird versucht, einen der wichtigen und wenig erforschten Aspekte hydrologischer Extreme anzugehen, nämlich die Beschreibung der Struktur der Abhängigkeit mehrerer Punkte in der Raum-Zeit von Variablen wie dem Niederschlag vor und während eines großen Hochwasserereignisses. Und zusätzlich die Möglichkeit, diese Struktur synthetisch durch numerische Simulationen zu erzeugen.

Betrachten Sie das folgende Problem. Im Einzugsgebiet eines bestimmten Flusses werden Niederschlag, Temperatur, Bodenbedeckung, Bodenart, Geologie, Evapotranspiration und schließlich der Abfluss des Flusses beobachtet. Von Interesse ist die Menge des resultierenden Abflusses im Fluss aufgrund von flüssigem Niederschlag oder Schneeschmelze. Ein (physikalisch basiertes oder konzeptionelles) Niederschlags-Abfluss-Modell wird gewählt, um die Wechselwirkungen zwischen den verschiedenen Variablen darzustellen und den Abfluss eines Flusses zu modellieren. Dabei gibt es zwei Hauptprobleme unter vielen anderen. Erstens ist es sehr unwahrscheinlich, dass das Modell das Einzugsgebiet vollständig genug abbildet, um brauchbare Ergebnisse zu erzielen. Zweitens wird das Einzugsgebiet nicht an jedem Punkt hinsichtlich Niederschlag/Temperatur überwacht. Für das gesamte übrige Gebiet mit Ausnahme der Überwachungspunkte müssen diese Variablen interpoliert werden. Um an Orten, an denen keine Messungen vorgenommen wurden, Schätzungen zu erhalten, werden geostatistische Verfahren wie Kriging und seine verschiedenen Varianten eingesetzt. Und genau hier beginnt das Problem. Die Temperatur schwankt, beispielsweise, über große Gebiete hinweg sehr gleichmäßig und nimmt mit zunehmender Höhe meist mit

einer nahezu konstanten Rate ab. Ein solches Verhalten kann leicht durch Kriging interpoliert werden, das den Effekt der Höhe berücksichtigt. Andererseits folgt der Niederschlag keinem solchen Verhalten, insbesondere auf kürzeren Zeitskalen, z. B. auf Minuten- oder Stundenskalen. Wenn das Ziel darin besteht, großräumige Niederschläge zu modellieren, ist es zwingend erforderlich, dieses Verhalten zunächst zu verstehen und dann zu modellieren. Dies wäre der Idealfall, der in der Praxis nicht immer möglich ist.

Der nächste mögliche Forschungsansatz könnte darin bestehen, die Bodenwahrheit, d. h. die Niederschlagsmessungen an den Überwachungsstandorten, zu untersuchen. Die Frage, die es zu beantworten gilt, lautet vor allem: Wie verhält sich der Niederschlag an mehreren Punkten vor und während eines großen Abflussereignisses? Fallen die Niederschläge an allen Pegeln, an einer Teilmenge oder an keinem? Es kann überraschen, dass der Abfluss eines Flusses ansteigt, ohne dass an den Messstellen Niederschlag in Form von Flüssigkeit oder Schneeschmelze zu verzeichnen ist. Dies kann sehr wohl der Fall sein, weil nicht das gesamte Einzugsgebiet überwacht wird. Die Möglichkeit eines solchen Szenarios wächst, je geringer die Dichte der Messstellen ist.

Es ist nicht trivial, das Problem der Niederschläge rückwirkend oder sogar in Echtzeit zu untersuchen, da das Einzugsgebiet nur spärlich erfasst ist. In einigen Fällen kann dieses Problem durch die Analyse von Abflusszeitreihen untersucht werden, die an mehreren Orten beobachtet wurden. Betrachten wir das folgende Problem der Abflüsse in mehreren Teileinzugsgebieten innerhalb eines Einzugsgebiets. Der Abfluss im Bach ist die kumulative Reaktion des Einzugsgebiets auf den Niederschlag, wobei alle möglichen Wechselwirkungen berücksichtigt werden. Jedes Teileinzugsgebiet kann als ein großes Niederschlagsmessgerät betrachtet werden. Wenn die Abflüsse verschiedener Teileinzugsgebiete des gesamten Einzugsgebiets überwacht werden können, lassen sich diese auch dazu nutzen, die Dynamik großer Hochwasser zu verstehen. So können zum Beispiel die Zeitpunkte der Spitzen und ihre relativen Größenordnungen analysiert werden. Es kann auch untersucht werden, ob alle großen Überschwemmungen an der Mündung des Einzugsgebiets darauf zurückzuführen sind, dass alle Einzugsgebiete große Abflüsse erzeugen, oder nur eine Teilmenge davon. Diese Informationen können für die Entwicklung von Worst-Case-Szenarien verwendet werden.

Um die multivariate Abhängigkeit in der Raum-Zeit zu beschreiben, wurden Copulas verwendet. Diese bieten den Vorteil, dass der Modellierungsprozess vom Rand- in den Rangraum übertragen wird, in dem die relative Reihenfolge und nicht die Größenordnungen der Werte von Bedeutung sind. Eine solche Konstruktion trägt dazu bei, die Auswirkungen von fehlerhaften Werten oder Ausreißern weitgehend abzuschwächen, da große Abflusswerte nicht physisch gemessen, sondern indirekt aus einer Bewertungskurve abgeleitet werden.

Die Stärke der Abhängigkeit zwischen zwei Variablen wird traditionell durch die Berechnung ihrer Pearson-Korrelation quantifiziert. Im Rangraum wird diese Korrelation als Spearman's oder Rangkorrelation bezeichnet. Darüber hinaus bieten Copulas auch die Möglichkeit, zwischen der Abhängigkeit von hohen und niedrigen Werten voneinander zu differenzieren. Diese werden durch die Ordnungs- und Richtungsasymmetrien beschrieben. Die Ordnungsasymmetrie beschreibt die relative Ungleichheit zwischen dem

relativen Auftreten von zusammen auftretenden niedrigen Werten gegenüber zusammen auftretenden hohen Werten. Ein solches Verhalten kann für die Zeitreihe an sich oder für zwei an unterschiedlichen Orten beobachtete Zeitreihen analysiert werden.

Da ein Hauptziel darin besteht, die Abhängigkeit in der Raum-Zeit durch die Analyse von Variablen an mehreren Orten zu verstehen, wäre es optimaler, die Abhängigkeit von mehr als zwei Punkten, d. h. die multivariate Abhängigkeit, zu beschreiben.

Der nächste Schritt bestand darin, Algorithmen zu entwickeln, die synthetische Zeitreihen mit vorgegebenen Eigenschaften simulieren können. In diesem Zusammenhang wurden Algorithmen auf der Grundlage der Fourier-Transformation eingesetzt. Bei diesen Algorithmen werden die Zeitreihen hauptsächlich in den Frequenzbereich verschoben, wobei nur die Phasen, die das Timing steuern, manipuliert werden, während die Magnitude, die die Varianz steuern, für alle Frequenzen konstant bleiben. Ihr Hauptvorteil ist ihre Geschwindigkeit und die Erhaltung der Auto- und Kreuzkorrelationen. Sie haben jedoch den Nachteil, dass andere Eigenschaften wie die asymmetrische Abhängigkeit nicht unbedingt in den resultierenden Zeitreihen vorhanden sind. Die seit langem existierenden Algorithmen Phasenrandomisierung, Amplitudenangepasste Fourier-Transformation und Iterative Amplitudenangepasste Fourier-Transformation wurden als Benchmarks verwendet. Die daraus resultierenden Reihen werden als *Surrogate* oder *Surrogat-Zeitreihen* bezeichnet.

Um den Nachteil der fehlenden asymmetrischen Abhängigkeit in den Surrogaten zu beheben, wurde der existierende Algorithmus *Phase Annealing* gewählt. Der Hauptvorteil dieses Algorithmus besteht darin, dass er während der Simulation beliebige Beschränkungen, z. B. punkt- und/oder flächenbezogen, berücksichtigen kann. Es fehlte jedoch noch ein Maß, das nicht nur die Gesamtasymmetrie, sondern auch die Variabilität der Asymmetrie in der Raum-Zeit beschreibt. Daher wurde ein neues Maß formuliert, das ein solches Verhalten berücksichtigt, indem die Werte als Zeitreihe genommen wurden, die die Ordnung und die Richtungsasymmetrie in der Raumzeit beschreiben. Als Zielfunktion wurde das Power Spektrum (das Quadrat der absoluten Werte der Fourier-Frequenzen) der Asymmetrie-Zeitreihen für die beobachteten und simulierten Reihen durch Minimierung ihrer quadratischen Differenzen im Phase Annealing-Algorithmus angepasst. Dies führte zu Reihen, die sowohl für den Niederschlag als auch für den Abfluss ein korrektes asymmetrisches Raum-Zeit-Verhalten aufwiesen. Da die Ordnungs- und Richtungsasymmetrien nur für Paare definiert sind, blieb das Problem der Berücksichtigung multivariater Asymmetrien bestehen.

Außerdem wurden zwei neue Algorithmen vorgeschlagen, IAAFT-PSC1 und IAAFT-PSC2, mit denen einige Aspekte des asymmetrischen Verhaltens der Abhängigkeit in neuen Zeitreihen relativ schnell erreicht werden können. Diese basierten auf der Vermischung von Zeitreihen der beobachteten Werte und ihrer Ränge in einem bestimmten Verhältnis. Die genauen Proportionen wurden mit Hilfe des Algorithmus Simulated Annealing bestimmt. Dies führte dazu, dass die Zeitreihen mit Pearson's und Spearman's Korrelationen näher an den beobachteten Korrelationen in der Raum-Zeit lagen als bei den Benchmark-Algorithmen, die nur eine der beiden Korrelationen beibehalten können.

Um unabhängig davon zu testen, ob verschiedene Algorithmen insgesamt eine hochdi-

mensionale Abhängigkeit in der Raumzeit erzeugen, die der beobachteten ähnelt, wurden auch Asymmetrien in d-Dimensionen entwickelt. Diese werden zusammenfassend als d-Asymmetrien bezeichnet. Es wurde ein geometrischer Ansatz verwendet, bei dem jeder d-Dimension ein Asymmetriewert zugewiesen wurde, der auf dem Mittelwert des Kubus der Abstände von Punkten zu einer bestimmten Linie basiert, die durch die gegenüberliegenden Ecken eines Einheitshyperwürfels verläuft. Daraus ergaben sich Vektoren für jeden Satz von Zeitreihen, die dann mit den Vektoren der beobachteten verglichen und auf die Differenz hin ausgewertet werden konnten.

Die d-Asymmetrien beschreiben die Gesamtabhängigkeit von Punkten in einem hochdimensionalen Raum, während die Frage nach der Beschreibung des Verhaltens bei großen Hochwasserereignissen noch offen blieb. Für diese Frage wurden Maße im Frequenzbereich entwickelt, da sie das zeitliche Verhalten vollständiger und transparenter darstellen können. Zunächst wurde die maximale Korrelation ($MXCORR$) definiert, die einem Worst-Case-Szenario entspricht, bei dem sich alle Werte ähnlicher Größenordnung in einer Reihe von Zeitreihen synchronisieren. Als Nebeneffekt ergab sich daraus auch ein Spektrum $CMPOW^d$, das zeigt, auf welchen Zeitskalen, z. B. täglich, saisonal oder jährlich, die verschiedenen Zeitreihen miteinander synchronisieren können. Zweitens, und das ist noch wichtiger, wurde ein weiteres Maß $CMCORR_p^d$ entwickelt, das das Verhalten bestehender Zeitreihen in höheren Dimensionen beschreibt, indem die Unterschiede der Phasen in das erste Maß einbezogen werden. Eine solche Konstruktion hilft bei der Bewertung der multivariaten Abhängigkeit auf eine vollständigere Art und Weise, indem das reale Verhalten bei der Verwendung verschiedener Algorithmen zur Simulation von Surrogaten berücksichtigt wird.

Eine weitere Methode wurde verwendet, um Zeitreihen von Abflüssen indirekt zu simulieren, indem synthetische Niederschlag-Abfluss-Modell-Inputs, nämlich Niederschlag, Temperatur und potenzielle Evapotranspiration, erzeugt wurden, die dann in Niederschlag-Abfluss-Modelle eingespeist wurden, die bereits anhand beobachteter Inputs kalibriert worden waren. Die daraus resultierenden Zeitreihen der Abflüsse wurden dann mit den beobachteten und den kalibrierten Modellabflüssen verglichen. Ziel war es, zu prüfen, ob es sinnvoller ist, synthetische Modellinputs zu verwenden und dann Abflusszeitreihen zu erhalten oder die Abflusszeitreihen direkt zu simulieren. Die Begründung für einen solchen Ansatz war, dass die Abflüsse in Flüssen das Ergebnis eines sehr deterministischen Prozesses sind, der mit statistischen Ansätzen nur schwer nachgeahmt werden kann. Andererseits erwiesen sich Niederschlag und Temperatur als wesentlich einfacher zu simulieren.

Die Zeitreihen des täglichen Abflusses, des Niederschlags, der Temperatur und der potenziellen Evapotranspiration sowie des stündlichen Niederschlags wurden mit allen oben genannten Methoden und mit verschiedenen Zielfunktionen für den Fall des Phase Annealing simuliert.

Um zu beurteilen, ob die Zeitreihen realistische Eigenschaften haben, wurden neben den bestehenden paarweisen und den neuen hochdimensionalen Maßen zwei weitere Tests durchgeführt. Der erste Test betraf die Verteilungen der Summierung von Zeitreihen im Raum (Raumsummen), der zweite Test die Wiederkehrperioden für Niederschlag und Abfluss. Für den Fall der räumlichen Summen, bei dem eine neue Zeitreihe erhalten wurde, bei der

der Wert bei jedem Zeitschritt die Summe aller Werte aller Reihen bei demselben Zeitschritt war, wurden die upper-tails mit den beobachteten Verteilungen verglichen. Der zweite Test der Wiederkehrperioden wurde so aufgebaut, dass eine Gruppe von vier Zeitreihen genommen wurde und die Wahrscheinlichkeiten, dass nur drei von ihnen Überschwemmungen mit Wiederkehrperioden von zwei Jahren oder mehr haben, mit allen anderen verglichen wurden, die Überschwemmungen mit Wiederkehrperioden von zwei Jahren oder mehr haben, für beobachtete und simulierte Fälle.

Schließlich zeigten die Ergebnisse eine Mischung aus verschiedenen Verhaltensweisen. Die neuen Methoden übertrafen die bestehenden, wenn man die verschiedenen Eigenschaften berücksichtigt, die für einen bestimmten Variablentyp typisch sind. Die Raumsummenverteilungen wurden jedoch von fast allen Methoden für alle Variablen sehr gut simuliert. Dies zeigte sich daran, dass die beobachteten Verteilungen sehr gut von den Verteilungen der Simulationen umhüllt wurden. Diese Umhüllungen sind Analoga zu den Unsicherheitsgrenzen von Verteilungen. Dies war der Fall für den täglichen Abfluss, den Niederschlag und die Modellabflüsse. Im Falle des stündlichen Niederschlags schnitt nur eine Methode, nämlich IAAFT-PSC2, akzeptabel ab. Dies bedeutet, dass die Simulation auf feineren zeitlichen Skalen noch weiter erforscht werden muss, da sich das Verhalten hier wahrscheinlich sehr von dem des täglichen Abflusses unterscheidet. Es wurde auch festgestellt, dass keine der Simulationsmethoden eine vollständige Reproduktion der verschiedenen Bewertungsmetriken erreichte und die Ergebnisse für viele von ihnen eine Unter- oder Überschätzung der Stärke zeigten.

Bei der Untersuchung simulierter Zeitreihen, die im Falle von Extremwerten eine gute Leistung hinsichtlich der zeitlichen Synchronisation und der Größenordnungen erbringen, wurde festgestellt, dass die einzige Eigenschaft, die alle gemeinsam hatten, darin bestand, dass ihre $CMCORR_p^d$-Spektren denen der beobachteten nahe kamen, insbesondere in den hohen Frequenzbereichen. Dies war bei den anderen Messungen, die sich auf das Gesamtverhalten konzentrierten, nicht der Fall.

1 Introduction

Uncertainty exists all around everywhere. Be it observations, measurements or routine decisions. Things almost never turn out as planned. In Hydrology measurement uncertainty, uncertainty due to small observation samples, low density monitoring networks and model output uncertainty is so common that they are not even considered, and even if they are, it is based on some assumptions that themselves are not necessarily fulfilled i.e., uncertain.

1.1 An example of uncertainty

Consider the case of stream discharge measurement. The standard practice is to measure velocity using various types of fluid velocity meters (current meters, pressure meters or ultrasonic sensors) along with water depth and cross-section measurements at different times of the year at a given location of interest. Velocity and cross-sectional area are then combined to yield discharge. Many observations are then used to derive a function that transforms depth of water to discharge for convenience, the so called stage-discharge or rating curve. Consider the number of error sources in the measurements that may affect the final value of the discharge. An analog/mechanical current meter converts the number of revolutions to velocity. The number of revolutions per unit time relate to velocity which in itself is also a fitted-function. A digital meter normally relies on the speed of sound using a transmitter and a receiver. How accurate are these? The friction increases over time due to corrosion inside the bearings or foreign particles entering the analog device. The speed of sound depends on the density of water for the digital flow meter. Depending on how much the flow meter is not parallel to the flow velocity introduces another error. The river bed is either soft or ragged but almost never straight except in seldom cases. The cross-section measurements are made at discrete intervals, where it is assumed that the riverbed uniformly increases/decreases in depth from one point to the next. Even at each point, the depth measurement itself is uncertain. The same applies to axial measurements along the cross-section. The velocity profile is curved while moving towards the surface. It is assumed constant for every polygon drawn inside the cross-section. All measurements are not made at the same instant of time. Finally, at least eight measurement errors are introduced into the calculations of stream discharge. All of these contribute to the final flow rate and consequently the rating curve. That is not considering if a surface roughness factor, e.g., due to vegetation, is used which in itself is highly uncertain. This final rating curve is used for obtaining hourly/daily discharge in the future. Later in time, the river bed might have changed by erosion or deposition. In the case of a major flood taking place after the fitting of the rating curve, how certain is it that the river bed kept its form similar to how it was before the flood? Similarly, the bed and banks may be affected by seasonal vegetation.

Another problem is due to the shape of the cross-section. It usually increases in width as distance to surface decreases, relative to the bed. An error of one unit at low stage produces less absolute volume error as compared to a higher stage. Finally, the largest source of error occurs when the depth of flow at a given time exceeds the range of the stage-discharge curve. There the function is simply used to extrapolate. For such flows, there is no guarantee that the fitted function holds outside its bounds. Later on during rainfall-runoff modeling, how often are the effects of all these possible errors considered while finding model parameters to the given data? To begin with, such information is never supplied with discharge time series. A similar case could be made for other environmental variables such as precipitation, temperature, evapotranspiration, land cover, geology, to name a few. The goal here is not criticism but rather the drawing of awareness towards the negligence of uncertainty.

Coming from field to a modeling environment; decisions about future actions are based on the present knowledge. In hydrology, similar to many other fields of science, models are used to help with the decision making. These models are calibrated using a finite length input dataset comprised of relevant variables. The model output uncertainty is affected by various factors. Firstly, the models themselves are coarse representations of reality. Different model parameters will produce different outputs. It may also be the case that dissimilar model parameters yield very similar type of model performances (sum of squared differences, for example), the so called *Equifinality*. Regarding inputs, there are two main types of problems. One is the aforementioned measurement error while the other being the length of the input series (e.g., the number of time steps of daily precipitation data). It should come as no surprise that using different time periods to calibrate the same model will result in different model parameters, which is also a function of measurement error indirectly to some extent. Generally speaking, using shorter time periods for calibration result in over-fitting i.e., the model performs well on calibration but worse on validation compared to the case if the input series were of a longer length. Another important aspect to keep in mind is that of the environment. It evolves continuously in time and space and therefore, using inputs from different time periods produce different model parameters up on calibration.

All the concerns mentioned above have to be dealt with, to make better decisions for the future. The sources of uncertainty will always remain to a certain degree while the understanding of the environment increases with time that in turn allows for creation of better models. However, the quality of their output is still linked to the quality and quantity of the inputs. The painted grim picture doesn't mean that hope should be given up. A more reasonable approach is to include confidence bounds on model outputs. For example, prediction of tomorrow's discharge in some river is expected to be 10 $m^3 \cdot s^{-1} \pm 1.5\ m^3 \cdot s^{-1}$. Prediction in such a form allows for the preparation of best and worst case scenarios in a much better manner depending on how wide the bounds of uncertainty are on the prediction. This thesis aims to address a similar problem.

1.2 The Objective

During decision making, suppose the case where the risk that a certain high precipitation/floods (extreme events) of given magnitudes occur simultaneously at multiple loca-

tions of interest has to be evaluated. Assuming that some observations from the past exist. Using them, with various models in practice, the probability of such a scenario can be calculated. The problem that needs to be addressed next is that of the uncertainty of the prediction. As the probability of an event becomes smaller which is inversely related to the magnitude of the event, the uncertainty becomes larger. To address this, existing models (e.g., copulas) that provide confidence bounds for all values. The objective of this thesis is to extend the knowledge on how these confidence bounds can be estimated in a better manner without making too many assumptions on the behavior of the variable in space-time. Mainly, the focus would be more on methods that are capable of simulating the spatial and temporal dependence of extreme values. Simulating values at a point that go beyond what was observed is not handled directly but rather the case of the sum of values considering multiple points that go beyond the sum of what was observed is dealt with.

Specifically, consider a catchment that has d rivers inside it, whose values are monitored at some pre-specified temporal resolution Δt. The discharge at the mouth of the catchment is the sum of the discharges of the said d sub-catchments. There are precipitation, temperature and potential evapotranspiration time series also available for all sub-catchments. Additionally, lumped rainfall-runoff models for each of the sub-catchments are calibrated using observed time series of precipitation, temperature and potential evapotranspiration. Then, questions that this research tries to answer are as follows:

1. While using simulated (synthetic) precipitation for d sub-catchments, are there cases where the sums of precipitation (that may or may not be aggregated in time) exceed those of what was observed till now?
2. Using simulated model inputs, are there cases where model discharges become larger than those that were modeled to-date or observed?
3. Using simulated discharge, are there cases where the sums of simulated values become larger than those of what was observed?
4. Lastly and most importantly, what properties of time series, that when conserved by the simulations, correspond to aggregates that become equal or larger than those of the observed cases?

These questions will be answered by exploring modified approaches, based on the existing ones, allowing users to generate multi-site/multivariate synthetic time series that perform *better* than the older ones. The meaning of what is considered better will become clear while progressing through the coming chapters. For this purpose, the chapters are structured in the following manner: First, Chapter 2 describes the state-of-the-art of existing time series generators which is then followed by Chapter 3 on the theory of Copulas and their properties that are important to describe time series in general. The study area, rainfall-runoff modeling and the experimental setup used to evaluate quality of time series is shown in Chapter 4. Next, existing and new time series generators are presented in Chapter 5. New bivariate quality metrics and objective functions are discussed in Chapter 6 which is then followed by Chapter 7 where spatial aspects of extremes, criteria for developing dependence description in d dimensions and new type of performance metrics that represent said dependence in d dimensions are introduced with potential applications. Chapter 8 shows

how various properties of time series may evolve in space-time for some observed cases. Results of using various time series and their evaluation for the extreme value cases are presented in Chapter 9. Finally, conclusions are drawn along with recommendations of topics that may be researched in the future in Chapter 10.

2 Stochastic time series generators

Stochastic time series generator are the main tools used in this thesis to simulate time series that have a desired set of properties. It makes sense to define the term *time series generator* formally, to avoid confusions. Here, it means a number generator be it uni- or multi-variate that generates a sequence of values that have some prescribed set of properties with some *stochasticity* included. As an example, consider the simpler case of temperature at a certain point on the surface of the Earth. It has a clear daily cycle due to the Earth's rotation around its own axis on to which the annual cycle, due to the Earth's revolution around the Sun, is superimposed to begin with. The rest of the effects that cause the variability are those due to the local climate, proximity to an ocean or a mountain range, masses of air that are heated by the Sun over the surface of the Earth. With the knowledge of these properties, one can formulate a time series generator that statistically, more or less, mimics these and all that without the knowledge of the physics of the entire process.

A time series generator can be formulated for almost any variable that can be observed. As a rule all of them try to reproduce the auto-correlation function of a given reference, while possibly preserving other properties. Purpose built generators exist in many areas be it finance, energy demand, traffic, supply and demand of goods, crop growth modeling and most importantly weather.

There are numerous use-cases for time series generators. The most important one being, for testing a system's response where no or limited observed data are available. Suppose the case of modeling a bridge or a road that is being designed and needs to be tested beforehand to see if it withstands the number of vehicles that may pass through/over it. Data from existing roads can be taken as a reference, but suppose that the situation is unique. For this purpose, a time series generator of vehicles is needed. From experience, it is known that it should create a temporal sequence of moving vehicles along with their weight and duration spent on the road/bridge. The main aim there would be to see if congestion would take place and how much it may be, for the worst case. Another purpose could be wear and tear of the structure, caused mainly by heavy vehicles and the weather. Simultaneously, time of the year along with temperature may be desired as freezing and too hot temperatures lead to an accelerated deterioration of the structure. Based on the system's response to *synthetic/stochastic* inputs, it can be evaluated if further changes to the design are required.

The idea of time series generators is very interesting while looking from the stand point of physics. A good analogy can be that of a scrambled egg. Suppose, 10 eggs are cracked open in separate bowls. If done carefully, the yolk stays separate from the white. The egg in each cup is uniquely different from its neighbors but for the observer they are unscrambled eggs. The property that separates an unscrambled (low entropy) one from a scrambled (high entropy) is that the yolk stays as a single mass inside the white regardless of their relative

size to each other. Now, suppose the eggs are scrambled i.e., each yolk is not a single mass any more. Hence, the yolk being a single mass can be seen as the property of the system which essentially defines it. What if the opposite process is required, i.e., unscrambling an egg.

The essence of time series generators is akin to unscrambling an egg. Any given time series uses pseudo-random numbers as inputs to create a time series whose properties are similar to those of a reference, albeit, using some rules/constraints. Full unscrambling of the egg for a time series generator would be the case in which the output of the time series generator is a copy of the reference. Such a generator would be pointless, as the reference exists already. What is mainly required is an output time series that has no correlation with the reference and only mimics its properties. In simpler terms, an unscrambled egg such that it does not look exactly like the original one is required, it should be different and more importantly, if an external observer is presented with an unscrambled egg (after scrambling) mixed with the other eggs (that were not scrambled to begin with), the observer should be unable to make a difference while deeming them all to be unscrambled eggs.

Time series of observed variables have properties such as correlations, asymmetric dependence and entropies among many others for both the auto- and cross-cases. Consider the distributions of precipitation and river discharge. A fundamental difference can be observed readily i.e., precipitation is intermittent (a mixed distribution) while discharge is, mostly, continuous. A time series generator built to simulate these variables has to explicitly deal with such behaviors. Such problems have been dealt with for decades. A excellent reference for beginners is that by Box and Jenkins (1976). The main question that has to be asked is: *which properties of the reference should the generator preserve?*

Numerous ways have been proposed and are in use to generate time series in various fields of science. The basic ones are presented first and afterwards the more sophisticated ones.

To begin with, rainfall-runoff models are also time series generators where equations describing nature can be seen as rules on how the output is formed. The simplest method to generate the maximum discharge was the Rational Method dating back more than 150 years (Beven, 2020) where, for each catchment, the amount of *design* discharge was equal to the an empirical constant multiplied by a given rainfall intensity. The science has progressed much farther since then. Continuous research has brought physically based models of catchments that take in a large set of inputs to compute states of various variables on a regular grid for any time step. Much simpler models, conceptual rainfall-runoff models, were also developed whose complexity lies between the empirical and the physically based models. These require inputs such as precipitation, temperature and potential evaportranspiration among others as inputs to produce a time series of discharge and require calibration of model parameters. Various models are capable of either a single event simulation or a continuous time series. These may be used in lumped or a distributed configuration. One such model namely the HBV (Bergström, 1992) is used in the thesis and will be described later in detail.

The simplest time series generators that keep the auto-correlation function are the Auto-Regressive (AR) and the Moving Average (MA) generators (Box and Jenkins, 1976). They work with the assumption of linear dependence between time steps. These use the auto-

correlation function that may or may not be extracted from observed data along with some Gaussian noise added to produce a time series. The product of the Gaussian noise and the correlation function, depending on the number of previous time steps used, is the value at the next time step. Infinitely long time series can be generated through this process. The main problem with these is that the correlation function which is assumed to be stationary throughout the time period of generation and is the only property that fully defines the system. This is generally not the case. Much cannot be done about this as it is generally not known how the correlation function evolves in time. Another problem is that of the assumption that the noise in space-time is Gaussian which is also, generally speaking, not the case. Several refinements of these processes exists namely ARMA, ARIMA and SARIMA (Box and Jenkins, 1976). A generator worth mentioning is the FFTMA (Ravalec et al., 2000; de Figueiredo et al., 2020), which is the spectral version of the MA process. Such processes do lack the imposition of certain behaviors such as the properties of precipitation. Its intermittent behavior needs to be explicitly accounted for by some form of a truncated Gaussian process. Similarly, discharge is also an issue as it has faster rising limbs and slower recession limbs. Based on the Gaussian nature of outputs of these generators, they are not so useful except for cases with long aggregation times such as the monthly and annual scales. Such models can also be modified for use as forecasting tools by fixing the observed states from previous time steps and then calculating the next few time steps by taking a distribution of noise values.

Another class of time series generators is that of the *resamplers*. Such methods involve creating time series that may or may not have a Gaussian distribution but at the end perform a transformation such that the distribution of the outputs is same as that of the reference data while still maintaining the properties of the reference, the so called *surrogates*. Theoretically, speaking any time series generator can be modified to behave in this way. The methods Phase Randomization, Amplitude adjusted Fourier transform, Iterative amplitude adjusted Fourier transform and multivariate iterative amplitude adjusted Fourier transform belong to this class and will be discussed in much more detail later. These methods were originally created to test a given time series for non-linear dependence but they also lend themselves to simulation of time series with linear or non-linear properties as a side-effect. An excellent reference that summarizes these type of generators is that by Lancaster et al. (2018).

Yet another method that is popular is that of the *wavelets* (Daubechies, 1992). These are one of the spectral methods to generate time series and involve distilling a given time series into components that have their periods oscillate with distinct lengths in time. Such methods give an insight in to various periodic components that a time series may or may not have. Wavelet iterative amplitude adjusted Fourier transform (Keylock, 2012) is one of these methods among many. One crucial feature that sets the wavelet-type time series generators apart from other is that they seem to be able to produce non-linear i.e., asymmetric dependence. An interesting approach of applying such methods to extreme discharge is that by Brunner and Gilleland (2020).

Numerous physically-based models that incorporate the real behavior/interaction of variables in the atmosphere, land and the oceans exist (Flato et al., 2013). These models lie in the realm of Physics and Meteorology and are not discussed here. Their main drawback being

that they are data intensive and require super-computers to run and even then may take months to finish. Due to this hurdle, stochastic weather generators have been developed.

As was stated in the introduction, the aim here is to evaluate the worst case scenario(s) in terms of bad weather i.e., stream flows that are large enough to cause damage to their surroundings that may cause a disruption in desired human activity, loss of property and or even worse, loss of life. So far, generic time series generators were discussed. These may or may not be able to mimic crucial properties of weather-related variables. To solve these problems, a class of stochastic generators exist i.e., *weather generators*. The main appeal of using such generators is that compared to using numerical global/regional climate models, stochastic weather generators are orders of magnitude less demanding in their computational requirements as they do not have to the solve all the partial differential equations that exist in a four dimensional grid and less input data requirements. Albeit, they only reproduce a subset of properties. The first formal weather generator can be attributed to Richardson (1981) where daily precipitation, minimum temperature, maximum temperature and solar radiation were simulated. Their methodology involved finding properties such as the probability of no-rain and rain and the distribution of rainfall. Next, properties of the remaining variables were calculated for the two states of rainfall. Then, rainfall was generated using a Markov-chain process and depending if rainfall took place or not, the other variables were drawn from an appropriate multivariate distribution. Provisions were also made to keep the seasonality of the variables. Another original approach was made by Racsko et al. (1991), where the weather was taken as a three dimensional process with the temperature being the most important one and the other two variables being the number of solar hours and precipitation. Their additional aim was to capture the length of dry precipitation periods more accurately for crop modeling. They separated the precipitation events into small, medium and large quantities with properties calculated subsequently for each group. Similarly, they fitted Gaussian distributions to temperature conditioned for every group of rainfall. Solar hours were also modeled similar to temperature but with a mixed distribution for the cases where the hours fell to zero. Racsko et al. (1991)'s approach can be seen as a refinement of the Richardson's approach. Modern weather generators are based on these two ideologies and are referred to as the *Richardson-type* and the *Serial-type* respectively. Since then, much more refined models have been in use that deal with very many variables at the same time. Some famous weather generators are WGEN (Richardson, 1984), CLIMGEN (Stöckle et al., 1999), CLIGEN (Nicks and Gander, 1994), WeaGETS (Chen et al., 2012), LARGSWG (Semenov and Barrow, 2002), CLIMA (Donatelli et al., 2009) and AWGEN-2D (Peleg et al., 2017). A comprehensive review of existing stochastic weather generators is given by Ailliot et al. (2015). Generators that can handle explicit change in climate (increase in temperature for example) such as those by Kilsby et al. (2007) and Schlabing et al. (2014) also exist.

Both physically-based and statistical variants of generators are in use. The former requires much more input and computational power while the later is much more efficient for generation albeit missing the true physical interaction of the state variables with each other.

Another interesting approach that is famous these days is that of using copulas. These allow for dependence modeling in a much more robust manner because they account for depen-

dence among variables based on their order with respect to each other and not the variable dependent marginal distribution. Proper weather generators with one exception, Schlabing et al. (2014), that are on the same level as the regular time series generators do not exist as of yet. Any existing generator can be modified to operate in the rank-space. However, univariate multi-site generators do exist. Their main application concerns precipitation and discharge in hydrology. An overview on the use of copulas in hydrology is given by Hao and Singh (2016). Nguyen and Chen (2022) formulated a generator that models continuous precipitation at a temporal resolution of 10 min. Vandenberghe et al. (2010) show a similar example. Gao et al. (2020) created a precipitation model that can be used to simulate both flood and drought behavior with minor drawbacks. Another generator worth mentioning is NiedSim (Bárdossy and Plate, 1992) and its subsequent variants Bárdossy et al. (2020); Bliefernicht and Bárdossy (2007); Beck (2013); Müller et al. (2017). It utilizes Fuzzy Logic combined with weather circulation patterns to produce precipitation time series that have properties that are difficult to capture otherwise. It is currently employed by various federal states in Germany for precipitation simulation. Once more, Ailliot et al. (2015) give an overview of some generators that employ copulas.

3 Copulas

A copula is a mathematical formulation that describes the dependence between/among at least two variables regardless of their magnitudes by only considering their order of occurrence. Before copulas are formally introduced, a traditional case of how a dependent and independent variables are modeled is presented. Suppose two variables X and Y with marginals (or observed values) x and y respectively and that Y depends linearly on X, then a straight line of the form $y = mx + c$ can be used, where x is the independent (or predictor) variable, y the dependent (or predictand) variable, m being the slope and c being the y-intercept. Such a line can be fitted to any two variables in which one is assumed to be linearly dependent on the other through a method called Linear Regression where the sum of squared difference of observed and predicted values of Y using X is minimized. Fig. 3.1 shows an example. Furthermore, a general (multivariate) form of it, Multiple Linear Regression (MLR) takes more than one independent variable to predict only one. In such a case, for each predictor variable there exists a separate slope or coefficient. Such a technique focuses on relation among variables in the marginals' domain i.e., using x and y.

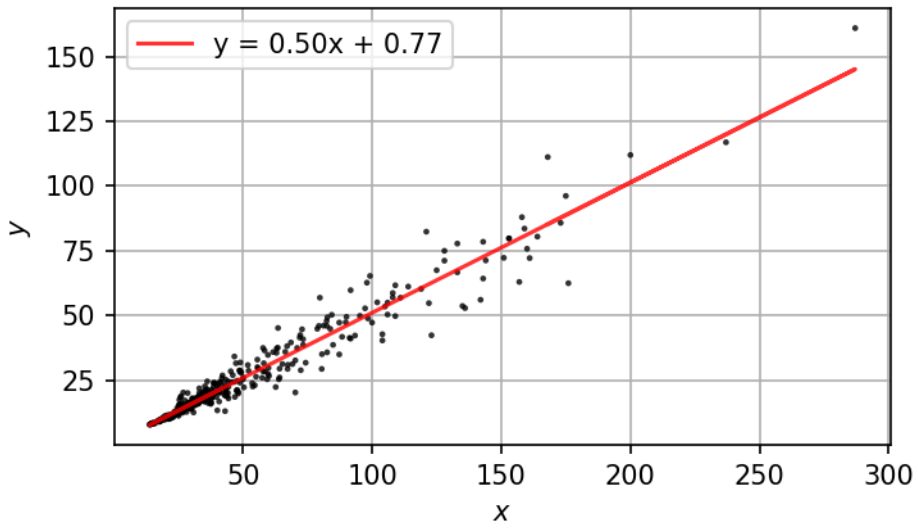


Figure 3.1: An example of simple linear regression with X as the predictor and Y the predictand.

The problem of prediction/interpolation/extrapolation can also be dealt with, by moving to a different type of space. Using the previous example, the corresponding cummulative probability distribution functions (CDFs) of the predictor and predictand variables, $F(x) = P[X \leq x]$ and $F(y) = P[Y \leq y]$ exist which are also referred to as non-exceedence

probabilities or *grades* of X and Y respectively, where P symbolizes probability whose range is between 0 and 1. These may be theoretical or empirical. The use of CDFs is central to statistics as it defines the values a variable may attain along with their relative frequencies. They are monotonic functions that express the probability of variables not exceeding a given threshold. CDFs are central to copulas as the non-exceedence probabilities are the entities that are used to construct them. To do so, $F(x)$ and $F(y)$ are used rather than X and Y. The main difference being that no lines are used; only special type of monotonically increasing surfaces inside a unit hyper-cube. And this is what a copula refers to. An entity that describes the dependence of these distribution functions. It could be stated that a copula describes the relation among variables in probability space, while fulfilling some conditions. Specifically, Sklar's Theorem (Sklar, 1959) states that any multivariate distribution can be re-expressed as a copula. Consider the joint probability of X and Y, $F(x, y)$ that may be written as,

$$F(x, y) = P[X \leq x, Y \leq y] \tag{3.1}$$

Then, Sklar's Theorem states that if $F(x)$ and $F(y)$ are continuous there exists a copula C that links the two CDFs and is unique. The joint probability $F(F(x), F(y))$ may be rewritten as,

$$F(x, y) = C(F(x), F(y)) \tag{3.2}$$

Sklar's Theorem also holds for the converse case i.e., if a copula is defined then $F(x, y)$ also exists. Instead of using X and Y, a d-dimensional vector X_d is used. It could represent d time series, for example. The probability form of such a vector can be written as,

$$(U_1, \ldots, U_d) = \Big(F_1(X_1), \ldots, F_d(X_d)\Big)$$

where $U_1, \ldots, U_d$ are the copula analogs of CDFs. Supposing that $P[X \leq x] = u$, a copula (C) is a joint cumulative distribution function that can be written as,

$$C(u_1, \ldots, u_d) = \text{Probability of } [U_1 \leq u_1, \ldots, U_d \leq u_d] \tag{3.3}$$

whose density function (c) is,

$$c(u_1, \ldots, u_d) = \frac{\delta^d}{\delta u_1 \ldots \delta u_d} C(u_1, \ldots, u_d) \tag{3.4}$$

where, C refers to the copula that describes the dependence among all involved variables in probability space. d refers to the number of variables considered. Marginals of the copula (u) are defined on the unit hyper-cube that has d dimensions and are uniformly and independent and identically distributed (i.i.d) between 0 and 1.

Formally, $C : [0,1]^d \rightarrow [0,1]$ is a d-dimensional copula (defined on the unit hypercube) if C is a joint cumulative distribution function of a d-dimensional random vector on the unit cube $[0,1]^d$ with uniform marginals (Nelsen, 2007). Furthermore, a d-dimensional C must satisfy the following properties:

1. $C(u_1, \ldots, u_{i-1}, 0, u_{i+1}, \ldots, u_d) = 0$ i.e., the copula yields a value of zero if any its arguments are zero.
2. $C(1, \ldots, 1, u, 1, \ldots, 1) = u$ i.e., the copula yields a value of u if only one of its arguments is u and the rest are 1.
3. C is non-decreasing as its arguments increase from 0 to 1, similar to a CDF. More specifically, for every orthotope/hyperrectangle drawn inside a copula whose volume is given by $V = \prod_{i=1}^{d} [x_i, y_i]$ is always non-negative. Where, x and y are the copula marginals with the former always less than or equal to the latter.

In other words, all the arguments on the left side of Eq. 3.3 can take values between 0 and 1 (both inclusive) and yield values between 0 and 1. Obviously, a copula needs at least two variables. The minimum that Eq. 3.3 can yield is a 0 while the maximum being a 1. Yet, another, simpler, way of looking at Eq. 3.3 is that it takes non-exceedence probabilities of multiple predictor variables to produce a conditional distribution that can be further transformed to marginals by the inverse of the respective predictand CDF i.e., for given $u_1, \ldots, u_{d-1}$ as the predictors to give a conditional distribution U_d.

An empirical copula of any two distinct variables X and Y having N values each and CDFs $F(x)$ and $F(y)$ can be constructed by an x-y scatter of $F(x)$ and $F(y)$. Fig. 3.2 shows the empirical copula of the same data used in Fig. 3.1. The advantage of using such an approach is that the relationship between/among variables is independent of the magnitude/range of their distributions and depends purely on their order/position with respect to each other, meaning that any monotonic transformation applied to the data does not affect the so called *dependence structure*. Such a form is useful for cases where a few outliers might dominate the entire covariance used in MLR or similar techniques. This property is also a double-edged sword i.e., a copula perceives values, whose difference could be trivial to the user, as distinct that have as much influence as the larger, more important, ones. Another main advantage of a copula is that instead of yielding a single value, a whole distribution of possibilities, or a not necessarily Gaussian conditional distribution, is yielded, which can then be used to give confidence bounds for the predictand.

In this thesis, only empirical copulas are utilized. The reasons will become clear while progressing through the chapters. Some theoretical copulas are mentioned nevertheless. Theoretical copulas exist that can be fitted to data e.g., Gaussian and Archimedean. The more interested reader is referred to Nelsen (2007). For any proposed one, be it parametric or non-parametric, certain conditions apply. The most important one being that for any interval, the densities cannot drop below zero i.e., a copula starts from zero, $C(0,0) = 0$, monotonically increases and stops at one, $C(1,1) = 1$. Furthermore, there exist the Fréchet-Hoeffding copula bounds, these limit the minimum and maximum $C(u_1, \ldots, u_d)$ values at any point in a copula. e.g., for a two-dimensional case, these are:

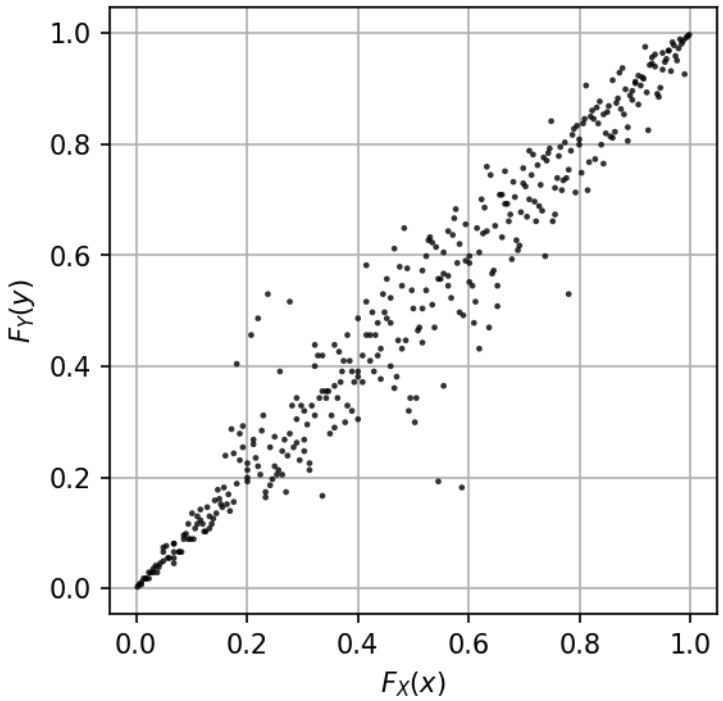


Figure 3.2: Empirical copula corresponding to data in Fig. 3.1.

$$max\{u_1 + u_2 - 1, 0\} \leq C(u_1, u_2) \leq min\{u_1, u_2\}$$

The lower bound refers to dependence between two variables that are perfectly negatively correlated (countermonotone) while the upper for perfectly positively correlated (comonotone). Important to mention are the independence and dependence copula. As their names give it away, the former refers to the case where u_1 is independent of u_2. It is given by,

$$C(u_1, u_2) = u_1 u_2$$

while the later refers to the case when both u_1 and u_2 have a 1:1 relation (the upper Fréchet-Hoeffding bounds). It can be written as,

$$C(u_1, u_2) = min(u_1, u_2)$$

It should be noted that the upper Fréchet-Hoeffding bounds exist for all copulas in any number of dimensions but the lower bounds are defined for the two-dimensional case only.

3.1 Auto- and cross-copula

A copula can be an *auto*-copula or a *cross*-copula. An auto-copula holds the dependence of a variable with itself. For example, for a time series of a variable X, an empirical copula can be constructed by taking u_t and u_{t+l}, where t is the time step and l is the number of steps the series is lead or lagged by, e.g., $\pm$ 2, steps. The resulting copula will have $N - |l|$ points for N number of time steps. A practical example would be to take a discharge time series, find its distribution function, lead a copy of it by one step and plot the resulting series. One can even take multiple leads or lags and construct a high-dimensional copula. This however, violates the assumption that the marginals of copula are independently distributed. Such a condition cannot be met for auto-copulas. A cross-copula holds the dependence between, at least, two distinct variables. This could be, for example, discharge time series at separate locations along a river. The rest of the idea is similar to that of the auto-copula but here the copula can be for the same time step as well i.e., $l = 0$. Fig. 3.3 shows some examples of empirical auto- and cross-copulas. Similar terminologies exist for the Pearson correlation i.e., auto-correlation and cross-correlation.

3.2 Pearson and Spearman correlation coefficients

For the sake of completeness, the Pearson correlation (ρ_p) between any two variables X and Y can be written as,

$$\rho_p = \frac{covariance(X, Y)}{\sigma_X \sigma_Y} \tag{3.5}$$

Similarly, the Spearman correlation (ρ_s) for probability distribution functions $F(x)$ and $F(y)$ is given by,

$$\rho_s = \frac{covariance(F(x), F(y))}{\sigma_{F(x)} \sigma_{F(y)}} \tag{3.6}$$

where, σ is the standard deviation. ρ_p and ρ_s can have values between -1 and +1, with -1 being a perfect negative linear relation, 0 being no linear correlation and +1 being a perfect positive linear relation between the two variables. Both, ρ_p and ρ_s, are good measures as long as the dependence between the two is linear with ρ_s used for cases where a monotonic relation is preferred and it is the main measure to calculate the strength of dependence of a copula. ρ_s can be viewed as an analogue of ρ_p when moving from marginals to grades. An example of getting a wrong impression, consider an empirical copula of two variables that has a donut shape. Here, ρ_p would be zero but actually there is a relation only that it is not a linear one; more precisely, for a given value of X there are two, almost, equally positive and negative values of Y, which result in a cancellation while calculating the covariance between the two, leading to the belief that they do not relate. A practical example would be to calculate auto-ρ_p and auto-ρ_s of a daily temperature series in which Y leads/lags X by three months. Fig. 3.4, 3.5 and 3.6 show hypothetical time series having diamond-, donut-

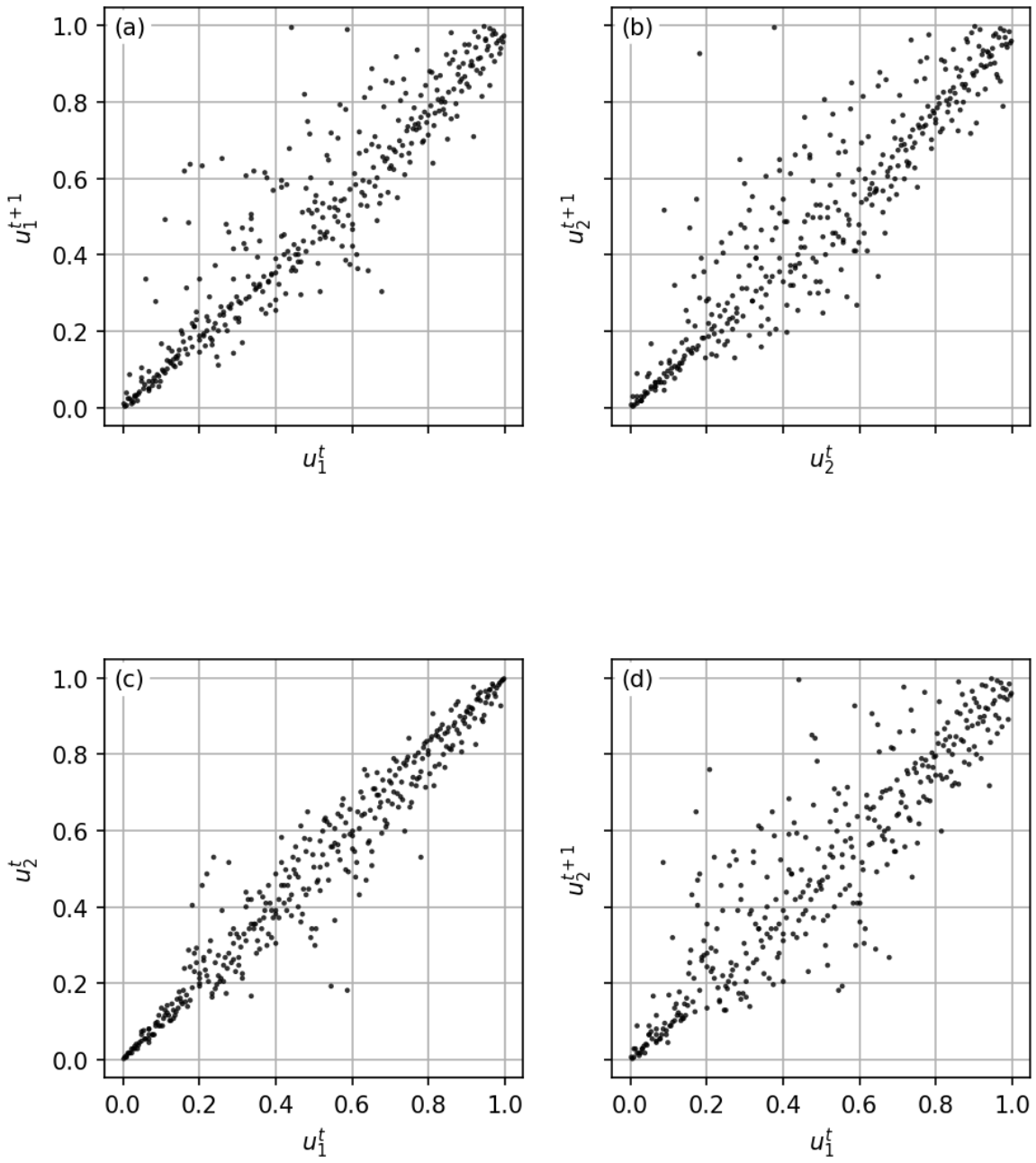


Figure 3.3: Examples of empirical copulas for various cases. (a) is an auto-copula for a variable U_1 with a lead of 1 step, (b) is an auto-copula for a variable U_2 with a lead of 1 step. (c) is cross-copula of U_1 and U_2 with no lag/lead steps. (d) is a cross-copula of U_2 having a lead of 1 step over U_1.

and V-shaped empirical copulas respectively with no significant ρ_p and ρ_s. Fig. 3.7 shows how the Spearman's auto-correlation function might look like for a time series. ρ_s is also called the *grade* correlation coefficient sometimes. ρ_s is not the only measure of the strength of dependence between variables. Other measures such as the Gini, Blomqvist coefficients and Kendall's Tau are also used. These are discussed in detail in Chapter 5 of Nelsen (2007).

3.3 Some additional terms

A "low"refers to a value, or its rank, in the lower tail of the distribution of a variable while a "high"refers to a value, or its rank, in the upper tail. The term "low-high"refers to the occurrence of a low followed by a high in a time series. A similar term with an opposite meaning is "high-low". Other terms are "low-low"and "high-high". These have contextual meaning and are used for clarification purposes only. Fig. 3.8 demonstrates these terms using two neighboring discharge stations' copula.

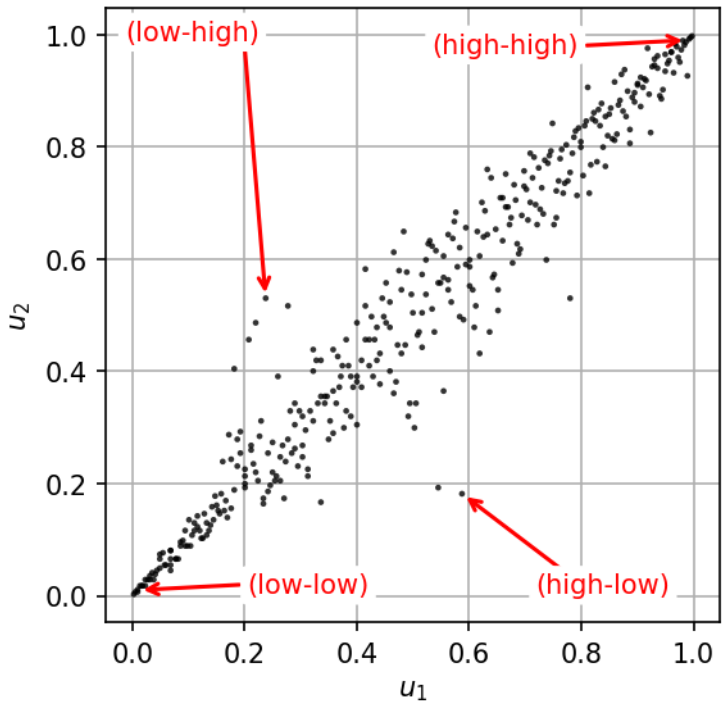


Figure 3.8: Demonstration of terms used to describe local copula behavior.

3.4 Order asymmetry

For a copula where there is no difference between relative densities in the regions $c(u_1, u_2)$, $c(1-u_1, u_2)$, $c(u_1, 1-u_2)$ and $c(1-u_1, 1-u_2)$ is said to be symmetric. For the opposite case where there is a difference between the four terms, the copula is said to be asymmetric. Theoretical copulas exist that do exhibit asymmetry and the same can be observed for

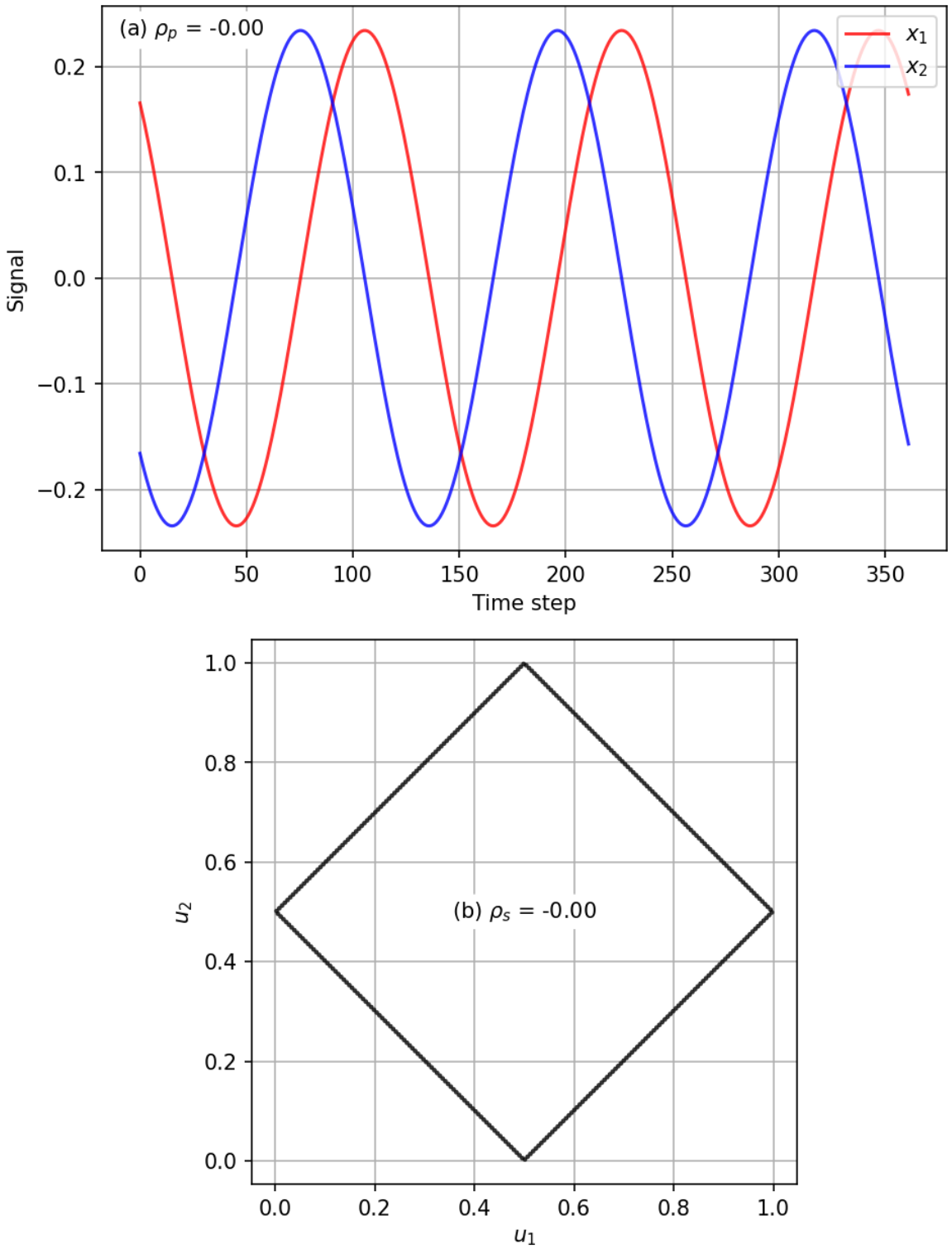


Figure 3.4: Example of two time series (a) that are similar to each other but having a lag/lead of 91 steps resulting in a diamond-like dependence (b) with no significant Pearson or Spearman correlation.

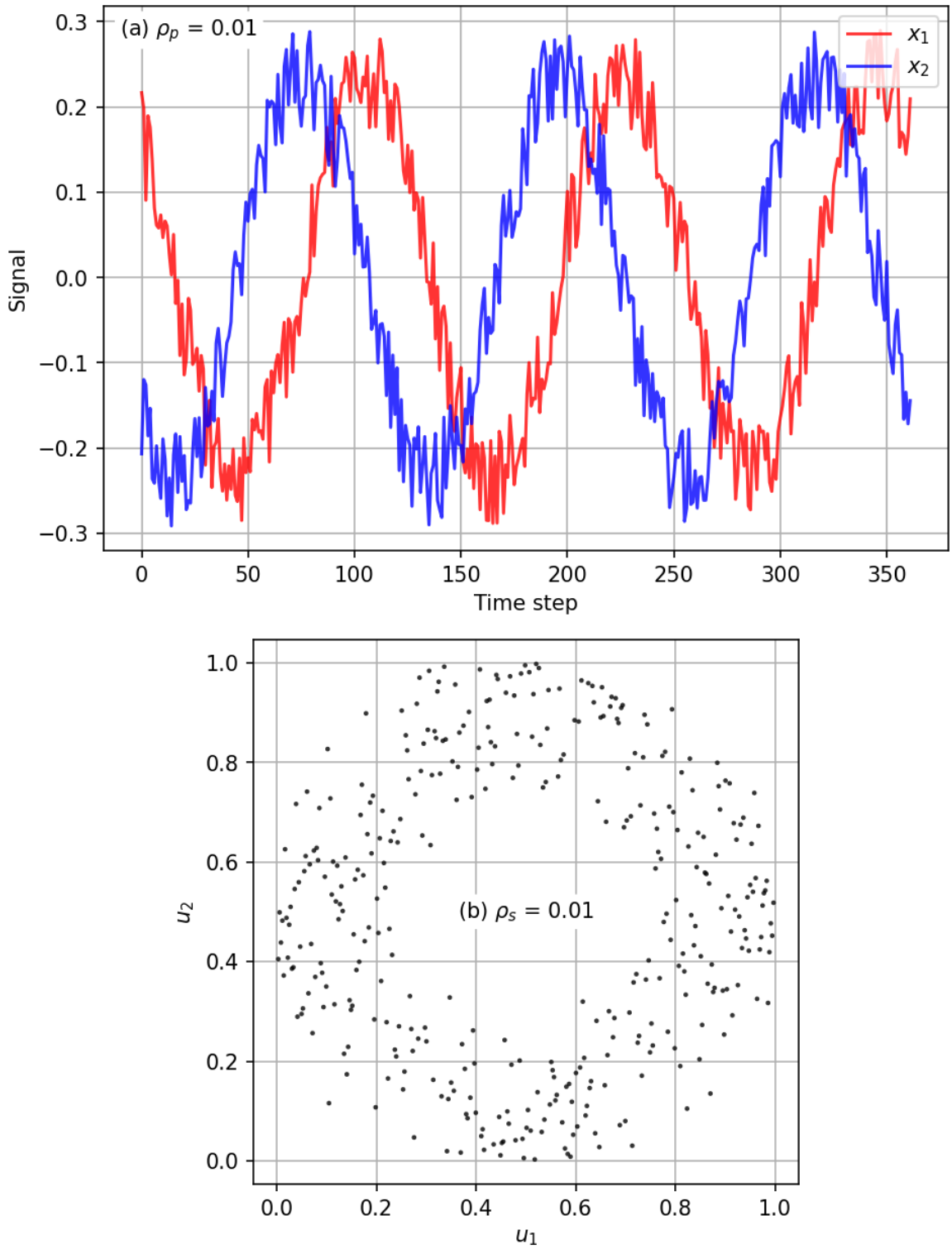


Figure 3.5: Example of two time series (a) that are similar to each other but having a lag/lead of 91 steps resulting in a donut-like dependence (b) with no significant Pearson or Spearman correlation.

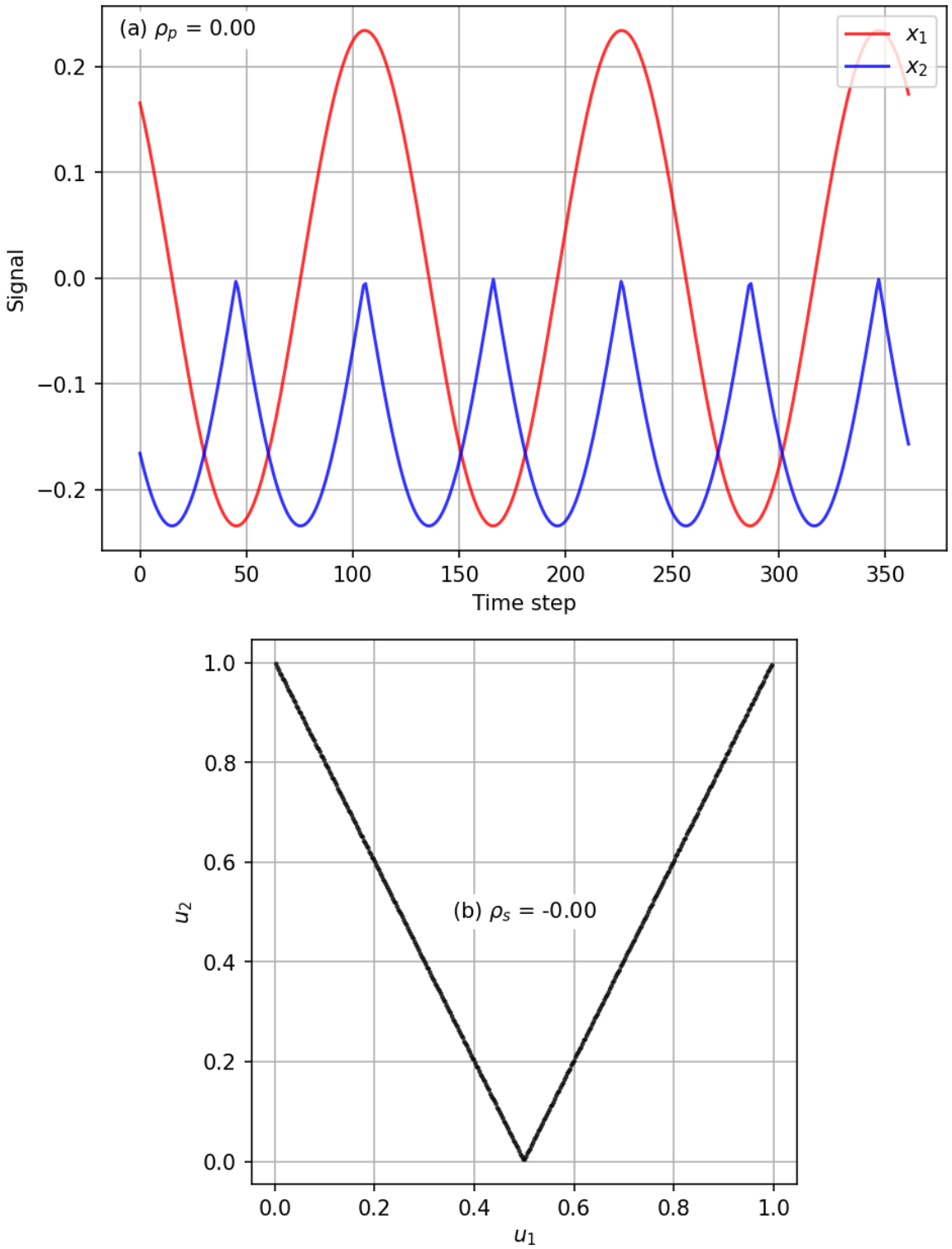


Figure 3.6: Example of two time series (a) that are similar to each other but having a lag/lead of 91 steps resulting in a V-like dependence (b) with no significant Pearson or Spearman correlation.

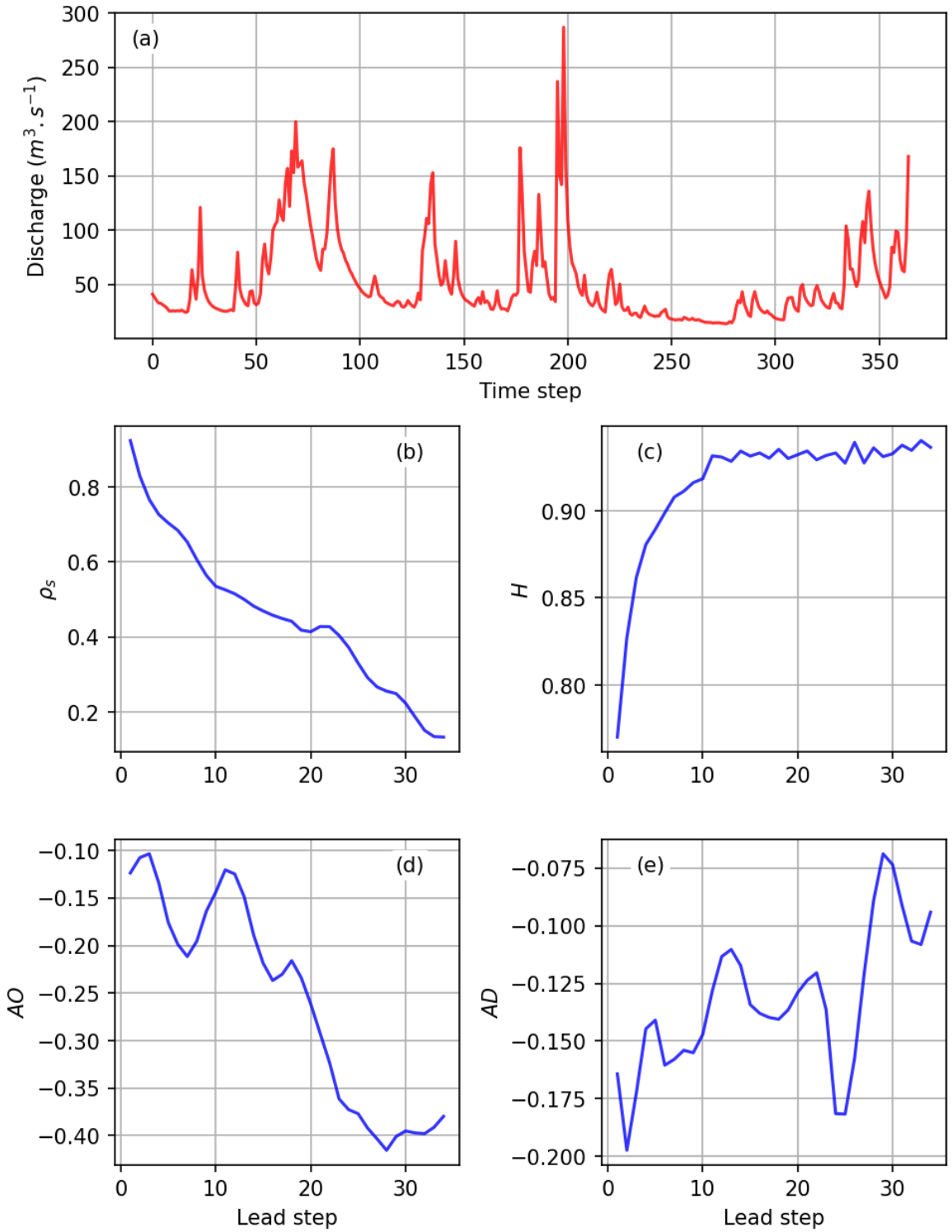


Figure 3.7: Properties of the empirical auto-copulas (b, c, d, e) of an observed daily discharge time series (a) for which U_2 leads U_1 by various steps.

variables in nature. For example, consider the example of discharge. The periods of low flows are generally much longer compared to peak flows that only last for a couple of days compared to the case of dry periods that may last for months. Measures to describe this asymmetry are important and provides information about the nature of dependence of the variables under consideration. For a bivariate copula, two types of asymmetries exist i.e., order and directional asymmetry denoted by AO and AD respectively.

AO refers to the relative occurrences of high-highs against low-lows of a copula for the bivariate case (Guthke, 2013). It is the expected value of the term $(U_1 + U_2 - 1.0)^3$ and is given as follows,

$$A_2 = E[(U_1 + U_2 - 1.0)^3] \tag{3.7}$$

or

$$A_2 = \int_0^1 \int_0^1 (u_1 + u_2 - 1.0)^3 c(u_1, u_2) du_2 du_1 \tag{3.8}$$

where, E refers to the expected value and A_2 is the theoretical order asymmetry. For the empirical case, raw order asymmetry (AO_{raw}) maybe written as,

$$AO_{raw} = \frac{1}{N} \sum_{t=0}^{N-1} \left(u_1^t + u_2^t - 1.0\right)^3 \tag{3.9}$$

where, t is the time step. In essence, it is the mean of the cube of the distances of points in a copula from the line $u_2 = 1 - u_1$. It is negative when the incidence of low-lows are relatively more than the incidence of high-highs and positive for the opposite case. It can be *auto* or *cross* depending on the copula. As a general example, order asymmetry of the Gaussian copula, for a sample large enough, is, theoretically, zero meaning that the dependence of low-lows and high-highs is equal. A more relatable example is that of discharge in a stream. It has, at least in the author's experience, a negative order auto-asymmetry if the auto-copula is that of u_2 leading u_1. In this case, for a very dry spell the flows approach a state of equilibrium i.e., low-lows for a long time. The opposite case is not true. It would be strange that a stream experiences consistently very high flows, i.e., no recession, for periods as long as the dry periods. It is always a faster rising limb with a slower recession for an event, depending on the size of the catchment.

Magnitudes of raw order asymmetries are meaningless when comparing them for two copulas with different ρ_ss. This is akin to the term in the denominator when calculating skewness of a sample. Hence, it is important to normalize them. This gives it the same range as ρ_p and ρ_s i.e., $\pm$ 1. Guthke (2013) calculated an equation to yield the maximum possible order asymmetry (AO_{max}) for a bivariate copula with a given ρ_s,

$$AO_{max} = 0.5 \times (1 - \rho_s) \times \left(1 - \left(0.5 \times (1 - \rho_s)\right)^{\frac{1}{3}}\right) \tag{3.10}$$

The minimum absolute order asymmetry is, of course, zero. Fig. 3.9 gives a visual representation of how location of points in the copula affect the order asymmetry. Finally, the order asymmetry used in this report is always a normalized one and is given by,

$$AO = \frac{AO_{raw}}{AO_{max}} \tag{3.11}$$

From here on, any reference made to order asymmetry would mean that it was calculated using Eq. 3.11 in two dimensions unless otherwise specified. Fig. 3.7 shows how it might look like for a time series. There are observed cases where AO goes beyond $\pm$ 1. These arise for when a significant number of values repeat in a time series such as those of the hourly precipitation.

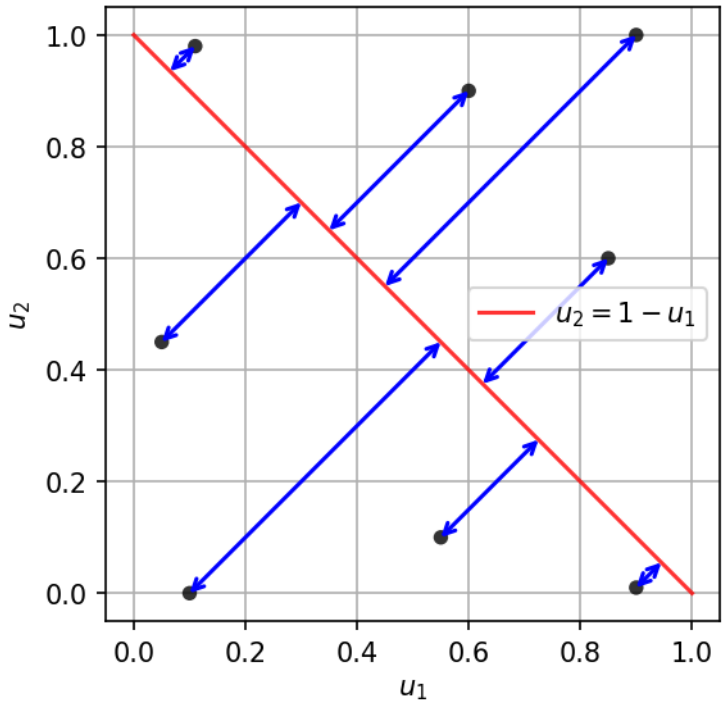


Figure 3.9: Demonstration of how distance of a point affects the order asymmetry. The greater the distance (blue line) from the red line the higher the effect.

3.5 Directional asymmetry

It refers to the relative occurrences of high-lows against low-highs (Bárdossy and Hörning, 2017). It is the expected value of the term $(U_1 - U_2)^3$ that may be written as follows,

$$A_3 = E[(U_1 - U_2)^3] \tag{3.12}$$

or

$$A_d = \int_0^1 \int_0^1 (u_1 - u_2)^3 c(u_1, u_2) du_2 du_1 \tag{3.13}$$

where, A_d is the theoretical directional asymmetry. Raw directional asymmetry (AD_{raw}) for an empirical copula may be written as,

$$AD_{raw} = \frac{1}{N} \sum_{t=0}^{N-1} (u_1^t - u_2^t)^3 \tag{3.14}$$

In essence, it is the mean of the cube of the distances of points from the line $u_1 = u_2$ in a copula. It is negative when low-highs happen more often than high-lows and positive for the opposite case. It can be *auto* or *cross* depending on the copula. Again, the Gaussian copula has a directional asymmetry of zero by definition, i.e., the chances that low-highs happen are similar to those of high-lows. This measure is more related and pronounced for the case of a daily discharge auto-copula, with u_2 leading u_1 by a couple of days, where a rising limb is more abrupt as compared to the recession limb. This gives discharge a negative directional auto-asymmetry. The opposite case is unusual i.e., the recession limb being shorter than the rising and happens rarely but is entirely possible for cases such as discharge rising slowly due to small precipitation for a long time which stops abruptly. A drawback of directional asymmetry is that it is very susceptible to *outliers* in the copula with a strong dependence (correlation- or entropy-wise). A single value having a large enough departure from the line $u_1 = u_2$ can change its sign and magnitude in a significant manner. Following the method used in Guthke (2013), to derive Eq. 3.10, the maximum possible directional asymmetry (AD_{max}) is,

$$AD_{max} = 0.5 \times (1 + \rho_s) \times \left(1 - \Big(0.5 \times (1 + \rho_s)\Big)^{\frac{1}{3}}\right) \tag{3.15}$$

The minimum possible absolute directional asymmetry is zero. Fig. 3.10 gives a visual representation of how location of points in a copula affect the directional asymmetry. Finally, the directional asymmetry used in this thesis is always a normalized one for the two-dimensional case and is given by,

$$AD = \frac{AD_{raw}}{AD_{max}} \tag{3.16}$$

From here on, any reference made to directional asymmetry would mean that it was calculated using Eq. 3.16 in two dimensions unless otherwise specified. Fig. 3.7 shows how it might look like for a time series. It is worth mentioning that order and directional asymmetries are also called reflection and permutation asymmetry in Rosco and Joe (2013) and Krupskii (2017). Although the equations presented therein are more general and do not include the normalization. A very similar formula that is used for the marginals is given by Keylock (2012).

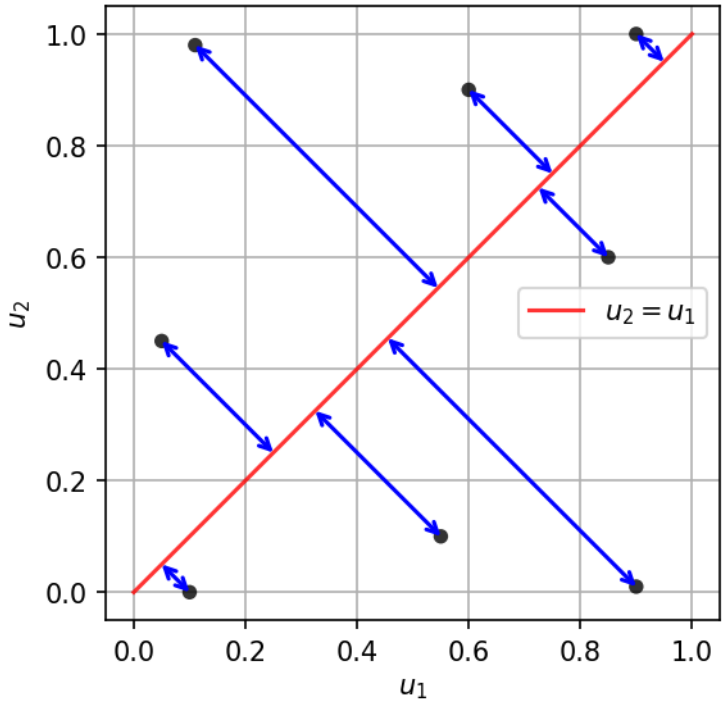


Figure 3.10: Demonstration of how distance of a point affects the directional asymmetry. The greater the distance (blue line) from the red line the higher the effect.

3.6 Copula entropy

Perhaps one of the most elusive measures, and difficult to get achieve through modeling, is the copula entropy. It, mainly, provides information about the *compactness* of copula densities. It expresses the strength of both linear and non-linear dependence. The information about how and where they are in a copula is missing. The same is true for the direction of dependence.

Other forms of entropy measures exist e.g., Shannon (Shannon, 1948), Tsallis (Tsallis, 1988), Renyi (Rényi et al., 1961) and Kullback-Leibler (Kullback and Leibler, 1951) entropies. Throughout this thesis the most widely known Shannon's entropy (H) is used for calculating the overall strength of dependence and Mutual Information (MI) (Cover and Thomas, 2006) while looking at finer temporal scales i.e., local entropy. For a bivariate copula (C_2) it may be written as follows,

$$H(C_2) = -\int_0^1 \int_0^1 c_2(u_1, u_2) ln\big(c_2(u_1, u_2)\big) du_2 du_2 \tag{3.17}$$

A form more suitable for practical purposes, i.e., the empirical case, is,

$$H_{raw} = -\sum_{j=0}^{M-1}\sum_{k=0}^{M-1} c[j,k]\ln(c[j,k]) \tag{3.18}$$

$$c[j,k] = \frac{1}{N}\sum_{t=0}^{N-1}\begin{cases} 1, & \text{if } \left(\frac{j}{N} < u_1^t < \frac{j+1}{N}\right) \wedge \left(\frac{k}{N} < u_2^t < \frac{k+1}{N}\right) \\ 0, & \text{else} \end{cases} \tag{3.19}$$

where, M is the number of bins of a bivariate relative frequency histogram of the empirical copula and $\wedge$ is the logical AND operator. $c[j,k]$ refers to the relative density in each cell similar to the density in Eq. 3.4. j and k are the j-th row and k-th column of the histogram and is calculated using Eq. 3.19. A lower entropy would mean that there are regions of higher density in the copula and the rest is mostly empty i.e., high dependence. A higher value would mean that values are more spread out i.e., the dependence is weak. Measuring the strength of dependence is the elusive part. Monotonic dependence can be easily captured by ρ_s but a non-monotonic dependence, e.g., a donut, a U or a V shape, will result in small ρ_ss with, still, a low entropy. Hence, the entropy provides the information that dependence exists only that it does not specify the type e.g., linear, non-linear, monotonic, or non-monotonic. The solution to such a problem while modeling is that the model should produce series that have similar entropies and no significant bias when considering an ensemble compared to the entropy of the observations. An important thing to take into account while using Eq. 3.19 is to discard all values of $c[j,k]$ that are zero because of the logarithm. Such an action is justified as the first term, $c[j,k]$, will be zero anyhow. Similar to copula asymmetries, entropies have to be normalized as well. The maximum possible entropy (H_{max}) is,

$$H_{max} = -\ln\left(\frac{1}{M^2}\right) \tag{3.20}$$

Note that unlike the asymmetries, H_{max} is independent of ρ_s. It represents the entropy of the independence copula. The minimum possible entropy is zero. This is counter intuitive, assuming that the minimum entropy possible is when all points lie in a number of cells that is equal to M (assuming same bin sizes in each dimension) of the copula which would result in a minimum entropy of,

$$-\ln\left(\frac{1}{M}\right)$$

However, it is also possible that almost all values lie in only one cell of a copula. This could be the case when two precipitation time series are compared and both of them have zeros only with the exception of a single time step. For that case, the entropy is almost zero. Similar to the asymmetries, the entropy used from here on is also normalized unless otherwise specified and is given by,

$$H = \frac{H_{raw}}{H_{max}} \tag{3.21}$$

Fig. 3.11 shows some empirical copulas along with their respective 2D relative histograms, ρ_p, ρ_s and entropies (H). The interesting case is that of the diamond, where the correlations are zero but the entropy is lower than the second case where the correlations are close to one. Fig. 3.7 shows how it might look like for a time series.

Similar to Shannon's entropy, another very similar type of entropy is the Mutual Information (MI). What sets it apart from H is that it can be computed locally, i.e., for each time step, with emphasis on the bin densities that given points correspond to, in their respective distribution functions. In essence, it assigns a *discrete* value to each time step, categorizing it in such a way that values that correspond to dense areas in their distribution functions weigh higher while values that are from less denser regions get a smaller weight, comparatively, because of the denominator in the logarithm. Sensitivity of this measure can be increased by increasing M i.e., points falling in bins with higher relative density result in an even higher local MI. Formally, for the empirical case it can be, along with some additional terms, written as follows. The formulation presented here is a modified version of the original MI that is adapted for use in copulas.

$$b[t] = \lfloor u^t M \rfloor \tag{3.22}$$

$$c[u^t] = \frac{1}{N} \sum_{t=0}^{N-1} \begin{cases} 1, & \text{if } b[t] = \lfloor u^t M \rfloor \\ 0, & \text{else} \end{cases} \tag{3.23}$$

$$c[u_1^t, u_2^t] = \frac{1}{N} \sum_{t=0}^{N-1} \begin{cases} 1, & \text{if } \left(b_1[t] = \lfloor u_1^t M \rfloor\right) \wedge \left(b_2[t] = \lfloor u_2^t M \rfloor\right) \\ 0, & \text{else} \end{cases} \tag{3.24}$$

$$MI(t) = c[u_1^t, u_2^t] \ln \left(\frac{c[u_1^t, u_2^t]}{c[u_1^t] c[u_2^t]} \right) \tag{3.25}$$

where, $\lfloor x \rfloor$ is the floor operation on x, $b[t]$ can range from 0 to $M-1$, $c[u^t]$ is the time series of the relative frequencies of $b[t]$ and $c[u_1^t, u_2^t]$ is the time series of empirical copula bin densities when u_1^t and u_2^t are in the same cell of the discretized bivariate copula.

In theory, $c[u^t]$ should be uniformly distributed. In practice, this does not hold and it does have an additional effect on MI. For example, consider the case of discharge. The stage is always rounded to the nearest centimeter or so to get the flow rate through the rating curve. The resulting time series has discretized numbers where the same flow rates get the same rank and, consequently, the same non-exceedence probability. This phenomenon becomes much more apparent for low flows that vary little overtime as compared to the high ones. For precipitation, it becomes much more important to take this in to account because of the zeros or very small values. Fig. 3.12 shows how MI might look like for a discharge time series where u_2 leads u_1 by a single time step. Notice how the low flow regions get larger values of MI as compared to the rest. It is because of the term in the denominator, which reduces the weight of smaller density cells much more as compared to Shannon's entropy. This property will be exploited to simulate better time series later.

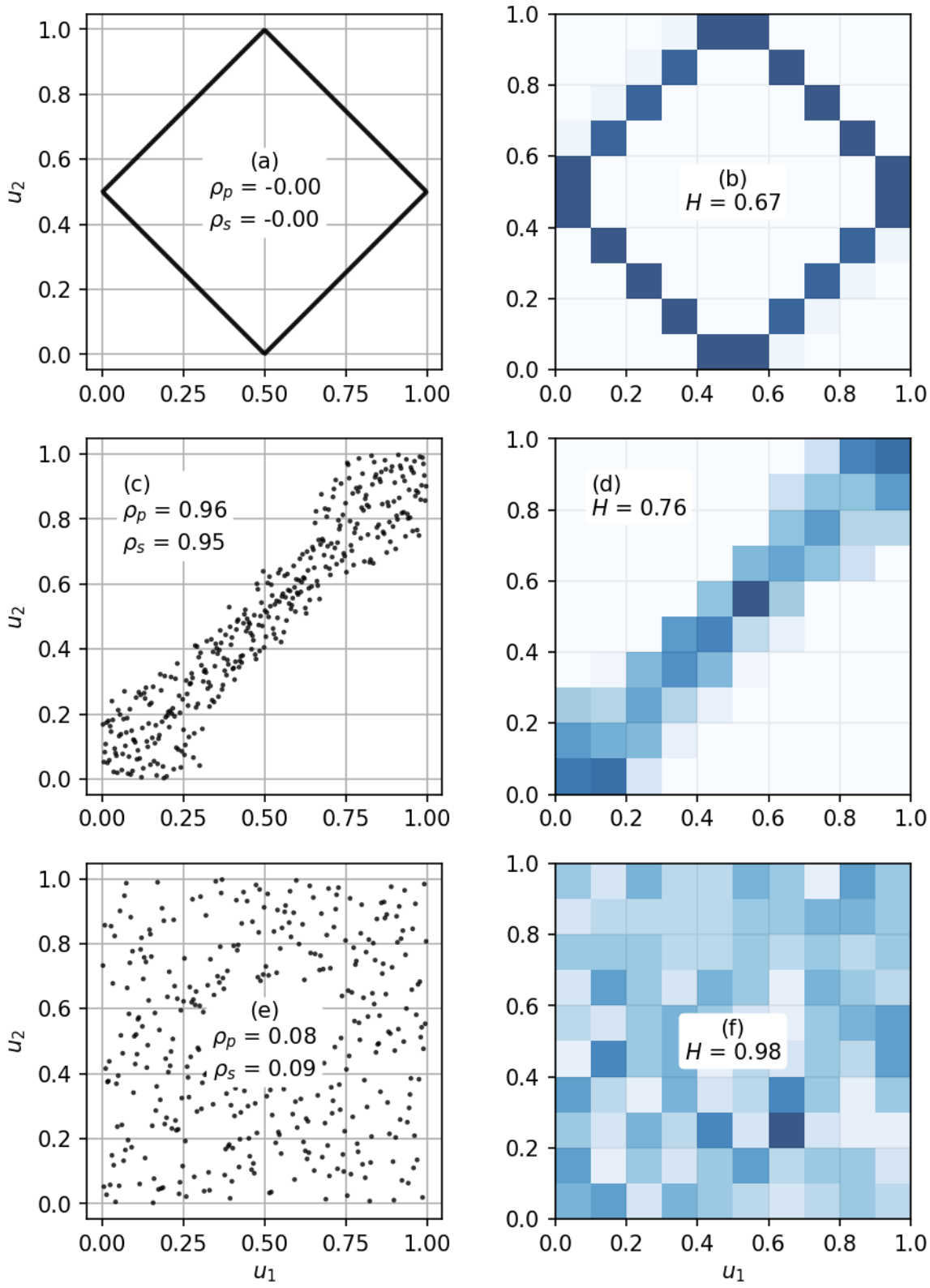


Figure 3.11: Empirical copulas with corresponding 2D relative histograms and the resulting entropies. Cases (a) and (b) are for the diamond shaped copula, (c) and (d) for two linearly dependent variables while (e) and (f) represent the no dependence case.

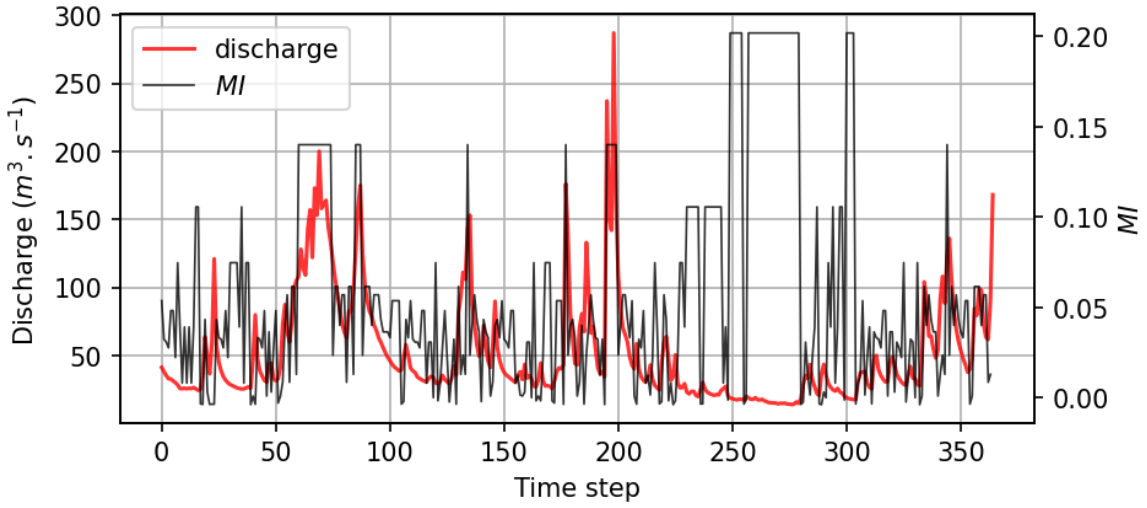


Figure 3.12: A hydrograph along with its corresponding Mutual Information (MI) time series calculated using a lead step of 1 with respect to itself.

3.7 d-dimensional Spearman correlation coefficient

So far, measures describing bivariate dependence were discussed. It is more interesting to have the strength of dependence for the case of d dimensions. For copulas, Wolff (1980) proposed a measure that can take into account a multivariate copula or time series of non-exceedence probabilities/grades and produce a single measure of the strength of dependence. Two versions of the d-dimensional Spearman's correlation coefficient exist. Their computationally friendly forms are given in Schmid and Schmidt (2007) which are derived from Wolff (1980) and Ruymgaart and van Zuijlen (1978). It can be viewed as the average distance of a copula from that of the independence copula. The equation for a d-dimensional copula is given as follows,

$$\rho_d = \frac{d+1}{2^d - d - 1}\left(2^d \int C(u)du - 1\right) \tag{3.26}$$

It collapses to ρ_s for the 2-dimensional case. The term $C(u)$ is difficult to compute in d dimensions even though it is more straight forward to understand. Later, Ruymgaart and van Zuijlen (1978) proposed another version which is much easier to calculate numerically. It is given as follows,

$$\tilde{\rho_d} = \frac{d+1}{2^d - d - 1}\left(2^d \int \prod(u)\, dC(u) - 1\right) \tag{3.27}$$

where, the term $\prod(u)$ refers to the product of all us at any given point in the copula. A problem exists while using ρ_d or $\tilde{\rho_d}$, namely that they are unequal except for the case when the copula is symmetric in all directions i.e., AO and AD are both zero. The mismatch in values can also be exploited to indicate if the copula of given data is non-Gaussian. Yet

another issue is that the lower Fréchet-Hoeffding bounds are still not defined for copulas in more than two dimensions. According to Wolff (1980), the lowest possible correlation among variables is not -1 at minimum, as one would expect, but a larger value for cases with d greater than two. As the number of dimensions increases the least possible correlation tends to 0 but cannot be explicitly defined. This issue has implications while normalizing copula asymmetries in higher dimensions due to the unknown lower limit of the correlation.

4 Study Area, Rainfall-Runoff Modeling and Experimental Setup

All the data used in this thesis is from locations in the Federal state of Baden-Würrttemberg (BW) of Germany. BW (Fig. 4.1) is the most South-western state of Germany that borders Switzerland to the South and France to West. To the East lies the Federal State of Bavaria with Rheinland-Palatinate and Hessen bordering it on the northern side. Its total area is about 36,000 km^2. The famous Black Forest area covers its south-western side. The River Danube originates in the Black Forest and flows out towards Bavaria. One of the largest rivers in Europe, the Rhine River, originates in the Swiss Alps which forms the border between Germany, France and Switzerland as it flows towards the Netherlands.

The Neckar River (Fig. 4.2) is a tributary of the Rhine River. It originates in the Swabian Alps that lie in the center of BW. The river then flows North towards Tübingen, then Stuttgart and then falls in to the River Rhine at the city of Mannheim. The catchment area of the Neckar River is ca. 14,000 km^2, which is about half the area of BW while its running length from source to mouth is ca. 300 km.

4.1 Data

Daily time series of precipitation, temperature and discharge are considered at various locations inside and around the study area. In some cases, hourly precipitation is also considered. The time period for daily data is from 1961 to 2015 while for the hourly it is from 2010 to 2015. Furthermore, the discharge gauge locations are solely considered in the Neckar River catchment as it lies completely inside the study area.

There are about 900 precipitation gauges having daily data records throughout the study area (Fig. 4.3). All of these are, of course, not active at any given time step. The total number of active gauges fluctuates depending on the time period. For example, 550 gauges were active for the period 1961-1970 with a steady decrease in the following years that reached to only 156 active gauges for the period of 2001-2010. The mean precipitation per year is 1000 mm with a relatively small annual cycle. The network used here is maintained by the government agency Deutscher Wetterdienst (DWD; the German Weather Service; DWD, 2019). The density of the hourly gauges (Fig. 4.4) is not as high as that of the daily. Temperatures go well below freezing in the winter with snow events between the months of December and March.

Fig. 4.3 also shows groups of five gauges that are used in the coming chapters for the demonstration of spatio-temporal dependence on the daily temporal scale. These series have very

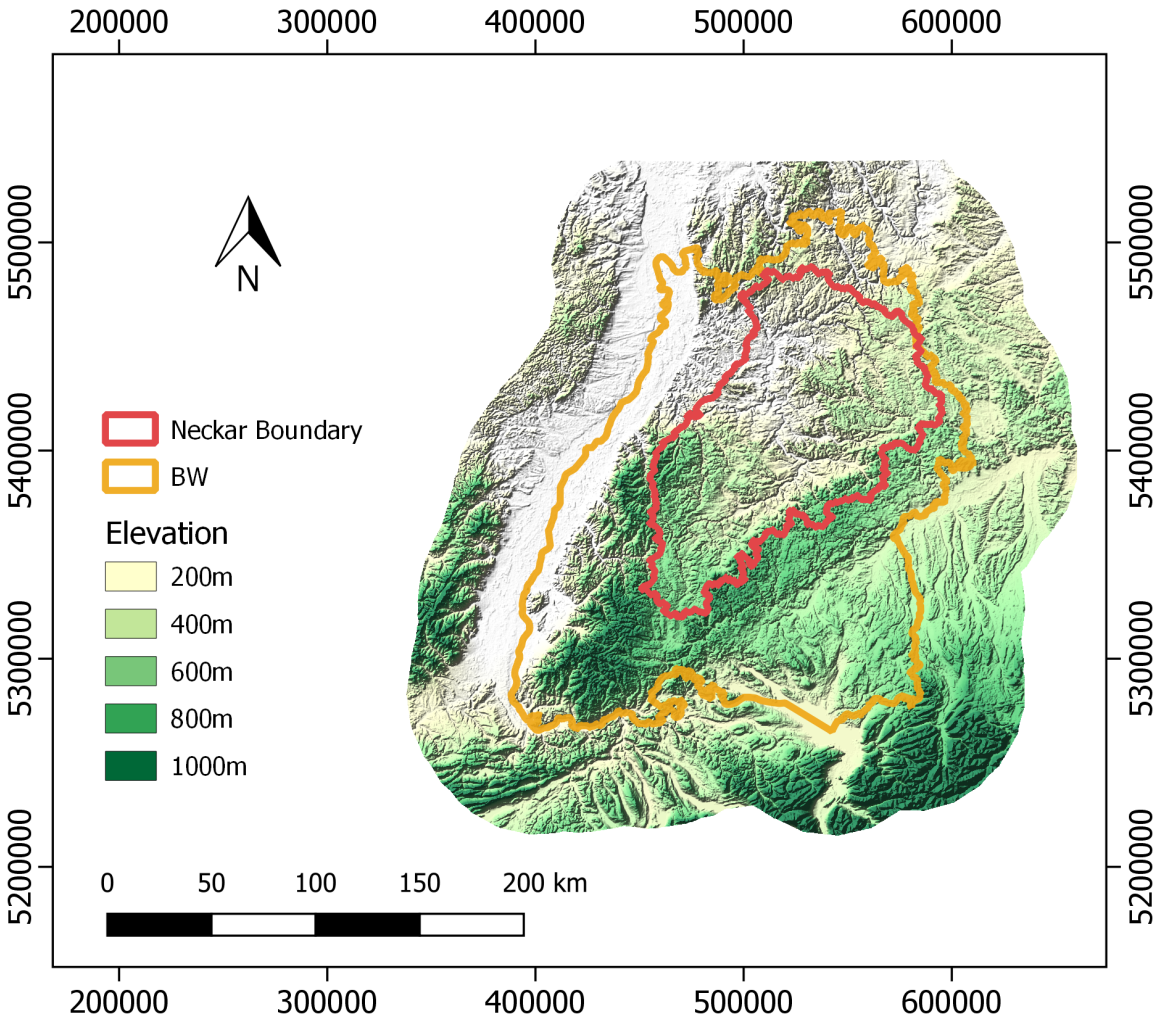


Figure 4.1: Study Area.

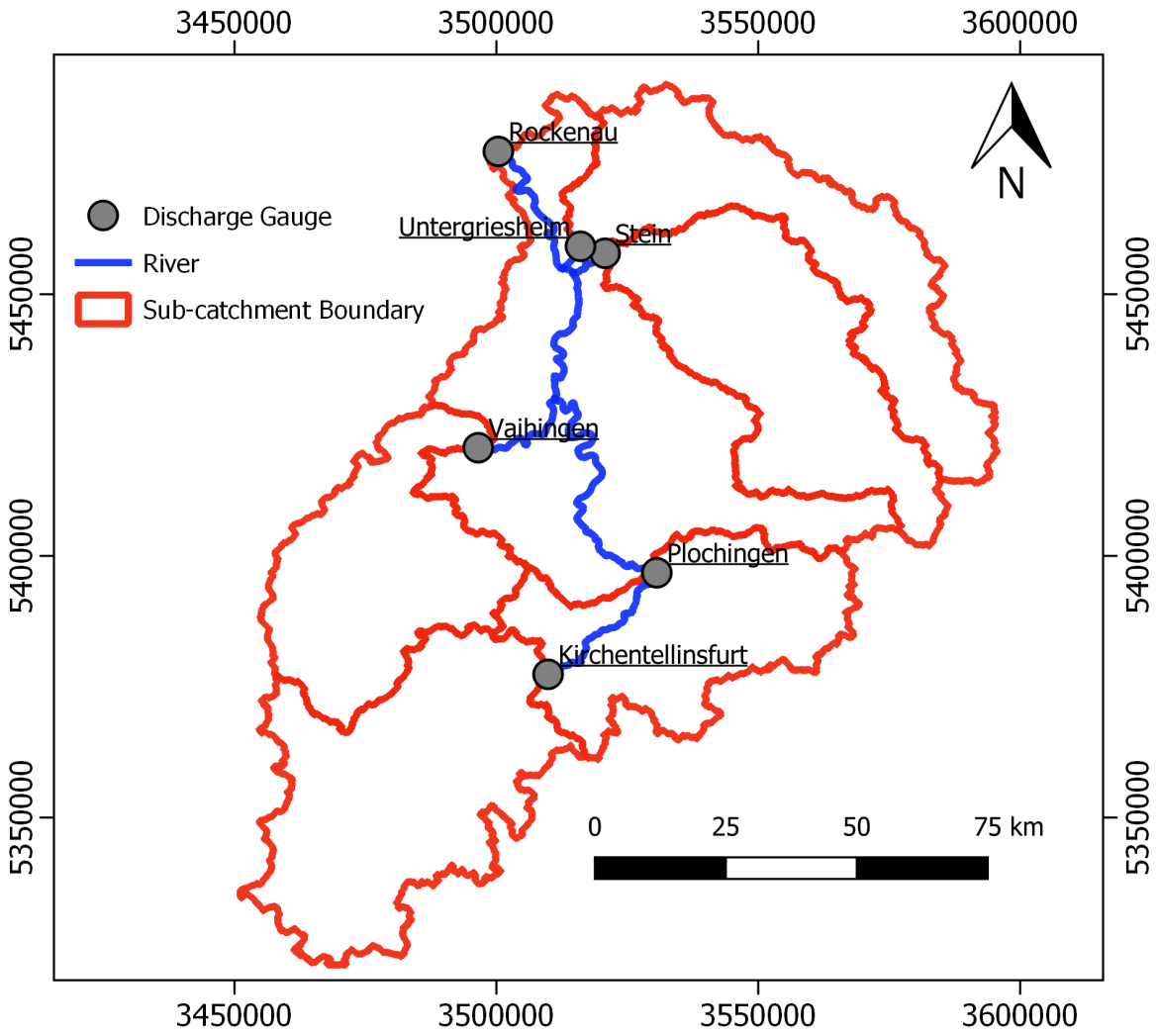


Figure 4.2: The Neckar River and its tributaries.

long continuous records i.e., 1961 to 2015, and are hence chosen for demonstration. Specifically, set A (red) has gauges that are located in the west of the study area and have a mean distance of ca. 39 km among them, while set B (blue) is located in the north-east having a mean distance of ca. 32 km among the gauges, while set C (black) contains gauges from all corners of the study area with a mean distance of 161 km among them. Note that set A and C share a gauge (second gauge from the top in set A and C). Similarly, Fig. 4.4 shows a set of gauges that is also used in further analysis for the hourly case. The time series of gauges in set D (pink) is from 2010 to 2015 with a mean distance of 33 km among the gauges.

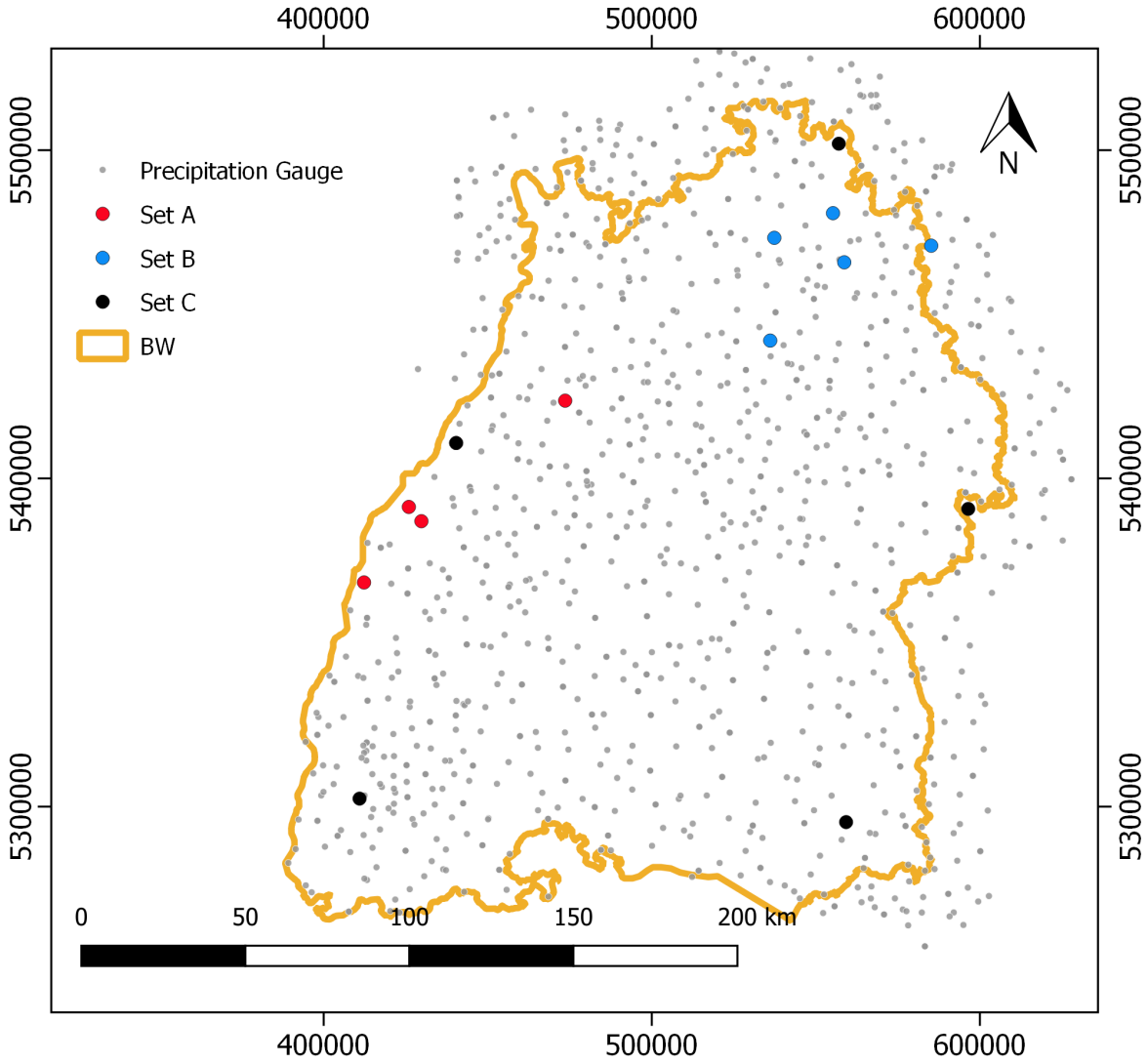


Figure 4.3: Distribution of precipitation gauges with daily records with various sets of gauges (sets A, B, C) that are chosen for demonstration of spatio-temporal properties of groups of points.

Sub-catchment discharge varies depending on the location of the gauging station (Fig. 4.2). The main section of the river from Plochingen to Rockenau is used for transport by ships and has structures built to regulate the depth of flow. This section, unfortunately, cannot be used

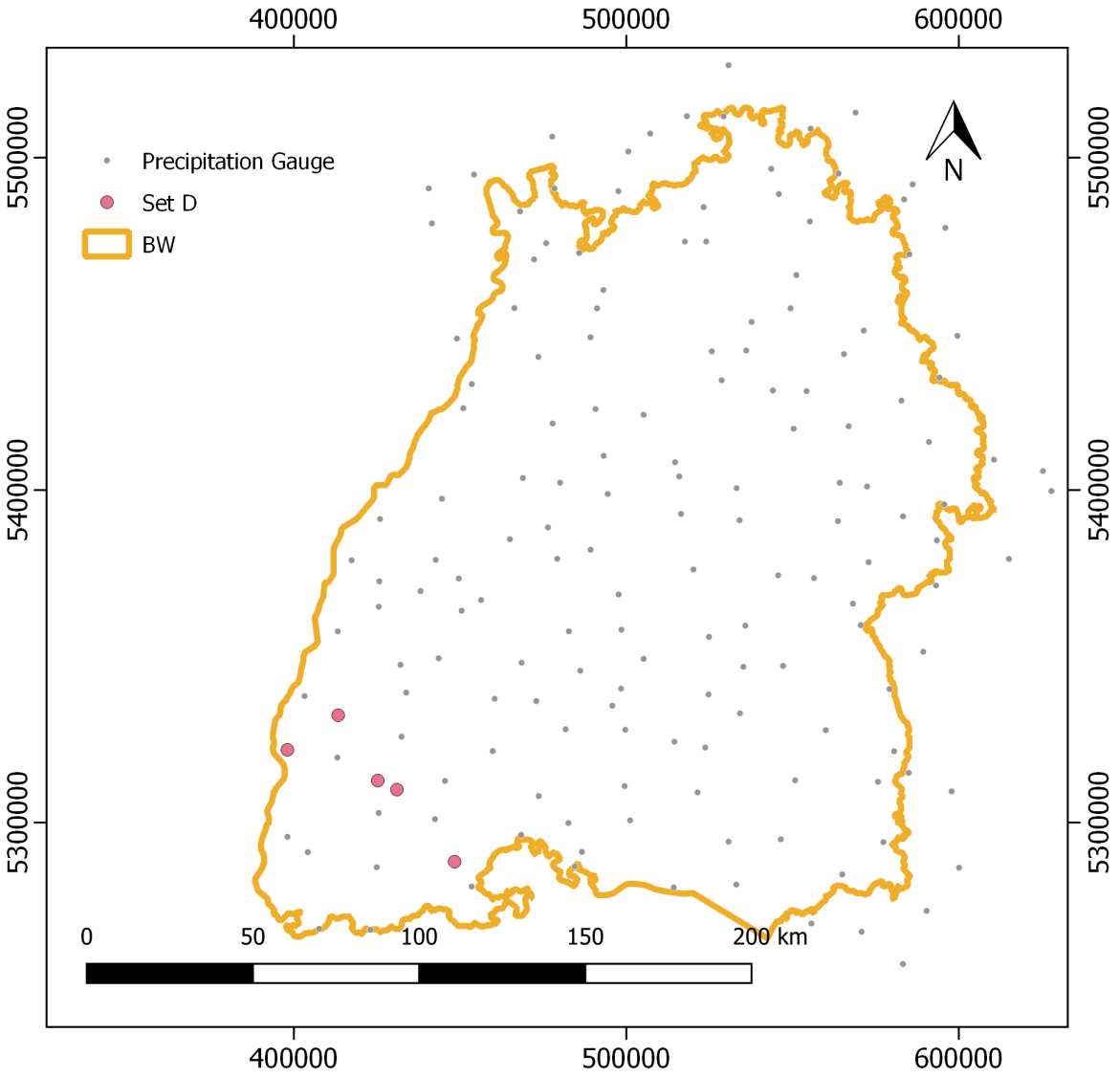


Figure 4.4: Distribution of precipitation gauges with hourly records with a set of gauges (set D) that is chosen for demonstration of spatio-temporal properties of groups of points.

for modeling as the flows do not respond correctly to rainfall/snowmelt events at the daily temporal resolution. Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg (LUBW; Environmental protection agency of Baden-Württemberg; LUBW, 2020) is the agency responsible for maintaining the gauges in BW. The time series for earlier time periods has a daily resolution but most gauges have been updated to produce an hourly or even a 15-minute resolution. Four sub-catchments of the Neckar were selected to serve as the basis for studying, comparison and simulation. These are Kirchentellinsfurt (Neckar), Vaihingen (Enz), Stein (Kocher) and Untergriesheim (Jagst). The criteria for selection of these four was the length of the records, the sub-catchment area, being a head-water sub-catchment, the quality of the time series and susceptibility to human intervention. For the case of daily resolution, it is important that the time of concentrations at any point of interest be 24 hours or more. If the catchments are so small that a peak is observed at the same time step at which the peak of a precipitation event took place then a rainfall-runoff model can not simulate the event at the correct time step as it needs at least one time step to propagate the effect of input precipitation as discharge. The details of these sub-catchments are shown in Table 4.1 and Fig. 4.2.

Table 4.1: Properties of the Neckar sub-catchments

Catchment	**Area** (km^2)	**Mean flow** ($m^3 \cdot s^{-1}$)	**Min. flow** ($m^3 \cdot s^{-1}$)	**Max. flow** ($m^3 \cdot s^{-1}$)
Kirchentellinsfurt	2320	26.9	0.9	538.0
Vaihingen	1656	19.8	3.0	424.0
Stein	1943	22.8	2.6	572.0
Untergriesheim	1829	17.7	1.3	504.0

For rainfall-runoff modeling, daily precipitation was interpolated over the whole Neckar catchment on a grid with 1 km spatial resolution using the External Drift Kriging (EDK; Ahmed and de Marsily, 1987) method with elevation as the drift. Similarly, minimum, mean and maximum temperatures were also interpolated on a daily basis. Potential evapotranspiration was then subsequently computed using these values with the Hargreaves-Samani equation (Hargreaves and Samani, 1982). These variables were then lumped as a daily time series for each considered sub-catchment that were then used as inputs for the rainfall-runoff model.

4.2 The HBV rainfall-runoff model

The HBV (Bergström, 1992) is a well known conceptual rainfall-runoff model (Fig. 4.5). Based on its history, e.g., Das et al. (2008); Götzinger and Bárdossy (2007); Hundecha and Bárdossy (2004) in this study area and simplicity, a slightly modified version that conserves mass is presented. It needs precipitation (P), temperature (T), and potential evapotranspiration (PE) time series as input. It can be run in a spatially lumped or a distributed configuration. A schematic diagram and the equations of a lumped configuration are

presented here. The Nash-Sutcliffe Efficiency (Nash and Sutcliffe, 1970) (NSE) and Log-Nash-Sutcliffe Efficiency (LnNSE) were used as objective functions while optimizing for best model parameters using the Differential Evolution (Storn and Price, 1997) optimization scheme.

Snow melt and accumulation

$$ME_i = max\left(0.0, \Big(CM_T + CM_P(P_i)\Big)(T_i - TT)\right) \tag{4.1}$$

$$SN_i = \begin{cases} SN_{i-1} + P_i & \text{if } T_i <= TT, \\ SN_{i-1} - ME_i & \text{else.} \end{cases} \tag{4.2}$$

$$LP_i = \begin{cases} 0.0 & \text{if } T_i <= TT, \\ P_i + min(SN_{i-1}, ME_i) & \text{else.} \end{cases} \tag{4.3}$$

where, the subscript i is the index of a given day, CM_{TE} is the snow-melt due to increase in temperature in $mm \cdot^{\circ} C^{-1} \cdot day^{-1}$, P_i is the precipitation in $mm \cdot day^{-1}$, CM_P is the snow-melt due to falling liquid precipitation in $mm \cdot^{\circ} C^{-1} \cdot day^{-1} \cdot mm^{-1}$, T_i is the temperature in $^{\circ}C$, TT is the threshold temperature below which the precipitation falls as snow, ME_i is the possible snow-melt in mm, SN_i is the total accumulated snow in mm, LP_i is the liquid precipitation in mm that might come from snow-melt or precipitation or both.

Evapotranspiration and soil moisture

$$AM_i = SM_{i-1} + LP_i\Big(1 - \left(\frac{SM_{i-1}}{FC}\right)^{\beta}\Big) \tag{4.4}$$

$$ET_i = \begin{cases} min(AM_i, PE_i) & \text{if } SM_{i-1} > PWP, \\ min\Big(AM_i, PE_i\left(\frac{SM_{i-1}}{FC}\right)\Big) & \text{else.} \end{cases} \tag{4.5}$$

$$SM_i = max(0.0, AM_i - ET_i) \tag{4.6}$$

where, SM_i is the soil moisture in mm, FC is the field capacity in mm, PWP is the permanent wilting point in mm, β is a unitless constant related to the soil's ability to retain moisture, AM_i is the available soil moisture in mm, PE_i is the potential evapotranspiration in $mm \cdot day^{-1}$, ET_i is the actual evapotranspiration in $mm \cdot day^{-1}$.

Upper reservoir runoff routing

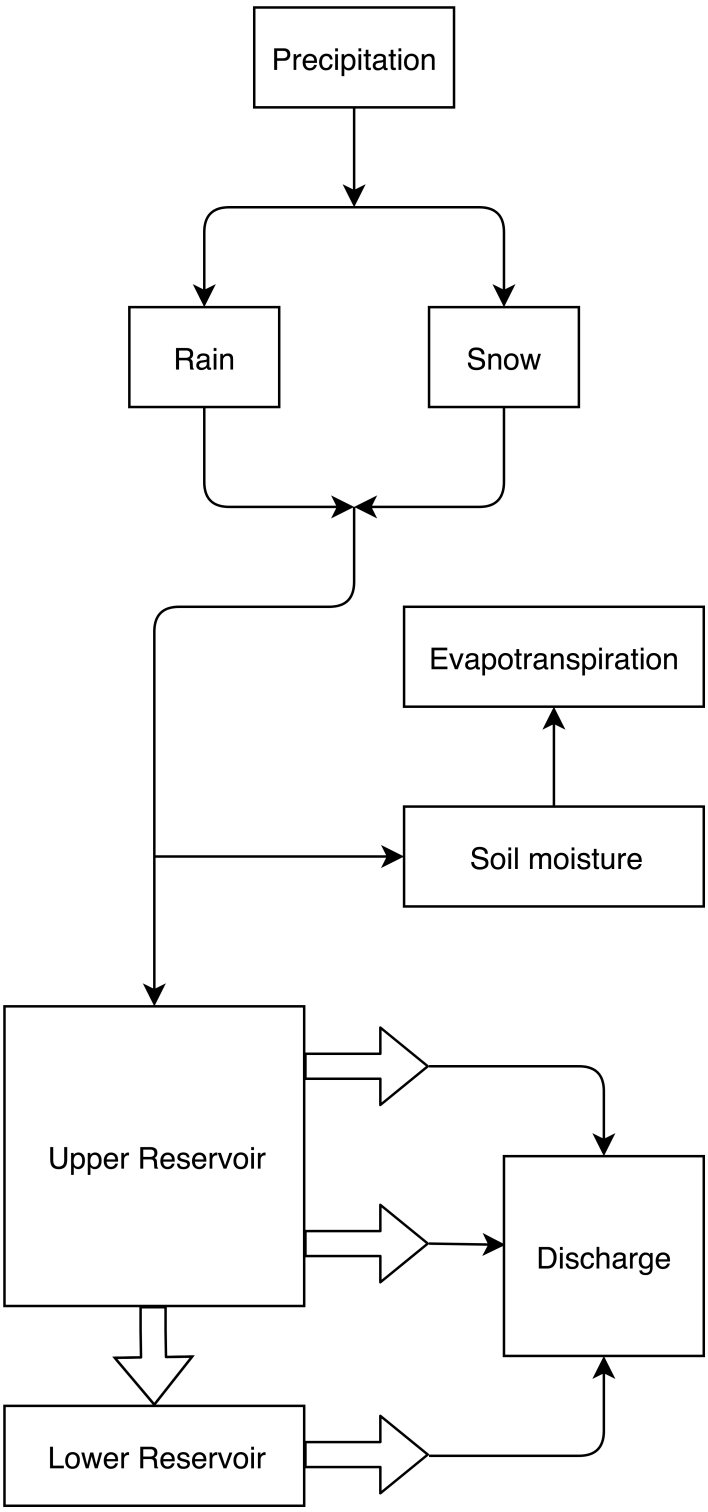


Figure 4.5: The HBV model.

$$RN_i = LP_i \left(\frac{SM_{i-1}}{FC} \right)^{\beta} \tag{4.7}$$

$$URUO_i = max\Big(0.0, K_{uu}(URST_{i-1} - UT)\Big) \tag{4.8}$$

$$URLO_i = max\Big(0.0, K_{ul}(URST_{i-1} - URUO_i)\Big) \tag{4.9}$$

$$URLR_i = max\Big(0.0, K_d(URST_{i-1} - URUO_i - URLO_i)\Big) \tag{4.10}$$

$$URST_i = max\Big(0.0, (URST_{i-1} - URUO_i - URLO_i - URLR_i + RN_i)\Big) \tag{4.11}$$

where, RN_i is the runoff in $mm \cdot day^{-1}$ i.e., the amount of water that is not retained by the soil and is available for routing through the model's reservoirs, $URST_i$ is the upper reservoir storage in mm, UT is the storage threshold in mm above which quick runoff from the upper outlet of the reservoir should take place. K_{uu} is the upper reservoir upper outlet's runoff coefficient in day^{-1}, $URUO_i$ is the runoff in $mm \cdot day^{-1}$ from the upper reservoir upper outlet, K_{ul} is the upper reservoir lower outlet's runoff coefficient in day^{-1}, K_d is the coefficient of runoff transfer from the upper to lower reservoirs in day^{-1}, $URLO_i$ is the runoff from the upper reservoir's lower outlet in $mm \cdot day^{-1}$.

Lower reservoir runoff routing

$$LRO_i = K_{ll}(LRST_{i-1}) \tag{4.12}$$

$$LRST_i = LRST_{i-1} + URLR_i - LRO_i \tag{4.13}$$

where, $LRST_i$ is the lower reservoir storage in mm, K_{ll} is the lower reservoir runoff coefficient in day^{-1}, LRO_i is the runoff from the lower reservoir in $mm \cdot day^{-1}$.

Simulated discharge

$$QS_i = CC(URUO_i + URLO_i + LRO_i) \tag{4.14}$$

where, CC is a conversion constant that converts mm/day to $m^3 \cdot sec^{-1}$, QS_i is the simulated discharge in $m^3 \cdot sec^{-1}$.

4.3 Setting up experiments to test for extremes

High flows are produced by high precipitation in one way or another. Be it large intensity precipitation for a short time or a medium precipitation for an unusually long time or a significant amount of snow melting due to rise in temperature. As was stated in the beginning of this thesis, the aim here is to have a time series generator that produces time series that when used by themselves or as inputs for a rainfall-runoff model would/should produce extremes, high precipitation or discharge, such that they match the observations in terms of statistics. Not only that, these generators should also preserve properties of time series that correspond to the natural behavior such as seasonal and annual cycles. More precisely, at any given point in space the distributions of the temporal aggregations, e.g., from hourly to daily or from daily to weekly or monthly, of the simulated series should be similar to those of the observed i.e., no consistent/significant bias towards very low of very high values. Similarly, at any given time step the distribution of the sum of all values of a given variable at various locations of all simulations should be similar to that of the observed with no significant biases. Although the word *extreme* includes both high flows and droughts, here only floods are dealt with.

As high flows are a result of precipitation (rainfall or snow-melt), precipitation and discharge are tested only. The distribution of high values can be assessed for both precipitation, which indirectly induces floods, and discharge, which is the main variable of interest in this study. These two variables require different setups due to their properties. Both of these are described in detail in the following subsections along with properties that the series should preserve.

4.3.1 Precipitation

Various *setups* are proposed here to test if the spatio-temporal behavior of the overall time series and especially that of the extremes is reproduced in a manner similar to that of a given reference, which in this case are the properties of observed time series of precipitation, temperature, potential evapotranspiration and discharge at daily and hourly resolutions. As was shown earlier, many gauges with long records in time are available, various combinations of properties can be used. It is proposed here that clusters of five stations that are not very far away from each other be simulated simultaneously and a group of five stations that are as far away as possible from each other in the study area. For the daily resolution, the time periods vary from 10 years to 55 years while 6 years only for the hourly. This affects the amount of gauges that can be used as all of them do not have continuous records for any selected time period. A *setup* is defined as a possible combination of various variables to be simulated either independent of each other simultaneously or by having a set of cross-properties among them. Properties considered during simulation could be the previously mentioned ones in the copula domain or the ones that are introduced next in the coming chapters. For example, precipitation is simulated such that the sum of the squared-differences of the auto-properties (Pearson's correlation (ρ_p), Spearman's correlation (ρ_s), Copula order asymmetry (AO) etc.) at various lags between the simulated and observed are minimized. Another example is that of precipitation, temperature and potential evap-

otranspiration simulated simultaneously such that both the sum of the squared-differences between the auto- and cross-properties of the reference and simulation are minimized. The details of how this is done exactly will be presented in the coming chapters. Each setup is simulated 100 times, so that enough values are produced to see a wide range of possibilities of extreme events i.e., the upper tails of various distributions.

After simulation, three properties are assessed namely, the distributions of various temporal aggregations at any location, the distribution of spatial-sums at any given time step and the distribution of spatial-sums of the aggregated time series. A spatial-sum series refers to the sum of the values of all series at each time step. It is assumed that properties that were used as objective functions during optimization should conserve the aforementioned three properties indirectly.

4.3.2 Discharge

Stream flow is a much more deterministic response of a catchment to precipitation and its environment. It is difficult to mimic it through statistics as compared to temperature and precipitation. An attempt is made nevertheless. Exactly the same setup as that used for precipitation is taken for discharge as well, albeit for the four aforementioned sub-catchments of the Neckar River. In addition, a much more elaborate experiment is also set up to verify the simulation results. This is detailed below.

Consider a simple rainfall-runoff modeling scenario. Precipitation, temperature, and potential evapotransipiration are prepared as inputs for a rainfall-runoff model. Observed discharge along with an optimization scheme is used to find the model parameters that produce output discharge of an acceptable quality . This model is then subsequently used either with forecasted data or with simulated data to see the system's response.

Suppose, the input data at each point is shuffled for all series independent of each other i.e., no spatial or temporal structure and then used as input and an attempt at model calibration is made afterwards. As the inputs have no temporal relation with the observed discharge, it is very unlikely that a good objective function value (sum of the squared-differences of the properties of reference and simulation) similar to the original case (the one with the measured inputs) is achieved. If it does achieve a performance similar to that of the original case then it means that either the inputs or the model or the observed discharge or everything was wrong to begin with. This case is also tested later and presented. What a time series generator does is that it also shuffles the values randomly, with one major difference that the shuffling is constrained (time series generators can be viewed as *constrained scramblers*). Constrained in such a way that the new series do have properties that are common with or similar to those of the observed. Here, the assumption is that if a set of properties is kept by the simulated time series and these are then used as inputs for the rainfall-runoff model, then the properties or performance of the model output discharge should be close to those of what was simulated by the model with observed input data. Next, the comparisons are made with the original model output discharge, rather than the observed, and the model output discharge that was obtained using simulated inputs. This is important because no model outputs match observations fully after calibration.

The aforementioned method does create some new problems. Instead of only using precipitation for simulation, other variables (temperature and potential evapotranspiration) have to be simulated simultaneously. Doing so serves two purposes. First, even though temperature and precipitation are not very related, the simultaneous simulation of these will result in whatever correlated behavior these two might have, thereby producing better series. Second, it allows for the freedom of using the same temperature and potential evapotranspiration for all model runs with changing precipitation, which may or may not bias the final model discharges.

4.3.3 Properties considered during simulation

One important aspect that was left out during the experimental setup is that of the properties used while simulating. One always has to probe the subset of properties to conserve given the large number of choices. Ideally, all should be conserved but this has two major drawbacks. One, not all properties can be kept using the state of the art generators. Two, the computational time increases exponentially with the number of properties that are to be conserved during simulation/optimization. Nevertheless, one has to find out manually which properties to conserve. Imagine that auto-correlation function which forms the basis of most generators turns out to be not that important. What if the Spearman's correlation is important to keep rather than the Pearson's or vice versa? Some new properties along with existing ones are presented in the later chapters in detail. In short and ideally, the properties considered during simulation should be a function of the application for which the generated data are used. For example, if the discharge volumes in a reservoir are of interest (e.g., for irrigation) then the monthly mean discharges values are of interest and their behavior at the daily or hourly scale is unimportant. However, for the same reservoir the peak flows may also be of importance so that the spillways could be dimensioned correctly. In this case, the maximum hourly peak flows would be of importance. Another very important aspect that has to be considered while simulating is that all the simulated series with a given set of properties should have the same chances of occurrence with respect to each other.

5 Phase Randomization: Existing and New Variants

Before presenting Phase Randomization, the Fourier transform, is defined because the former depends on later. Similar to what was described in Chapter 3, a transition from marginal to frequency domain is made. Any given time series can be written as a composition of several frequencies, each having its respective amplitude and phase. The amplitude defines the relative influence of the frequency on the time series while the phase defines its starting point. This formulation was described by Baron Jean-Baptisite-Joseph Fourier (1768-1830) in Fourier (1822). In essence, the forward Fourier transform decomposes any given signal in to waves such that when combined at any point in time, the result is the value of the original signal. Throughout this thesis, the Fourier transform of a one dimensional real valued series is utilized.

5.1 The Fourier transform

Suppose a variable X, composed of N real values, then its forward Fourier transform (FT) is given by,

$$FT_X(k) = \sum_{t=0}^{N-1} x(t) \exp\left(\frac{-2\pi jtk}{N}\right) \quad \text{for } k = 0, \ldots, N-1 \tag{5.1}$$

And the backward transform is given by,

$$x(t) = \frac{1}{N} \sum_{k=0}^{N-1} FT_X(k) \exp\left(\frac{2\pi jtk}{N}\right) \quad \text{for } t = 0, \ldots, N-1 \tag{5.2}$$

where, k is the frequency, $x(t)$ is the value of X at time step t, and j is $\sqrt{-1}$ with each term in Eq. 6.1 being a complex number except for the first term which is the sum of all the values of X. Throughout this thesis, the first value is not used as it only serves to shift the location of all the values and also because the focus is on the temporal/spatial variance of the variables.

Each term on the left side of Eq. 6.1 is called a *Fourier coefficient* of the given frequency and is a complex number. Two spectra, namely the magnitude spectrum (MAG) and the phase spectrum (PHS), are given by the following equations respectively,

$$MAG_X(k) = ||FT_X(k)|| \quad \text{for } k = 1, \ldots, N-1 \tag{5.3}$$

$$PHS_X(k) = \angle FT_X(k) \quad \text{for } k = 1, \ldots, N-1 \tag{5.4}$$

Each frequency's phase determines where the wave starts on the y-axis and its magnitude determines how high or low it may be. Similarly, another spectrum that defines the variance of the series in terms of frequencies is the power spectrum (POW) and can be written as,

$$POW_X(k) = \big(MAG_X(k)\big)^2 \quad \text{for } k = 1, \ldots, N-1 \tag{5.5}$$

The power spectrum is also called the *periodogram*. Eq. 6.2 has an interesting property i.e., it relates to the auto- or cross-correlation function of time series. In other words, a change in the phases can be made without disturbing the mean, variance and the auto-correlation function of the series. The equation that neatly presents this relationship is called the Wiener-Khinchin theorem (Chatfield, 2016) and is written as follows for converting the auto-correlation function ($\rho_p(t)$) to the power spectrum,

$$POW_X(k) = \sum_{t=0}^{N-1} \rho_{p_X}(t) \exp\left(\frac{-2\pi jtk}{N}\right) \quad \text{for } k = 0, \ldots, N-1 \tag{5.6}$$

where, $\rho_{p_X}(t)$ refers to the auto-ρ_p computed with t lead steps. Another way of visualizing this relationship is that the power spectrum is the forward Fourier transform of the auto-correlation function of a time series. Similarly, the auto-correlation function can be derived by taking the backward Fourier transform of the power spectrum after setting the value of the first coefficient (at frequency 0) to zero. It is given as follows,

$$\rho_{p_X}(t) = \frac{1}{N} \sum_{k=0}^{N-1} POW_X(k) \exp\left(\frac{2\pi jtk}{N}\right) \quad \text{for } t = 0, \ldots, N-1 \tag{5.7}$$

The values in Eq. 5.7 may need to be normalized by the first term in the spectrum to get the correlations in the range of $\pm\,1$ as the transformation actually produces the auto-covariance, which depending on the values may or may not have the same range as ρ_p. This theorem also has an important benefit that is outside the scope of this thesis but is mentioned anyway. For time series with missing values, it is not possible to compute the Fourier transform. But it is possible to get the auto- correlation or covariance function. Assuming that the distribution of the data is the same as if it were having no missing values, the power spectrum and subsequently the magnitude spectrum can be computed. Now to get the missing values, a constrained optimization scheme can be written in such a way to find the phase spectrum such that it would produce exactly the same values at the known time steps while constructing the values at the missing time steps as a side-effect.

Visualizing the Fourier transform

Consider a time series of a variable X (Fig. 5.1). Its magnitude spectrum is symmetrical around the center. This is a useful side-effect of the Fourier transform. It wouldn't be the case if the inputs were composed of complex numbers. For the phase spectrum, this is not exactly the case. There, the corresponding phases are shifted by 180°. This is yet another feature i.e., the Fourier coefficients at indices k and $N - k$ are complex conjugates of each other for real valued inputs e.g., the coefficient at index 1 is the complex conjugate of the coefficient at the index $N - 1$. This property is exploited by algorithms when computing the Fourier transform of a series because the first half of the terms are with the opposite sign of the second half. The resulting magnitude and phase spectra using these algorithms yield only half of the spectrum whose length is *almost* half of N, depending on the number of steps being odd or even. For odd steps it is $0.5N + 1$. Throughout this thesis only series with even steps i.e., the last step is dropped if the steps are odd which makes the length of the magnitude and phase spectra one half of the original length of the input time series.

5.2 Phase Randomization

Following what was observed in Eq. 6.2, a new series can be generated that has the same first- and second-order moments as the original input i.e., an arbitrary phase spectrum can be generated (except for the phase at index 0) and used to recalculate the Fourier coefficients and a new time series can be created using Eq. 5.2. This procedure is called *Phase Randomization (PR)*. Generating new time series aka *surrogate series* is, among others, one of its uses (Theiler et al., 1992). It is this property that is exploited to generate series. Similarly, many series at multiple locations can be generated while conserving their auto- and cross-correlation functions. This may be achieved by generating one single random phase spectrum that is added to all the existing phase spectra. This results in the same difference of the phases between multiple locations and consequently their cross-covariances are also preserved. Details of this procedure is presented in the coming subsections. Fig. 5.2 shows three realizations, surrogate series, that were obtained using PR. It should be noted that surrogate values follow the Gaussian distribution regardless of the distribution of the input values (on the mean). This is of course not desired while simulating series that do not follow the Normal distribution. For this purpose, a further step is taken to get the original distribution back. It is detailed shortly in a coming subsection.

It is important to mention how the randomized phase spectra affect the resulting surrogate series' properties. For any given frequency, increasing or decreasing the phase by 180° results in direction reversal of the wave. For example, if a wave starts with a positive value, after adding 180° to it will result in it starting from the same value but with a negative sign. If all the waves are shifted by 180° then, the resulting time series flips its direction in time. For discharge, it would mean that the rising limbs would be slower than the recession limbs. Some drawbacks of using PR will be presented later.

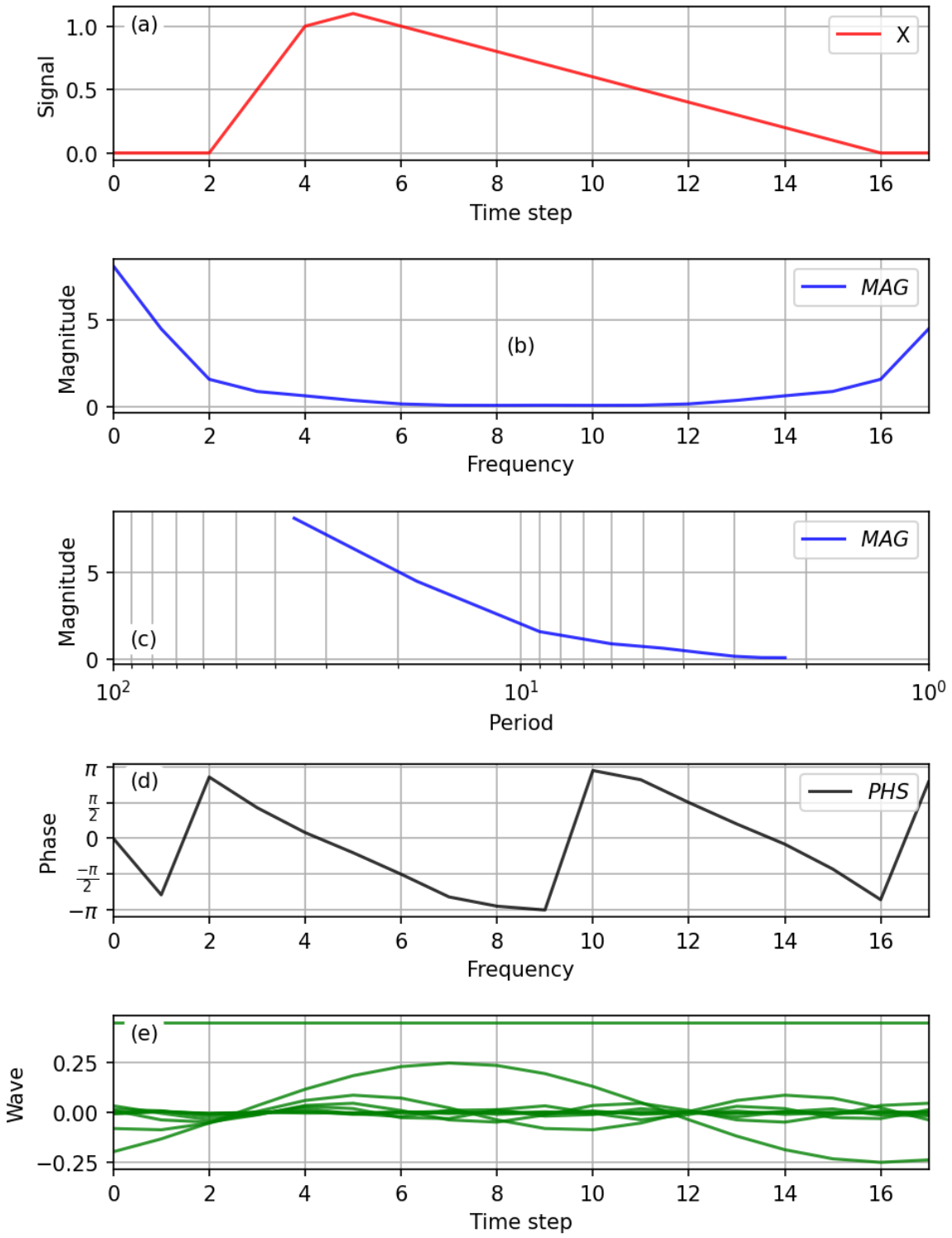


Figure 5.1: Magnitude (b and c) and phase (d) spectra of a hypothetical time series (a). (c) shows a different way of visualizing the magnitude spectrum by using periods of waves which are easier to understand as compared to frequencies. Due to the magnitude spectrum symmetry, only the first half is shown in (c). It is important to mention that the first value of the magnitude spectrum represents the sum of the values of X. (e) shows waves with various frequencies that sum up to give the original time series X.

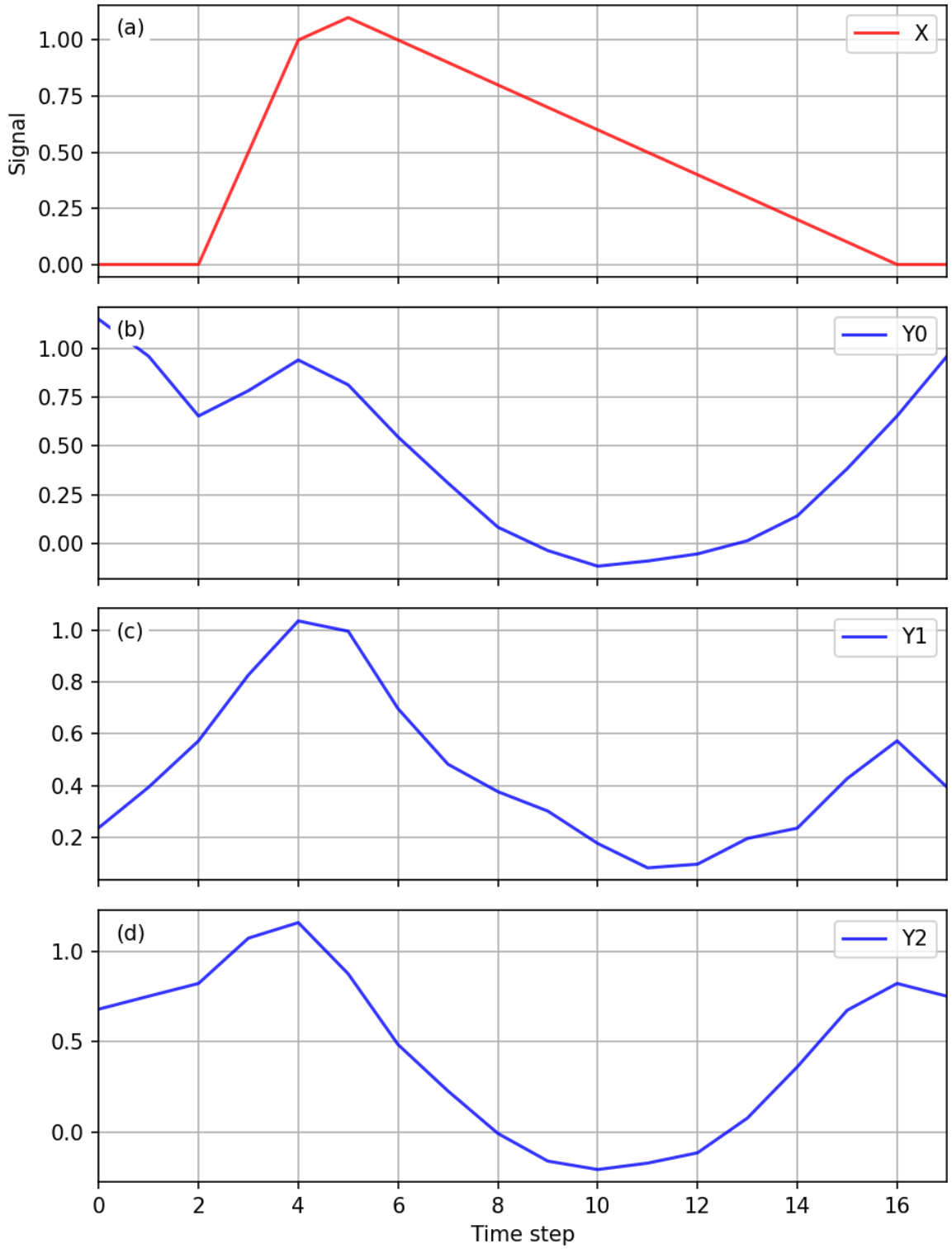


Figure 5.2: A reference series (a) and its three corresponding surrogate series (b, c, d) generated using Phase Randomization.

The algorithm

An algorithm describing how a multivariate series X_M with M variables and N time steps (an N $\times$ M matrix) can be Phase Randomized such that the covariance matrix of the variables remains undisturbed. This algorithm requires N to be even.

1. Compute the Fourier coefficients at the respective frequencies for each variable in X_M using Eq. 6.1.
2. Using Eq. 5.3 obtain the magnitude spectra.
3. Using Eq. 5.4 obtain the phase spectra.
4. Draw $\frac{N}{2} - 1$ random phases between $+\pi$ and $-\pi$.
5. Create another set of phases by adding $+\pi$ to all the phases from the previous step and then flipping them in order.
6. Add the phases from step 4 to all the phase spectra from step 3 at indices 1 to $\frac{N}{2} - 1$.
7. Add the phases from step 5 to all the phase spectra from step 3 at indices $\frac{N}{2} + 1$ to $N - 1$. The phases at indices 0 and $\frac{N}{2}$ remain unchanged for all spectra.
8. Compute the new Fourier coefficients by using the modified phase spectra from the previous step and the magnitude spectra from step 2. This can be done by taking the product of the magnitude and the cosine of the respective phase at each frequency as the real part of the coefficient. For the imaginary part take the product of the magnitude and the sine of the phase.
9. Using the new Fourier coefficients for each series, get the surrogate series through Eq. 5.2.

Drawbacks of using Phase Randomization as a time series generator

It was shown how simple it is to generate series that conserve the auto- and cross-correlations. However, there are quite a few drawbacks to using PR.

Firstly, series that have the same length as that of the reference series can be generated only. Simple concatenation of series in to a longer one results in the destruction of the continuity of dependence in time, because the ending of one series and the beginning of the next have no relation to each other for observed series necessarily.

Secondly, it only keeps the mean and overall correlation in time and space. Other statistical properties such a skewness and kurtosis are not kept necessarily.

Another disadvantage is that the surrogate series have values that follow the Gaussian distribution. This becomes a serious problem when the variables in question have different distributions. For example, there are no negative discharges, precipitation and evapotranspiration. Precipitation has a mixed distribution where zero precipitation may dominate a good portion of the overall distribution. At least two ways exist to remedy this problem. Using the first method, all input values are transformed to Gaussian and then these

are Phase Randomized. After PR, the original values are reordered such that they follow the same order in time as that of the surrogate series. The later method, involves fitting proper distributions to all inputs. Transforming the inputs to Gaussian (same as before). Afterwards, the back transform is used to come back to the range of the reference values (QQ-transforming or rank-order matching). This has the advantage of having values in the simulations that may go beyond the range of the reference. Both these methods are variations of the Amplitude adjusted Fourier transform (AAFT) algorithm described in Theiler et al. (1992). However, the former method is opt for in this thesis because the main concern here is simultaneous occurrences of extremes rather than their magnitudes. Unfortunately, doing so (using either method) leads to another problem. Namely, PR keeps the auto- and cross-correlations in the Gaussian space. The resulting series are almost guaranteed to lose auto- and cross-ρ_p.

Yet another, and perhaps the biggest, limitation is that of the loss of important copula properties namely the Spearman's correlation, order- and directional-asymmetries and entropy. Upon Phase Randomizing, information about timing of the waves carried by the phase spectra are lost. For example, the waves with a period of one year and those that correspond to seasons, always happen at, almost, fixed times in any year for observed data (at least for environmental variables). The timing of the beginning of a hydrological year and seasons are lost when the phase spectra are randomized.

Fig. 5.3 and 5.4 show the example of a daily discharge series. It has some special properties that the surrogate series should also preserve. Properties such as the difference of dependence in time of the low flows and the high flows i.e., order asymmetry (AO) and the difference in the rising and recession limbs of various events i.e., directional asymmetry (AD). Notice how AO becomes significantly different at larger lead steps while the AD is significantly different at the first few lead steps only. The simulated series lose both these properties i.e., the opposite corners of the empirical copulas show very similar densities. Furthermore, the copula entropies (H) also increase i.e., loss of dependence that was not captured by auto-correlations.

By observing the empirical copulas and the properties that the new series have, it can be concluded that they all correspond to that of the multivariate Gaussian copula i.e., they are symmetric. It could be said that PR is in fact a multivariate Gaussian distribution sample generator in disguise, where the means, the correlation matrix (in case of a multivariate scenario) and the auto-correlation function correspond to that of the given reference.

An important property of weather-related variables are their annual and seasonal cycles along with daily, for the case of hourly data, that are lost as a result of randomizing all phases. The solution to this was formally presented in Galka (2000); Nakamura et al. (2006), where phases of important cycles were kept undisturbed. This results in simulations that show consistent patterns of time series in time. This approach is also used for all time series simulated in this thesis.

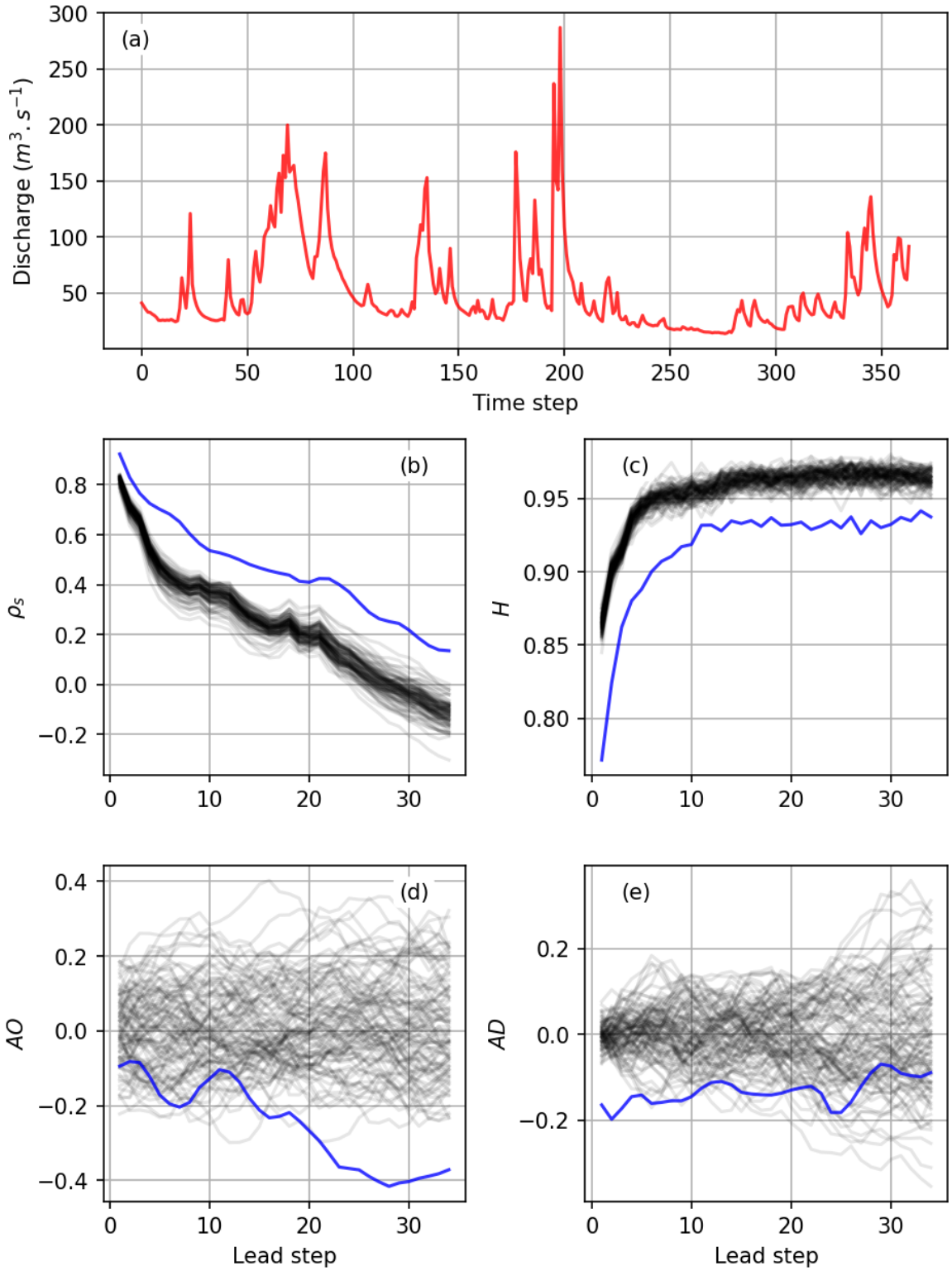


Figure 5.3: Comparison of copula properties (b, c, d, e, f) of a reference discharge series (a) against those obtained using Phase Randomization. The color blue represents the reference while black represents properties of Phase Randomized series. The actual surrogate series are not shown here. Pearson correlations are also not compared as they are equal for all series.

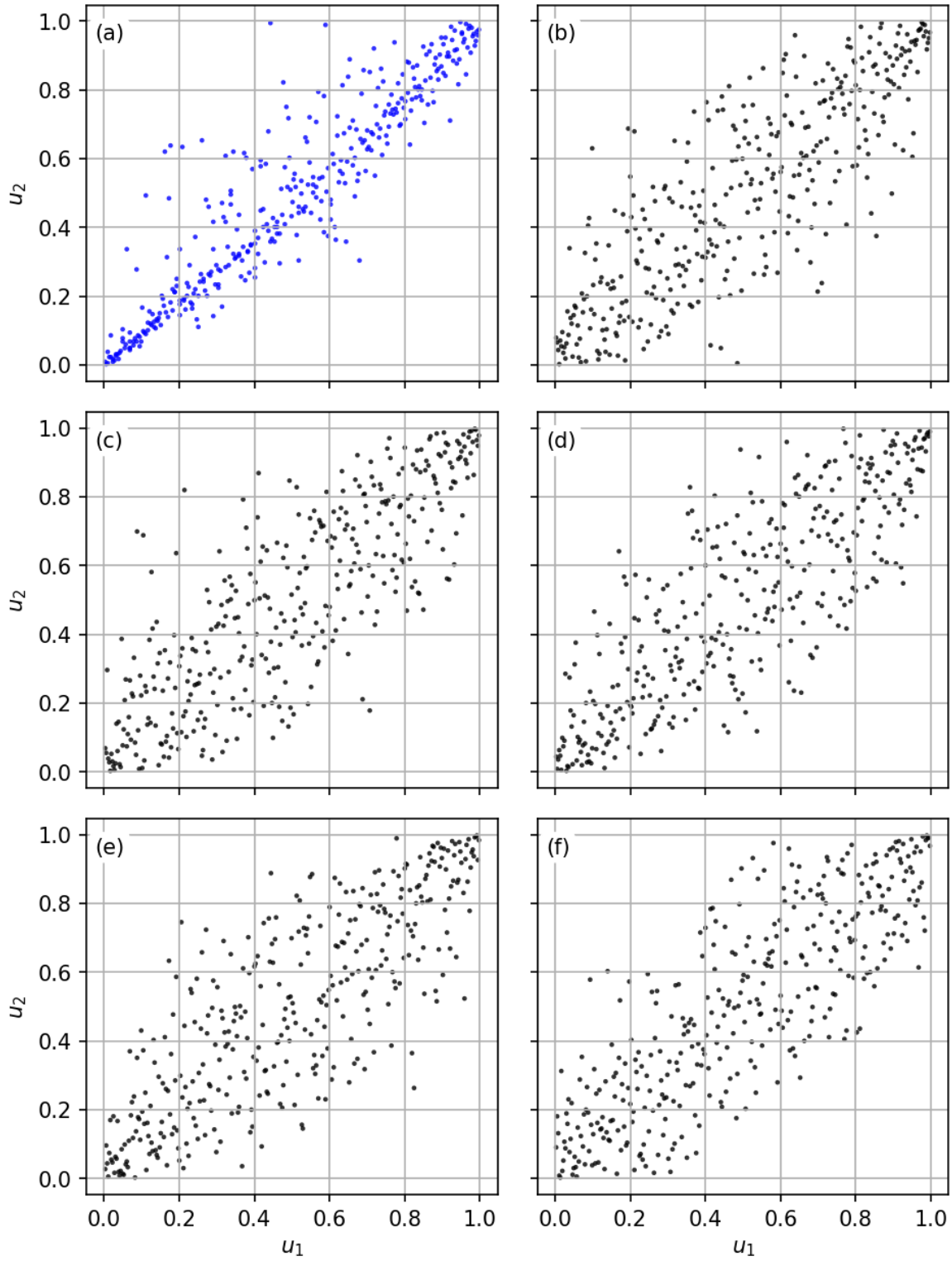


Figure 5.4: Empirical copulas of reference (a) and surrogate (b, c, d, e, f) series. u_2 leads u_1 by 1 step.

5.3 Amplitude adjusted Fourier transform

Amplitude adjusted Fourier transform (AAFT) aims to get the original distribution back by using a simple transformation from the Gaussian values to the original values (Theiler et al., 1992). This method is called the QQ-transform (Quantile-Quantile transform) or Rank-order matching. This variant of PR adds one additional step. One method to achieve this is as follows:

1. Sort the values of the original time series.
2. Find the ranks of the values in the surrogate time series.
3. Use these ranks as indices in the sorted time series to get them in to new order that corresponds to that of the surrogate but has now the values whose distribution is exactly that of the original time series.

It should be noted that AAFT is the default form of Phase Randomization that is used through out this thesis because the distributions must match those of the reference. For example, a Phase Randomized series will have values that follow a Gaussian distribution i.e., negative discharges, which can never be the case. Same applies to precipitation and potential evapotranspiration. For this reason whenever time series are mentioned that were created using PR, what they actually refer to is that the AAFT algorithm was used.

Drawbacks of using AAFT

AAFT solves the problem of the incorrect data distribution but now creates another one namely that the auto-Pearson's correlation (ρ_p) is disturbed because the magnitude spectrum was only equal to that of the original for the surrogate Gaussian values. This also has a consequence for the distribution of the first-order differences. All the properties in the copula domain remain as they were before because the QQ-transformation is monotonic. All the other drawbacks due to PR still remain.

5.4 Iterative amplitude adjusted Fourier transform

To address the issue of the lost auto-ρ_p function, yet another addition to the AAFT algorithm, Iterative amplitude adjusted Fourier transform (IAAFT), was proposed by Schreiber and Schmitz (1996). This variant involves replacing the resulting *incorrect* magnitude spectrum from AAFT by that of original series iteratively. The algorithm is as follows:

1. Compute the magnitude spectra (MAG_X) of the original time series X.
2. Let Y be the surrogate. Randomly shuffle values for each series (this approach was recommended by Venema et al. (2006)) in Y. This results in the time series having a nearly flat power (i.e., all values in the magnitude spectrum become very similar). This is the series that is iteratively improved to become the final surrogate series.
3. Compute PHS_Y.

4. Compute the surrogate Y by using MAG_X and PHS_Y in Eq. 5.2.
5. Apply the QQ-transform to get time series Y with the correct distributions.
6. Skip this step if it is the first iteration. Otherwise, compute the Spearman's correlation (ρ_s) of the new time series with that of the the previous iteration. If it is above a threshold, say 0.999, then stop the algorithm if not then go to step 3 and repeat till the threshold is reached.

The final series resulting from the above algorithm should have a power spectrum very similar to that of the original series along with the distribution that is exactly that of the original. This is guaranteed by the construction of the algorithm as was demonstrated by its creators.

Drawbacks of using IAAFT

IAAFT addresses a major short coming of the previous variants by iteratively finding a time series that has the correct properties. These new time series may have properties in the copula domain that are clearly better than those of the AAFT. However, as only the magnitude spectrum of the original time series was used, there is no guarantee if other required properties are kept. It does not allow any control over what the properties of the resulting time series should be except that the auto-ρ_p function is kept.

5.5 Multivariate iterative amplitude adjusted Fourier transform

So far, time series could be generated that had the correct auto-ρ_p function. The next problem that needs to be addressed is that of the pairwise cross-ρ_p. Keylock (2012) proposed a method, Multivariate iterative amplitude adjusted Fourier transform (MIAAFT), to generate multivariate time series by extending the IAAFT algorithm with one more step. The idea is simple, elegant and efficient. Previously, it was mentioned that the cross-correlation of two sites is dependent on the phases of the Fourier coefficients. The closer the phases to each other the stronger the cross-correlation. It is this fact that is used to generate time series that have the additional property of having correct cross-correlations. The algorithm is as follows:

1. Randomly select an index m from the M number of time series in the input X_M.
2. Find the phase differences (ΔPHS_m) for each series by subtracting the phase spectrum of series X_m ($PHS_X m$) from each phase spectrum in PHS_X. This should yield M number of phase spectra in which all the values in the column m are zeros.
3. Similar to IAAFT, start by having a time series Y in which all the values are randomly shuffled.
4. Compute the phase spectrum of the series Y_m ($PHS_Y m$).
5. Compute the new phase spectrum (PHS_Y) by adding the respective difference spectra from ΔPHS_m to $PHS_Y m$.

6. Compute the surrogate Y by using MAG_X and PHS_Y in Eq. 5.2.
7. Rank-order match Y.
8. Skip this step if it is the first iteration. Otherwise, compute ρ_s of the new time series with that of the the previous iteration. If it is above a threshold, say 0.999, then stop the algorithm if not then go to step 4 and repeat till the threshold is reached.

Drawbacks of using MIAAFT

Using this method does solve the problem of cross-ρ_p with one extra step. However, during testing this algorithm it was observed that quality of the outputs was not uniform for the all the simulated time series. The problem is due to the selection of the series, Y_m, as the reference for phase differences. Whichever station was selected, had the auto-ρ_p function preserved better than those of the other series. Apart from that, the problem of dependence loss in the copula domain still remains.

5.6 Modified multivariate iterative amplitude adjusted Fourier transform

In order to deal with the over-fit for one series in MIAAFT, only a small adjustment is required to rectify the bias. This adjustment merely involves cycling through series while the MIAAFT algorithm loops. The algorithm, Modified multivariate iterative amplitude adjusted Fourier transform (MMIAAFT), is as follows:

1. Start by having a randomly shuffled time series Y.
2. Set the series counter (m) to zero.
3. Compute ΔPHS_m.
4. Compute $PHS_Y m$.
5. Compute PHS_Y by adding $PHS_Y m$ to ΔPHS_m.
6. Compute the surrogate Y by using MAG_X and PHS_Y.
7. Rank-order match Y.
8. Increment m by 1. If it equals M then reset it to zero.
9. Skip this step if it is the first iteration. Otherwise, compute ρ_s of the new time series with that of the the previous iteration. If it is above a threshold, say 0.999, then stop the algorithm if not then go to step 3 and repeat till the threshold is reached or if ρ_s begins to oscillate (increasing and decreasing periodically) then stop the algorithm.

Drawbacks of using MMIAAFT

It is a slightly better version of MIAAFT. A small drawback was observed though. Due to the constant change of m for which the phase differences are taken, the algorithm never converged but rather oscillated after reaching a certain high correlation. Another way of circumventing the non-convergence could be to use a dummy series that has no correlation to the other series and using that as the pivot for the phase differences in the MIAAFT algorithm. Nevertheless, results showed that even if the algorithm never converges, it does produce spectra that are very similar to those of the reference and are less biased than MIAAFT. Problems in the copula domain still remain.

5.7 Simulated Annealing

Before introducing a very powerful variant of Phase Randomization, it is important to discuss the optimization scheme of Simulated Annealing which is at the heart of Phase Annealing. Suppose the following problem. An observed time series of a variable X with N time steps is provided for which surrogate series have to be generated. To simulate the process generating X a model G with a parameter vector β having a length of M and K model input variables Z_K also in a series form is chosen. Let the simulated output time series for any β using G be Y. What remains to be solved is β such that when used in the model, the overall difference between X and Y is below a specified threshold.

Finding such a β may or may not be trivial. It could be that no solution exists as the model is too simple or just an incorrect representation of the process generating X; or many such vectors exist, the so called Equifinality; or there exists a unique solution indeed. Given the complexity of natural systems, simpler models are chosen to represent them. Generally, it is not trivial to find the optimal β by analytical methods. Hence, *optimization schemes* are used. It involves finding out the best β iteratively. This is traditionally done by starting from a randomly drawn β from the parameter space and moving it slowly in an M-dimensional space while evaluating the objective function till a desired solution is achieved. The *movement in parameter space* in search of the optimum is what sets apart all optimization schemes. Some move based on slope, some based on covariance matrices, some on mixing different βs and some on random walks. They optimization schemes have two main types: *local* and *global*. Local refers to schemes that start from a point, keep evaluating the differences till they become smaller and smaller and stop once it starts to become large; whereas, global refers to schemes that tend to evaluate the differences all over the parameter vector space and stop when they have no more space to evaluate, in theory that is, all while retaining the best β. All optimization schemes have one thing in common and that is the objective function. Its value represents the difference in reference and simulated variables e.g., sum of squared difference between a reference and the model output, Nash-Sutcliffe Efficiency (NSE) etc. Furthermore, it is possible that an objective function value might need to be maximized or minimized depending on the problem. Traditionally, optimization schemes aim at minimization.

Among the thousands of optimizations schemes available, *Simulated Annealing* (SA) is used

in this thesis to find the best surrogates. It is a global optimization scheme that aims to find the minimum of any given objective function by applying random changes to β using a prescribed *cooling schedule*. This technique is a mathematical version of the method used to slowly cool hot metals to make them soft and release any stress that has built up inside the metal lattice i.e., Annealing. The cooling has to be slow enough such that the atoms of the metal can align themselves in a target lattice. A faster cooling rate will result in atoms locking in positions that are not optimal, resulting in a brittle structure. The main idea i.e., the connection between statistical mechanics and multivariate optimization was first presented by Kirkpatrick et al. (1983) while the basic algorithm of SA is given in detail by Černý (1985). Their ideas were based on those of Metropolis et al. (1953), who formulated an algorithm that described transition of a system of atoms/molecules from a high to very low energy (entropy) state.

The algorithm

Following steps are involved in finding the solution of a given problem using the SA algorithm.

1. Define the objective function that results in a value O. The aim is to evaluate/minimize the sum of the squared differences between some reference (X), that has to be matched, and simulation (Y).
2. Define the maximum number of iterations (Q).
3. Define the minimum absolute difference value between successive iterations (Ψ). The algorithm stops if the absolute difference drops below this for some given number of iterations continuously.
4. Set the initial annealing temperature (T_0). It has to be greater than zero and less than Infinity.
5. Define a *cooling ratio* (C) which is applied after every J iterations. C is a value greater than zero and less than one. This is called the *cooling schedule*.
6. Draw a random β_0 and compute O_0 by running G using β_0 and using X with the resulting Y in the objective function.
7. Set the iteration counter (i) to one.
8. Draw a new β_i and compute O_i same as that in step 6.
   a) Calculate the absolute difference (ψ_i) between O_0 and O_i.
   b) If O_i is smaller than O_0 then replace β_0 with β_i and O_0 with O_i and go to step 9.
   c) Otherwise,
      i. Compute the Boltzmann's Probability (P_b),

$$P_b = \exp\left(\frac{O_0 - O_i}{T_i}\right) \tag{5.8}$$

ii. Draw a random number (P_r) between zero and one.

iii. If P_r is less than P_b then β_0 is replaced by β_i and O_0 by O_i.

iv. Else, reject β_i and go to step 9.

9. Increment i by one.
10. If Q iterations are completed then go to step 15.
11. If ψ_i is less than Ψ then go to step 15.
12. If J iterations have passed then update the temperature,

$$T_i = C \times T_{i-1} \tag{5.9}$$

13. If $T_i \approx$ zero then go to step 15.
14. Go to step 8.
15. β_0 is the optimal solution and O_0 is its corresponding objective function value.

Two main aspects of the algorithm stand out i.e., the cooling schedule and the Boltzmann's probability. Apart from that still some questions remain e.g., how to sample a new β_i. These are explained next.

Some remarks

The initial temperature (T_0), the cooling schedule and the Boltzmann's probability (P_b) are tightly coupled in SA. To begin with, a T_0 is required such that when used in Eq. 5.8 results in a rejection rate that is 100% exactly (or slightly below for practical purposes, say 80 to 90%). This would mean that the knowledge that every rejected solution is accepted when the algorithm begins and by reducing the temperature after J iterations (cooling) exists beforehand. In practice, such a situation is not possible. Paradoxically, finding a suitable T_0 would require another optimization scheme. A simple method to get an acceptable value is to evaluate the acceptance rates of randomly sampled temperatures that involve running the optimization scheme for a few hundreds of iterations, without updating the temperature. This can be done simply by increasing the temperature from some known small value at which the acceptance rate is very close to zero (say below 20%) and stopping at a temperature that yields an acceptance rate between 70 to 80%. Once a suitable temperature is reached, SA can begin. The second issue is that of the cooling schedule. If the cooling ratio (C) is too small (used in Eq. 5.9), then there is a high chance of obtaining a local optimum. C very close to one might result in the optimization taking too long to converge to a global optimum. Unfortunately, there is no straightforward way of knowing the best C. This too has to be calculated iteratively once T_0 is known. One way to do so is to run the algorithm in parallel many times, starting from a small C and then evaluating if all runs converged to similar objective function values (O). If yes, then this is the optimum C otherwise keep increasing it till all runs result in Os that are within an acceptable tolerance of each other. C is also linked to the length of the vector β. The larger the length, the closer C should be to

one. The third issue, and a unique feature, of SA is P_b. Due to the *exp* and its negative input it is always between zero and one. This is an unusual feature of SA because it allows for acceptance of βs that result in an O that is worse than the previous one. This is done in the hopes that even worse performing βs can lead to better solutions subsequently. It changes dynamically as the temperature is reduced. For example, initially when temperature is high P_b is also closer to one due to the larger temperature in the denominator. Thereby, resulting in relatively higher chances of acceptance of worse solutions. As the algorithm progresses, P_b gets closer to zero, thereby resulting in smaller chances of acceptance of worse solutions. Such an approach improves the chances of not getting stuck in a local optimum. Finally, the issue of sampling of the vector β for subsequent iterations needs to be addressed. Nominally, it can be sampled from anywhere in the M-orthotope (a hyper dimensional rectangle) but that is very likely to produce strong fluctuations in the objective function resulting in high rejection rates and slow convergence. A more reasonable approach is to dynamically reduce the sampling space around β as the algorithm progresses. For example, a variable can be defined that is reduced proportionally to the relative number of iterations or the acceptance rate. This variable may be updated whenever the temperature is reduced. Hence, in the beginning when the temperatures are high along with the acceptance rate, β can be sampled from anywhere in the M-orthotope thus sampling almost uniformly and then as the algorithm progresses, based on the acceptance rate, the distance of the next β relative to the last is reduced. This ensures that the next sampled vector will result in an objective function value that is not very different than the last. Hence, resulting in a smoother and a more efficient convergence towards the global optimum.

The aspect of convergence of any optimization scheme is of the main importance. For a global optimizer, the focus is more on the efficient *sifting* of the parameter space to find a point that produces the least objective function value. This, of course, can not be always guaranteed for all schemes. But for SA it is guaranteed provided the condition that the cooling schedule is set to change slowly enough. The main reason that this can be guaranteed is due to P_b, which allows the acceptance of solutions that are worse than the existing ones. Meaning that optimization does not get stuck in a local minima necessarily. This was mentioned in Metropolis et al. (1953). At the end, the question of the cooling schedule still remains. If it is too slow, it is likely that the same points (or neighborhoods) are visited repeatedly. If the cooling schedule is too fast then the chances of obtaining a local optimum increase. The only remedy for was mentioned before, i.e., the optimum cooling schedule has to be found empirically.

5.8 Phase Annealing

Phase Annealing (PA) is a technique that aims to generate series/fields similar in properties with respect to a given reference by making changes to the phase spectrum (or spectra) using Simulated Annealing (SA). It was recently presented in Hörning and Bárdossy (2018) to generate spatial random fields that must have set of prescribed properties. In this thesis, the algorithm is used to generate time series for a single variable at multiple locations or multiple variables at the same location or multiple variables at multiple locations or a univariate

series at a single location, while retaining a set of properties in the simulations similar to those of a reference.

To make things more understandable, suppose the following two problems: A time series of a univariate reference variable X is provided. It has properties such as auto-correlation (ρ_p and ρ_s) functions, copula auto-asymmetries (AO and AD) and copula auto-entropies (H) in time. It is required to have M series such that they all have properties very close to those of the reference and that all of them have, possibly, very little correlation in time with the reference and with each other. This would be the *auto*-case. Similarly, there can be two reference series X and Y, where the properties for the *cross*-case are also required. One approach can be the use of theoretical copulas that can be fitted to various arrangements of the variables. Here, an approach is used that allows for preservation of properties non-parametrically. This method is used because using theoretical copulas to simulated series involves implicit assumptions on the dependence structure rather then a general solution that frees the users from making any assumptions and produces series that preserve the properties of the reference. Another problem with fitting theoretical copulas is that only few exist that are truly multivariate such as the Gaussian copula which lacks the asymmetric dependence properties that is very visible when analyzing observed time series. Vine copulas suffer from the same drawbacks.

Before proceeding further, the curse of dimensionality is discussed. Suppose two points in three dimensions e.g., [0, 0, 0] and [1, 1, 1] as a sample. Considering only the first dimension of both the points the distance between them is 1.0 units. For the first two dimensions it is 1.414 units and for all of them it is 1.732 units. In other words, the density of points in space reduces as dimensions increase while keeping the number of points constant. To obtain the same density of a mean distance of 1.0 units more points have to be added. This can be done by dividing the M-orthotope into smaller equally sized M-orthotopes and then calculating the mean density of points in each of them. The discretization should be based on the one-dimensional case's distances. Now, more points are added till the density equals to that of the reference case. Unfortunately, this is very unlikely to happen in a real world scenario because the length of the time series are restricted and getting a longer series is not possible.

As shown in Sec. 5.2, Phase Randomization (PR) can be used to generate new series easily. The problem that remains to be solved is that many important properties are not reproduced which is directly linked to the random phase spectra that are drawn. As a phase spectrum is a very high dimensional vector, the probability that a randomly drawn spectrum also yields the desired properties is very small. Hence, an optimization scheme is required to find such a spectrum. Using the Simulated Annealing (SA) algorithm to do so is what is referred to as Phase Annealing (PA).

The algorithm

Following steps are involved in finding suitable phase spectra for a multivariate series (X) with M variables and N time steps. Here, the SA algorithm, described in Sec. 5.7, adapted to generate a multivariate time series is shown.

1. Compute the Fourier coefficients (FT) of X, its corresponding magnitude (MAG) and phase (PHS) spectra using the formulas shown in Sec. 5.1.
2. Compute all the desired properties that have to be kept such as correlations, asymmetries and entropies using formulas from Chapter 3. These will be used as a reference in the objective function and these are the ones that the simulated series should have at the end.
3. Define the objective function(s). The differences between the reference and simulated series' properties are evaluated. These could be for the auto- and the cross-cases. For example, the sum of the squared differences of the auto-asymmetries of the reference at desired lag/lead steps and the simulations at the same lag/lead steps. Let the value of the objective function be O.
4. Define the maximum number of iterations (Q).
5. Define the minimum absolute O difference between successive iterations (Ψ).
6. Set the initial annealing temperature (T_0).
7. Define a cooling ratio (C) which is applied after every J iterations.
8. Draw a random phase spectrum similar in length to that of PHS. With all the phases in between +π and -π. Add these to PHS at the corresponding frequencies to produce new phase spectra (PHS_0). For example, at frequency k the same phase is added to all the phases in PHS. Use Eq. 5.2 to obtain a new X_0. The procedure is similar to that used in sampling a new phase spectrum in Phase Randomization. Use the new series to compute O_0.
9. Set the iteration counter (i) to one.
10. Draw a random phase (ϕ_i) and frequency index (k_i). Compute the new phase spectrum (PHS_i) by adding ϕ_i to PHS_0 at the frequency index k_i. Use PHS_i to get X_i and then compute O_i, same as that in step 8.
    a) Calculate the absolute difference (ψ_i) between O_0 and O_i. If ψ_i is less than Ψ then go to step 17.
    b) If O_i is smaller than O_0 then go to step 11.
    c) Otherwise,
        i. Compute the Boltzmann's Probability (P_b),

$$P_b = \exp\left(\frac{O_0 - O_i}{T_i}\right)$$

        ii. Draw a random number (P_r) between zero and one.
        iii. If P_r is less than P_b then move to step 11.
        iv. Else revert the changes applied in this iteration at step 10 and move to step 12.

11. Replace O_0 with O_i, PHS_0 with PHS_i and X_0 with X_i.
12. Increment i by one.
13. If Q iterations are completed then go to step 17.
14. If J iterations have passed then update the temperature,

$$T_i = C \times T_{i-1}$$

15. If $T_i \approx$ zero then go to step 17.
16. Go to step 10.
17. PHS_0 are the optimal phase spectra and O_0 is their corresponding objective function value and X_0 is the multivariate *Phase Annealed* time series.

Some remarks

Caveats that were valid for the original SA algorithm apply here as well. Furthermore, in PA the phase spectrum is not fully analogous to that of a model parameter vector β.

At least two variables are dealt with simultaneously, namely the phase (ϕ) and the frequency index (k). Choosing these does affect how efficiently the high-dimensional phase space is traversed while looking for a suitable solution. Suppose that PHS has a length of one thousand. Selecting one phase randomly from it and changing its value will have very little effect, in general, on the resulting properties. The same strategy that was recommended for β can be used here as well i.e., dynamically reducing the change applied to phases and the number of frequencies selected for modification. The implementation can be as follows: At the start when temperatures and acceptance rates are high, generate more frequency indices. Similarly, generate phase differences that are also larger in magnitude. As the optimization progresses and the acceptance rates drop, the absolute value of the new phase differences along with number of phases to modify is reduced. This allows for movement through the phase space more efficiently in the beginning to look for a neighborhood that might hold phases such that the resulting properties are more suitable rather than moving very slowly by changing a single phase at every iteration.

Another problem that might arise is that of the relative magnitude of the objective functions. For example, the order asymmetries (AO) of the simulated series being very similar to that of the reference while the directional asymmetries (AD) being wildly different. This results in AD dominating the objective function value with the final AO being the best. For such a case, weights need to be assigned to each property in the objective function. The weights can be obtained by comparing the relative magnitudes of differences in the objective function by using randomized phase spectra for a few hundreds of iterations. The measure with a relatively small difference gets a larger weight and vice versa for the ones with larger difference magnitudes. The same may be done while comparing the objective function values of time series of multiple variables, where some might dominate the rest.

5.9 A better MMIAAFT: IAAFT-PSC1

So far, none of the IAAFT variants that were discussed here conserve properties in the rank space. One such approach is presented here. The idea is as follows: It is clear that MIAAFT conserves auto- and cross-ρ_p functions. Which depends on the distribution of the observed values. If the distribution changes, so does ρ_p. This was evident with the use of PR. This fact can be exploited. Suppose that instead of using the marginals of a time series, their grades, $F_X(x)$, are used and then the MMIAAFT algorithm is used. The generated series will have the correct ρ_p but as the grades were used as inputs, the new time series with the correct ρ_s are simulated. However, after QQ-transformation to get the time series with the correct distribution, it can be observed that the ρ_p has in fact changed. The crux of the matter is that the type of input influences the properties of the outcome. The next question that arises is: Do any transformed series exist that results in keeping both ρ_p and ρ_s when used as input for MMIAAFT? If using the marginals results in correct ρ_p and using the grades results in correct ρ_s then it could be that a series that is a mixture of the two produces results that have both the properties. It may be that even other copula properties such as asymmetries and entropy are also kept. This is what is proposed here. A variant of MMIAAFT that aims to do both. In essence, it uses a mixture of the observed values and their grades as input rather than any single one of them. The trick lies in how the two are mixed. On the account of the Pearson's and Spearman's correlation preservation (PSC), this algorithm is referred to as IAAFT-PSC1. The steps are as follows:

1. Compute the grades (U) of the input series X for each column m in M columns. The grades are non-exceedence probability values obtained from the empirical distribution function of a given variable.
2. For M random αs between 0 and 1 (one for each column) do the following:
   a) Start by having a randomly shuffled series Y.
   b) Set m to zero.
   c) Compute grades (V) of Y.
   d) Compute ΔPHS_m for Y and V .
   e) Compute PHS_Ym and PHS_Vm.
   f) Compute PHS_Y by adding PHS_Ym and PHS_V by adding PHS_Vm to their respective ΔPHS_ms.
   g) Compute the surrogate Y by using MAG_X and PHS_Y. Similarly, compute the surrogate V by using MAG_U and PHS_V.
   h) Rank-order match Y. Similarly, Rank-order match V.
   i) Normalize Y and V by subtracting theirs respective means from them and then dividing them by their respective standard deviations. This normalizes both to a range of 0 to 1.
   j) Set the new series Y_m = $\alpha_m Y_m$ + (1 - α_m)V_m for all columns m in M.

k) Rank-order match Y.

l) Increment m by 1. If it equals M then reset it to zero.

m) Skip this step if it is the first iteration. Otherwise, compute ρ_s of the new time series with that of the the previous iteration. If it is above a threshold, say 0.999, then stop the algorithm if not then go to step 2c and repeat till the threshold is reached or if ρ_s begins to oscillate then stop the algorithm.

n) Store optimized Y and its grades V as separate variables elsewhere.

3. After finding series with many αs, compute the properties such as auto- and cross-ρ_p and ρ_s and select the series that give the least error as the final solution.

For the single-site case, steps 2d to 2f can be ignored while taking the surrogate phase spectrum for the single column similar to what is done in IAAFT. Similarly, many time series can be simulated simultaneously while keeping no cross-correlation. This setting is also relevant to IAAFT-PSC2, which will be discussed next. The algorithm is guaranteed to converge as the α_ms are optimized using SA. Also, for each iteration the algorithm is basically IAAFT which is also guaranteed to converge. This was also evidenced by the tests (not shown) performed while developing it.

Drawbacks of using IAAFT-PSC1

IAAFT-PSC1 does address the problem of the ρ_s loss albeit at the cost of much more computational time. The algorithm is effectively an optimization scheme where the αs are the model parameters that have to be optimized. Simulated Annealing (Kirkpatrick et al., 1983; Černý, 1985) was used to find the optimum αs. An important point to note is that the final model parameters depend on the randomized sequence of Y that the algorithm was started with. This algorithm does not solve all the problems though. Results showed that order asymmetry was kept while the directional asymmetry was not. This means that it is suitable for cases where the directional asymmetry is negligible. For keeping the directional asymmetry, something much more sophisticated is required.

5.10 IAAFT-PSC2

While analyzing the results of IAAFT-PSC1, it was observed that as the ρ_s is preserved to a certain degree, it had a problem in many cases. The problem being that the auto-ρ_s function of the series was above than that of the reference. It stayed well below and approached almost that of the reference. Furthermore, it was observed that for simulations where the auto-ρ_s function was well preserved, the auto-ρ_p function was shifted higher than that of the reference. For any α, there was always an offset between the auto-ρ_p and ρ_s. So far, the phase spectra of the surrogate series were not disturbed during the optimization. Next, the phase spectra of the marginals and grades were swapped with each other before the backward Fourier transform was applied. This, seemingly, illogical step produced series that resulted in no difference between the auto-ρ_p and ρ_s functions on average. Furthermore, it was also

observed that the order asymmetry was much better conserved. IAAFT-PSC2 involves twice the calculations of IAAFT-PSC1, as it does the phase swapping for the case where no cross-correlations are kept. When the phase swapping was tried for the case with cross-correlation conservation, then the resulting series were of lower quality. In essence, IAAFT-PSC2 is IAAFT-PSC1 run twice for the original setting and once with a setting where multiple station are taken but no cross-correlation is kept and the phase spectra of the marginals and grades are swapped. The algorithm is as follows:

1. Compute the grades (U) of the input series X for each column m in M columns.
2. For M random αs between 0 and 1 (one for each column) do the following:
   a) Start by having a randomly shuffled series Y.
   b) Set m to zero.
   c) Run the IAAFT-PSC1 from the steps 2c to 2j in IAAFT-PSC1 in a cross-correlation preservation setting. Let the resulting series be A.
   d) Run the IAAFT-PSC1 from the steps 2c to 2j in IAAFT-PSC1 in a non-cross-correlation preservation setting. Let the resulting series be B.
   e) Compute $Y = A + B$.
   f) Rank-order match Y.
   g) Increment m by 1. If it equals M then reset it to zero.
   h) Skip this step if it is the first iteration. Otherwise, compute ρ_s of the new time series with that of the the previous iteration. If it is above a threshold, say 0.999, then stop the algorithm if not then go to step 2c and repeat till the threshold is reached or if ρ_s begins to oscillate then stop the algorithm.
   i) Store optimized Y and its grades V as separate variables elsewhere.
3. After finding series with many αs, compute the properties such as auto- and cross-ρ_p and ρ_s and select the series that give the least error as the final solution.

For the single-site case, step 2c can be ignored. Similar to IAAFT-PSC1, IAAFT-PSC2 also converges to a point where no significant changes in the temporal order of the variables cannot be observed. But given that the phase spectra are swapped repeatedly, it may take many more iterations as, say, compared to IAAFT.

Drawbacks of using IAAFT-PSC2

IAAFT-PSC2 addresses the correlation bias issue of IAAFT-PSC1. All of this is at additional computational cost. There was one major drawback observed for cases where there was significant directional asymmetry in time i.e., discharge time series. As both variants of the IAAFT-PSC cannot preserve the directional asymmetry, it was observed that for simulations the resulting time series had low-highs that correspond to the observed cases but an equal

number of cases of high-lows and that also in close succession in time. Meaning that the series had an oscillating behavior, which is unnatural. This drawback could only be observed when the simulated time series were inspected directly. By looking only at the ρ_p, ρ_s and both asymmetries such behavior cannot be detected. Hence, it is recommended that this method be used for cases where the directional asymmetry is not significant. Precipitation and temperature can be simulated with this algorithm.

6 New Objective Functions in the Spectral Domain

Objective functions or diagnostic measures to assess quality of simulations that operate in the spectral domain, Fourier, are proposed in this chapter. In essence, the time series of properties of a variable are dealt with rather than only using the time series of variables themselves. These depend on the cumulative power spectrum ($CMPOW$), a measure proposed in this thesis to describe dependence in the frequency domain, which in essence is the power spectrum (or periodogram) of a series that are derived from reference time series. The philosophy behind using such an approach is that single-valued descriptors of distributions and relationships e.g., mean, variance, skewness, kurtosis, Pearson's correlation (ρ_p), Spearman's correlation (ρ_s), order asymmetry (AO) and directional asymmetry (AD) suffer from the drawback that they can be influenced by tails of their distribution (or outliers). On the other hand, taking a full spectrum to represent a time series is a much stronger constraint. However, it is also not completely immune to outliers. The coming topics assume that working with the proposed power spectra lead to better simulations while ignoring the phase spectra completely. For the sake of completeness, all terms involved are given below for a generic one-dimensional case.

Suppose a variable X, composed of N real values, whose discrete forward real Fourier transform (FT) is given by,

$$FT_X(k) = \sum_{i=0}^{N-1} x(i) exp\left(\frac{-2\pi jik}{N}\right) \quad \text{for } k = 0, \ldots, N \tag{6.1}$$

where, k is the frequency, $x(i)$ is the ith value of X, and j is $\sqrt{-1}$ with each term in Eq. 6.1 being a complex number.

The power spectrum (POW) of the series X can be written as,

$$POW_X(k) = |FT_X(k)|^2 \quad \text{for } k = 1, \ldots, M \tag{6.2}$$

$$M = \begin{cases} \frac{N}{2}, & \text{if } N \text{ is even} \\ \frac{N+1}{2} - 1, & \text{if } N \text{ is odd} \end{cases} \tag{6.3}$$

Where, $|\ |$ returns the absolute value of its argument. The cumulative power spectrum ($CMPOW$), can then be defined as follows,

$$CMPOW_X(k) = \sum_{i=1}^{k} POW_X(i) \quad \text{for } k = 1, \ldots, M \tag{6.4}$$

Using a proper transform of a time series X in Eq. 6.1 to compute $CMPOW$, which corresponds to the dynamic behavior of the transformed data, e.g., copula asymmetry in any given direction, through time. Such a behavior is missing when using only one value to represent the entire time series' behavior because there are a large number of possibilities using many values to produce only one. Through out this thesis, $CMPOW$ is always normalized by dividing it by the sum of all values of its respective POW. This helps in comparing multiple series where the overall shape is important rather than the absolute values and also does not allow one objective function to dominate the others while optimizing.

Depending on X, $CMPOW$ has a different meaning. For example, if X is the first-order difference of a series then the spectrum represents dependence with respect to X itself such as the auto-correlation or order auto-asymmetry or directional auto-asymmetry or auto-entropy at a lag/lead of l steps. If X is a convolution of the same variable at different locations for the same period of time, e.g., temperature or discharge at two stations, then the spectrum represents properties such as the cross-correlation or order cross-asymmetry or directional cross-asymmetry or cross-entropy. In general, various transforms of data can be used to describe the transform-dependent behavior of the data in the spectral domain.

During optimization, the squared difference of $CMPOW$ of a reference (corresponding to an observed series) and the simulated series in the objective function is minimized to produce time series that have properties similar to that of the observed.

6.1 Marginal cumulative periodogram

Even though, the reference and simulated series have the exact same variance, the amount of this variance is, generally, not exactly the same for each frequency in both the series. This can be easily corrected. As shown before, the Wiener-Khintchin theorem provides the conversion from the power spectrum of a series to the auto-covariance and vice versa. Whether the covariance or its normalized form, the correlation, is used does not matter as long as the same is used everywhere. This means that to keep the auto-correlation function of a variable being simulated, its own must be matched with that of the reference. This is computationally much more efficient as compared to computing the auto-correlation function explicitly using Eq. 3.5 for each lag/lead step. However, there is one drawback to this that the Fourier-transform assumes that the input time series is periodic i.e., the beginning and ending values are correlated in the same manner as the values are elsewhere in the time series. This is mostly not the case, as was pointed out by Nakamura et al. (2006), who then advised for a truncated Phase Randomization (PR). Fortunately, any weather related variable is periodic due to the annual cycle but there exist other signals such as trends and seasonal cycles which may not repeat every year. Even the hydrological year can be shown to not have the exact same length for all years and consequently its starting and ending time. In general, if a series is long, say more than 10 years at a daily scale, then this does not pose a

great problem and the auto-correlation functions computed using the Wiener-Khintchin theorem and explicit calculations are very similar. The $CMPOW$ of the marginals is denoted by $CMPOW_p$ where, p is for the Pearson's correlation. For a single time series $CMPOW_p$ is obtained when X is the time series of the simulated marginals in Eq. 6.4. Notice that k begins at 1 instead of 0. M is computed using Eq. 6.3. This form is normally known as the *cumulative periodogram*. During optimization, the sum of the squared differences of $CMPOW_p$ have to be minimized between the simulated and the reference. Another important point to mention is that in this case the PHS is completely ignored. This means that even if the auto-covariances are matched, the direction may become negative. There is a 50% chance of this happening. To circumvent this problem, some phases are not randomized, for example those of the annual- and seasonal-cycles. This ensures that summer of the simulated series is observed during the summer of the reference series and not in the winter. This was already mentioned in Sec. 5.2.

6.2 Grade cumulative periodogram

Similarly, the auto-Spearman correlation (ρ_s) can also be preserved by using the grades or non-exceedence probabilities of the series in Eq. 6.4 during optimization. It is denoted by $CMPOW_s$ where, s stands for the Spearman's correlation.

6.3 Copula order asymmetry

Low-lows or high-highs occurring together is what affects the order asymmetry. To describe this measure in the spectral sense, suppose a time series of a variable X with N values having a probability distribution function $F(x)$, a series of components of the order auto-asymmetry (AO_{AUTO}),

$$AO_{AUTO}(t) = \Big(F(x_t) + F(x_{t+l}) - 1.0\Big)^3 \quad \text{for } t = 0, \ldots, N - l - 1 \tag{6.5}$$

can be calculated. Where $F(x_t)$ is the grade of X at time step t and l is the number of lag/lead steps. The series resulting from Eq. 6.5 has l less steps as compared to the original series X. Then, $CMPOW$ of AO_{AUTO} is,

$$CMPOW_{AO\_AUTO}(k) = \sum_{i=1}^{k} POW_{AO\_AUTO}(i) \quad \text{for } k = 1, \ldots, M \tag{6.6}$$

Fig. 6.1 shows AO_{AUTO}, POW_{AO_AUTO}, $CMPOW_{AO\_AUTO}$ of daily discharge time series for two different years at the same gauging station with a lag of 28 days i.e., dependence between low and high flows on a monthly basis. To demonstrate and clarify the meaning of this measure, two series are taken as inputs. Generally speaking, departure of $AO_{AUTO}(t)$ from zero signifies dependence in time for low and high flows. Values near zero have a

double meaning, one that mean flows are coming together and the other being the combination of a high and a low value. Both cases do not affect the order asymmetry. Furthermore, negative values show a stronger dependence of low flows while positive signify a stronger dependence of high flows. POW_{AO_AUTO} and $CMPOW_{AO\_AUTO}$ in Fig. 6.1 show that the overall shape is similar for both, with the annual cycle (period=365) being the dominating frequency followed by the seasonal cycles. Another important detail to observe here is the difference in frequencies of periods shorter than six months i.e., the monthly dependence is more correlated to time of the year rather than time of the month. In other words, it signals that relatively short term dependence of high and low flows is not as stable as that of the annual, at least in this case.

Similarly, for two distinct time series X and Y with N values having probability distribution functions $F(x)$ and $F(y)$ respectively, a series of components of the order cross-asymmetry (AO_{CROSS}),

$$AO_{CROSS}(t) = \Big(F(x_t) + F(y_t) - 1.0\Big)^3 \quad \text{for } t = 0, \ldots, N-1 \tag{6.7}$$

can be calculated. Where $F(x_t)$ is the grade of X and $F(y_t)$ is the grade of Y at time step t. Then $CMPOW$ of AO_{CROSS} is,

$$CMPOW_{AO\_CROSS}(k) = \sum_{i=1}^{k} POW_{AO\_CROSS}(i) \quad \text{for } k = 1, \ldots, M \tag{6.8}$$

To compute two distinct AO_{CROSS} series corresponding to that in Fig. 6.1, a total of four input time series are needed. Fig. 6.2 shows AO_{CROSS}, POW_{AO_CROSS}, $CMPOW_{AO\_CROSS}$ of daily discharge time series for two different years at two gauging stations on the same river (X being downstream of Y). Series X and A are for the same station with a different time period while Y and B are for the same station with a different time period. The time period of X and Y is similar while the time period of A and B is similar.

An important difference while comparing AO_{AUTO} in Fig. 6.1 to AO_{CROSS} in Fig. 6.2 is the range of the values. The latter shows almost twice the range, signifying that the dependence of lows and highs for two neighboring stations on a daily scale is much higher than that of a single station on a monthly scale. This can also be seen in the relatively larger values of POW_{AO_CROSS}. Consequently, this information can also be used to find out the distance and the period of time, over a catchment, in which the dependence between highs and lows is very similar. Such an experiment might not work, mainly because gauges are spaced at irregular points in space. Nonetheless, this information is useful while simulating in space-time as only one $CMPOW_{AO\_CROSS}$ has to be preserved. Similar to $CMPOW_{AO\_AUTO}$, the overall shape of $CMPOW_{AO\_CROSS}$ is similar for both with the annual cycle (period=365) being the dominating frequency followed by the seasonal cycles. Another important detail to observe here is that the difference in frequencies of periods shorter than six months are, again, not similar. Meaning that the spatial dependence is more a function of the time of the year. Yet another interesting thing to note is that the beginning terms of POW_{AO_CROSS} and

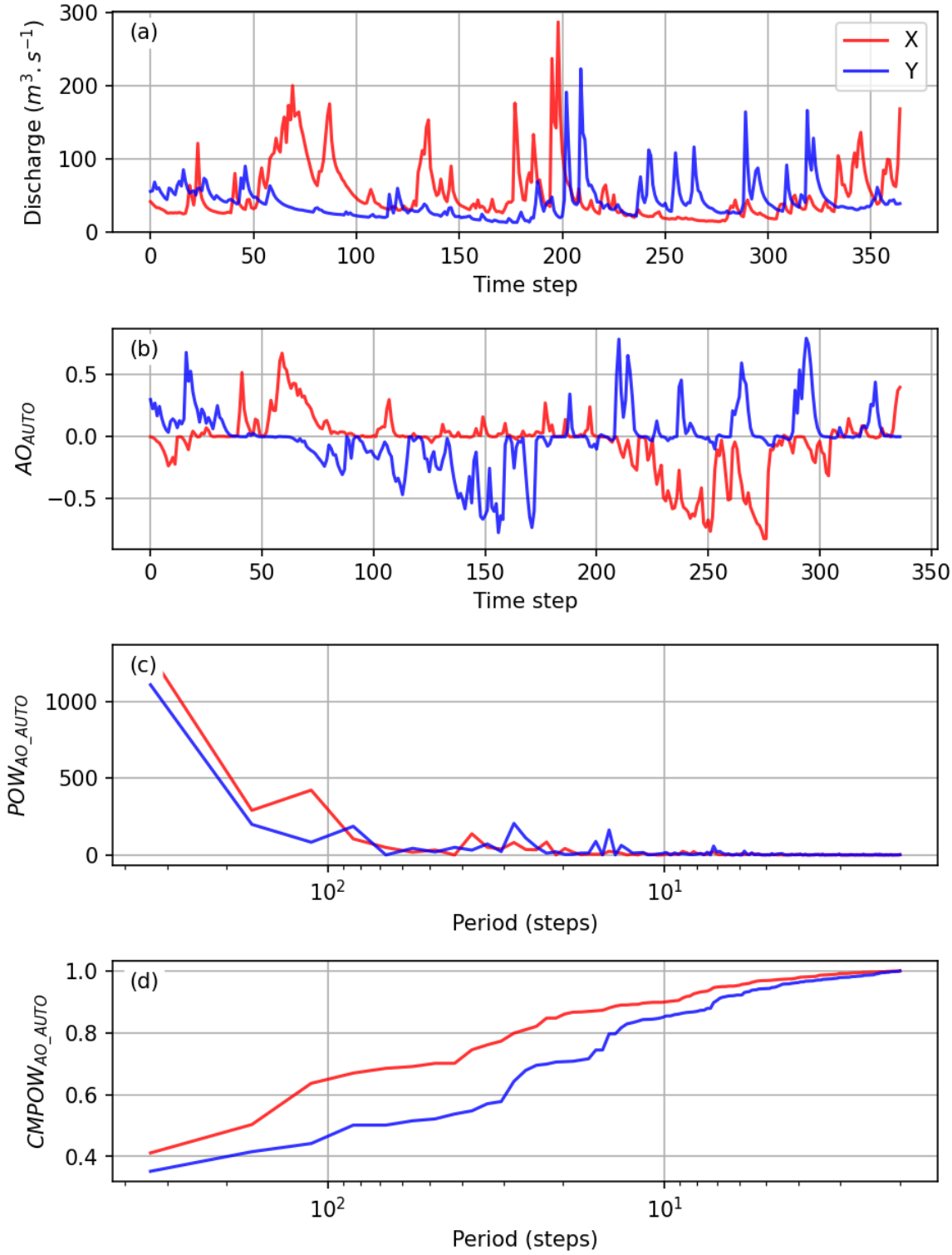


Figure 6.1: Cumulative power spectrum of the order auto-asymmetry ($CMPOW_{AO\_AUTO}$) of two discharge series for the same station but different years computed using lead steps of 28 days.

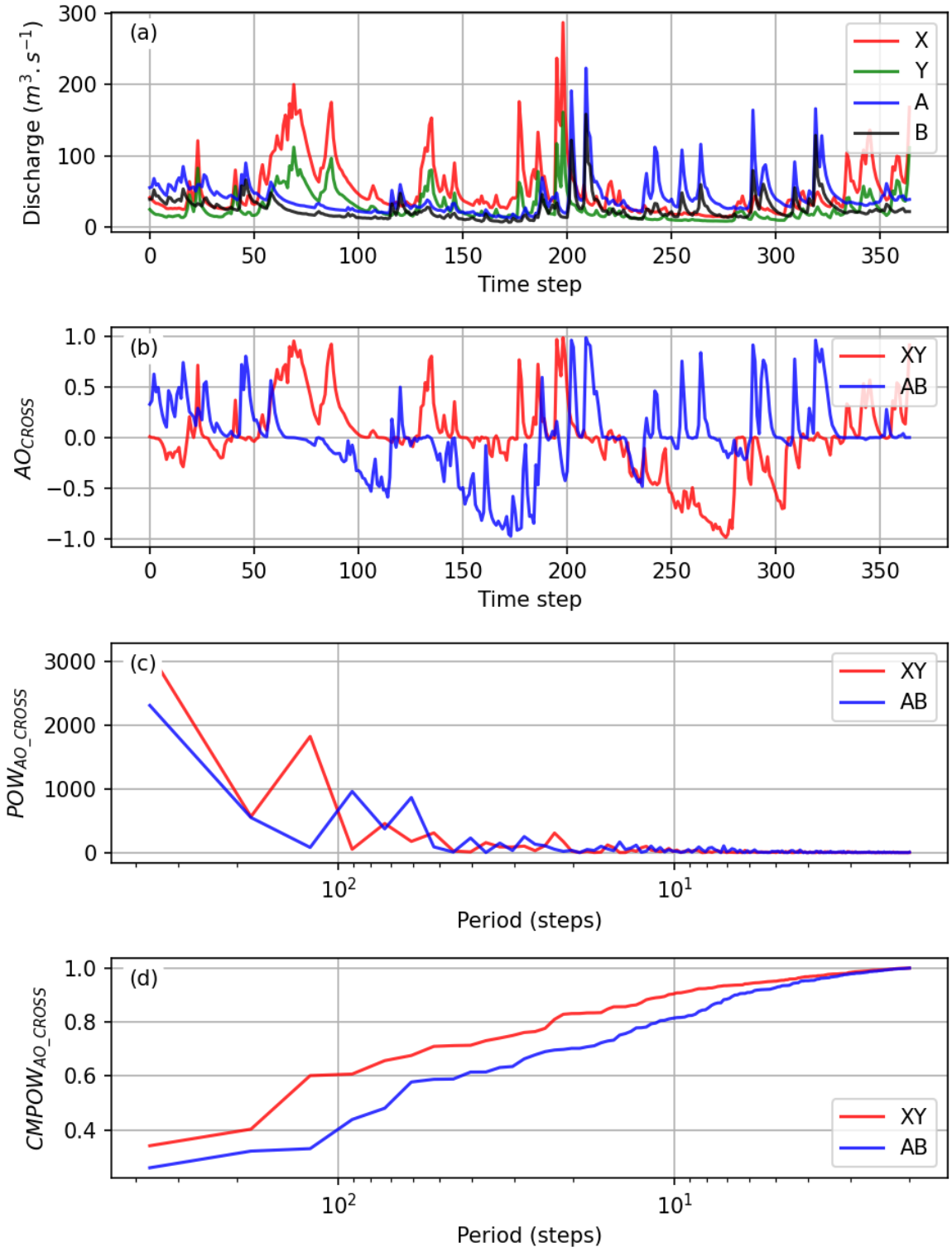


Figure 6.2: Cumulative power spectrum of the order cross-asymmetry ($CMPOW_{AO\_CROSS}$) of four discharge series.

$CMPOW_{AO\_CROSS}$ are very close to zero, this shows that in space the relative dependence of low and high flows is very similar.

6.4 Copula directional asymmetry

Lows followed by highs or vice versa is what affects the directional asymmetry. To describe this measure in the spectral sense, suppose a time series of a variable X with N values having a probability distribution function $F(x)$, a series of components of the directional auto-asymmetry (AD_{AUTO}),

$$AD_{AUTO}(t) = \Big(F(x_t) - F(x_{t+l})\Big)^3 \quad \text{for } t = 0, \ldots, N-l-1 \tag{6.9}$$

can be calculated. Where $F(x_t)$ is the grade of X at time step t and l is the number of lag steps. The series resulting from Eq. 6.9 has l less steps as compared to the original series X. Then, $CMPOW$ of AD_{AUTO} is,

$$CMPOW_{AD\_AUTO}(k) = \sum_{i=1}^{k} POW_{AD\_AUTO}(i) \quad \text{for } k = 1, \ldots, M \tag{6.10}$$

Fig. 6.3 shows AD_{AUTO}, POW_{AD_AUTO}, $CMPOW_{AD\_AUTO}$ of daily discharge time series for two different years at the same gauging station with a lag of 1 day i.e., dependence between low and high flows on a daily basis. To demonstrate and clarify the meaning of this measure, two series are taken as inputs.

Generally speaking, departure of $AD_{AUTO}(t)$ from zero signifies occurrence of lows followed by highs, or vice versa, such as the rising limb of an event in a hydrograph, the greater the difference between the successive values the larger the asymmetry. In a certain sense, directional asymmetry is the opposite of order asymmetry. The main difference is that very different values following each other are what constitute directional asymmetry while everything else has a relatively weak effect. Values near zero are a result of similarly ranked values following each other e.g., low-low or high-high. Both cases do not affect the directional asymmetry. Furthermore, negative values mean a small value followed by a large one and vice versa for the positive ones. POW_{AD_AUTO} and $CMPOW_{AD\_AUTO}$ in Fig. 6.3 show that there is no dominating frequency as compared to order asymmetry.

Similarly, for two distinct time series X and Y with N values having probability distribution functions $F(x)$ and $F(y)$ respectively, a series of components of the directional cross-asymmetry (AD_{CROSS}),

$$AD_{CROSS}(t) = \Big(F(x_t) - F(y_t)\Big)^3 \quad \text{for } t = 0, \ldots, N-1 \tag{6.11}$$

can be calculated. Where, $F(x_t)$ is the grade of X and $F(y_t)$ is the grade of Y at time step t. Then, $CMPOW$ of AD_{CROSS} is,

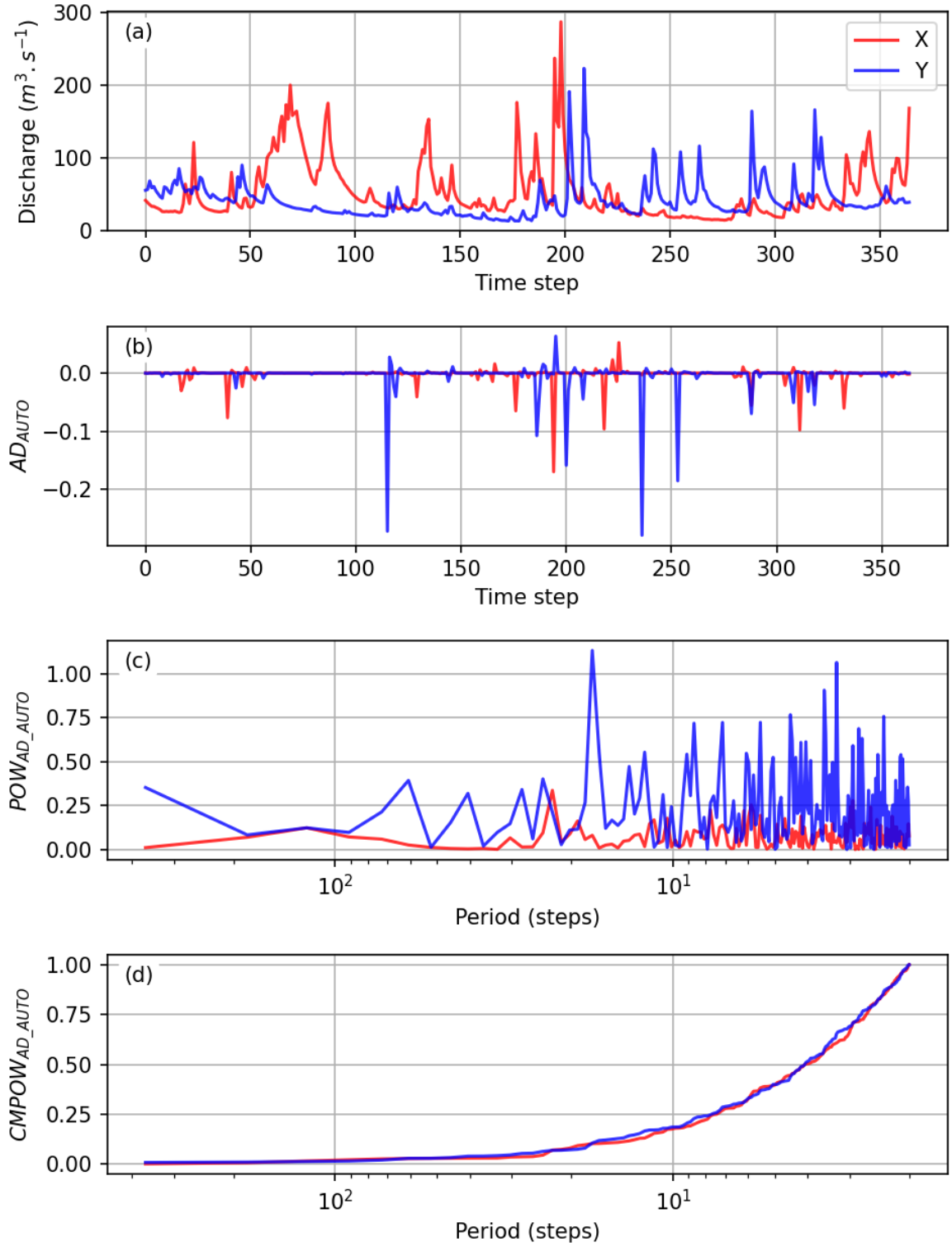


Figure 6.3: Cumulative power spectrum of the directional auto-asymmetry ($CMPOW_{AD\_AUTO}$) of two discharge series for the same station but different years computed using a lead step of 1 day.

$$CMPOW_{AD\_CROSS}(k) = \sum_{i=1}^{k} POW_{AD\_CROSS}(i) \quad \text{for } k = 1, \ldots, M \tag{6.12}$$

To compute two distinct AD_{CROSS} series corresponding to that in Fig. 6.3, a total of four input time series are needed. Fig. 6.4 shows AD_{CROSS}, POW_{AD_CROSS}, $CMPOW_{AD\_CROSS}$ of daily discharge time series for two different years at two gauging stations on the same river (X being downstream of Y). Series X and A are for the same station with a different time period while Y and B are for the same station with a different time period. The time period of X and Y is similar while the time period of A and B is similar.

An important difference while comparing AD_{AUTO} in Fig. 6.3 to AD_{CROSS} in Fig. 6.4 is the direction of the values. The latter shows predominantly positive values i.e., station X experienced larger flows at the same time step as compared to Y but not the other way around. It makes sense as X is downstream of Y and could possibly get more discharge from a smaller catchment connecting to the stream while Y experienced smaller runoff. Most of the time both stations experienced similarly ranked flows, meaning that the weather over both the gauges was frequently similar. The relatively smaller values of POW_{AD_CROSS} as compared to POW_{AD_AUTO} show that directional asymmetry is much weaker in space as compared to time. POW_{AD_CROSS} of both the series show very different concentration of dominant frequencies. This could be an indication that directional cross-asymmetry is highly uncertain when moving in time. It is noteworthy that for one time period (XY) the asymmetry is relatively independent of time but for the next time period (AB) it is dominated by longer and some shorter waves, even though both the series represent the same points in space.

6.5 Copula entropy

Considering the same case of daily discharge from the previous two sections, using $CMPOW_{AD\_AUTO}$ and $CMPOW_{AD\_CROSS}$ as objective functions have a major drawback while simulating series that may change very slowly as long as no rainfall/snow-melt occurs and exhibit a dissimilar change in dependence as more water suddenly accumulates in the stream. The resulting asymmetries are dominated by such events where a low-high takes place that is far away from the diagonal in the copula. For example, Fig. 6.3 shows that the time series of $AD_{AUTO}(t)$ is very often close to zero and deviates significantly upwards from zero at a few time steps only. These peaks dominate the variance of the time series and consequently its magnitude spectrum. Using such a measure for simulation will aim at getting the tails of the distributions right while neglecting the central values. These central values (that change slowly over time) give copula of a discharge series an entropy that is smaller than that of the Gaussian with a similar correlation matrix. The smaller entropy is the result of values clustered around the line $u_1 = u_2$ with very few values that are far away from it. Were it not for these outliers, the correlations would have been much stronger. This is where copula entropy becomes helpful.

Copula local Mutual Information (MI) can be used to simulate series that have entropies close to that of the reference. Similar values occurring in close succession in space-time lead

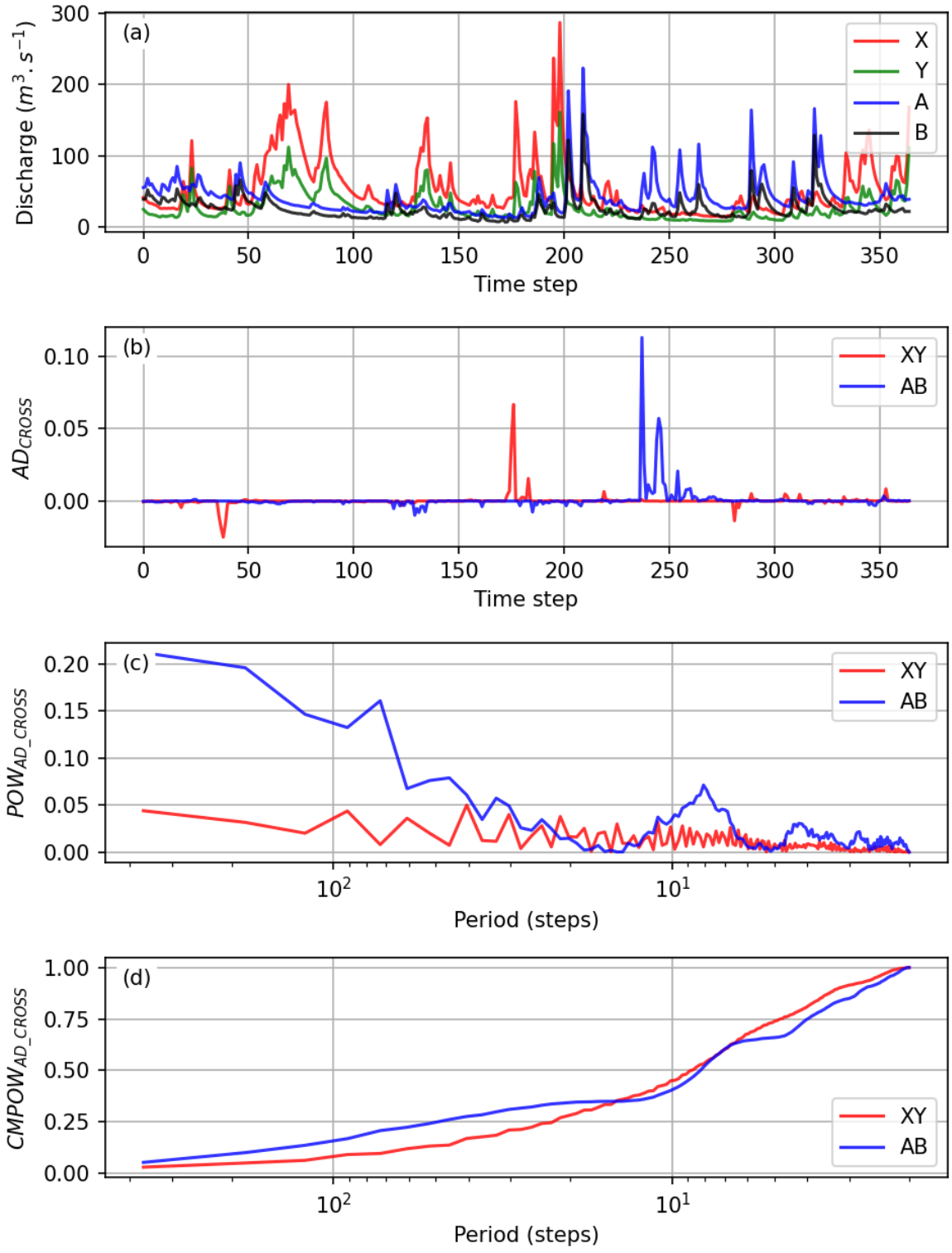


Figure 6.4: Cumulative power spectrum of the directional cross-asymmetry ($CMPOW_{AD\_CROSS}$) of four discharge series.

to high local MI while values that belong to low density regions of a copula get a smaller value. This results in a time series whose variance is dominated by the high local entropy values, exactly what the previous measures tended to ignore. To bring this concept to a usable form, consider a time series of a variable X with N values. Then using Eq. 3.25, $MI(t)$ with u_2 leading u_1 for various time steps can be computed, for example. Using the same methodology developed in the previous two subsections, $CMPOW$ of MI_{AUTO} can be written as,

$$CMPOW_{MI\_AUTO}(k) = \sum_{i=1}^{k} POW_{MI\_AUTO}(i) \quad \text{for } k = 1, \ldots, M \tag{6.13}$$

Fig. 6.5 shows MI_{AUTO}, POW_{MI_AUTO} and $CMPOW_{MI\_AUTO}$ of daily discharge time series for two different years at the same gauging station with a lead step of 1 day. Interesting to note here is when MI_{AUTO} becomes large, locally. Independent of the magnitude of the discharge, if changes in subsequent values are comparatively small MI_{AUTO} becomes large e.g., slowly increasing flows and stable low flows. MI_{AUTO} is smallest when abrupt changes happen in subsequent ranks e.g., a sudden rise in discharge or fluctuating low flows. In general, it could be stated that local entropies are high when a given point in the copula belongs to a bin with a higher relative frequency.

Similarly, if u_1 and u_2 represent two distinct time series in Eq. 3.25 then, $CMPOW$ of MI_{CROSS} can be written as,

$$CMPOW_{MI\_CROSS}(k) = \sum_{i=1}^{k} POW_{MI\_CROSS}(i) \quad \text{for } k = 1, \ldots, M \tag{6.14}$$

Fig. 6.6 shows MI_{CROSS}, POW_{MI_CROSS} and $CMPOW_{MI\_CROSS}$ of daily discharge time series for two different years at two gauging stations on the same river (X being downstream of Y). Series X and A are for the same station with a different time period while Y and B are for the same station with a different time period. The time period of X and Y is similar while the time period of A and B is similar. MI_{CROSS} increases if both the time series go high or low proportionally to each other. In other words, MI_{CROSS} represents the relative timing of the two time series locally. Highly synchronized events at both locations lead to higher local entropy.

6.6 Drawbacks

However, even after taking local entropies into account along with other measures, still all statistical properties of the time series are not recovered. Local MI does suffer from a drawback. For the case of AD_{AUTO} or AD_{CROSS}, relatively few events dominated the variance but for MI there may be many (increasing the number of copula bins might help). This as a result might lead to some smoothness in the simulated time series. Matching the $CMPOW$s does alleviate the problem but not entirely. This is, unfortunately, true for any

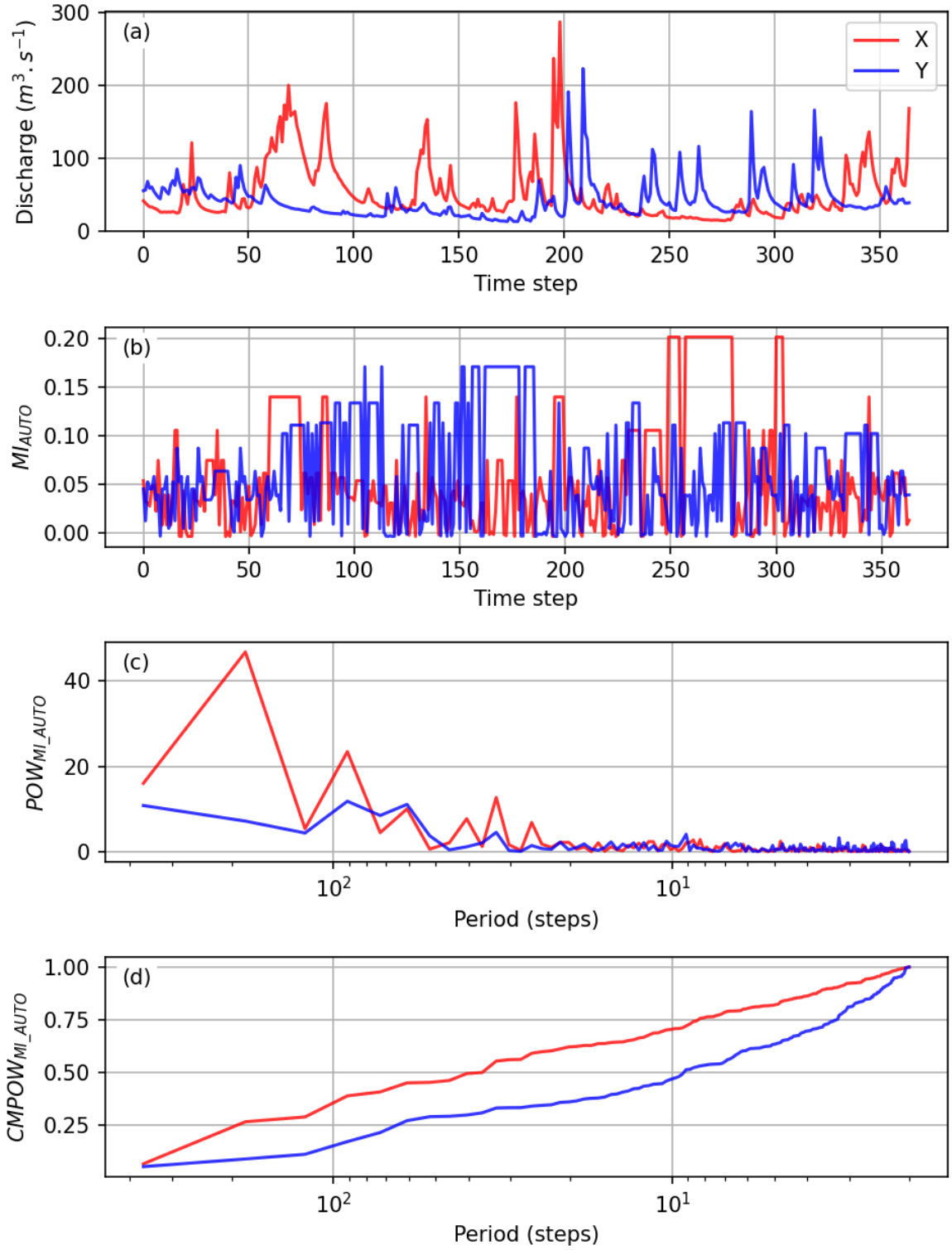


Figure 6.5: Cumulative power spectrum of the local Mutual Information ($CMPOW_{MI\_AUTO}$) of two discharge series for the same station but different years computed using a lead step of 1 day.

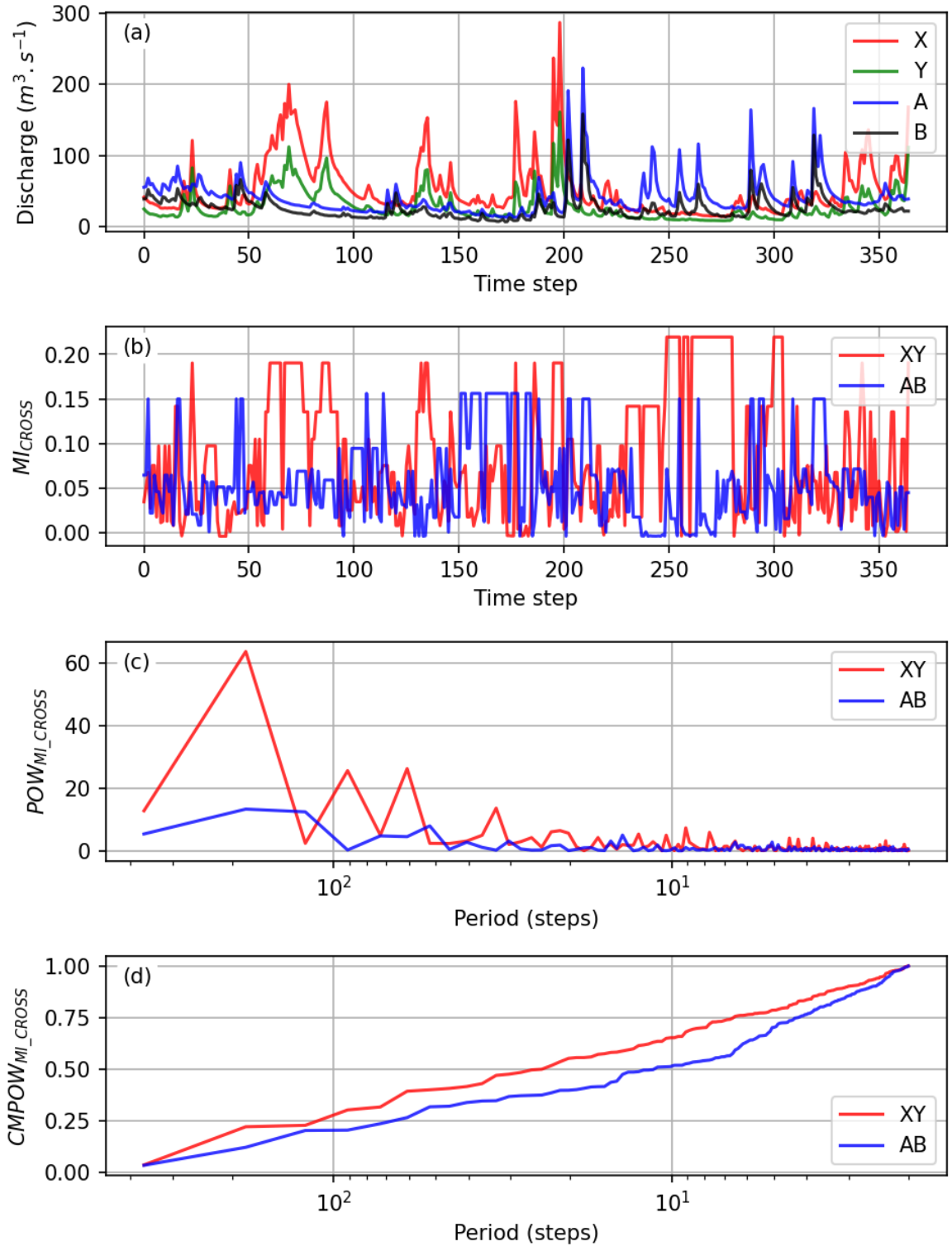


Figure 6.6: Cumulative power spectrum of the local Mutual Information ($CMPOW_{MI\_CROSS}$) of four discharge series.

single measure. According to Fernández (2013), to describe all the features of a time series, very many (almost equal to the number of data points) single-valued statistical descriptors are needed. This would of course lead to an over-fit. What is more appropriate is that the utilized descriptors must be somehow related to the properties that the simulations must preserve.

7 Spatial Aspects of Hydrological Extremes

The main focus of this thesis is on identifying, describing and simulating hydrological extremes. Mainly, *extreme floods* occurring *simultaneously or parallely* at multiple locations. Here, the term simultaneous refers to close proximity in time. Suppose the question, if a hundred year floods occurs at station A then what is the probability that another flood of a given return period happens at station B within one day or what happens at station B in general? Where, the threshold of one day is decided upon by the decision maker and any events happening within this window constitute a simultaneous occurrence. The same question can be formulated for M number of locations of interest. An extreme flood refers to flows in streams due to heavy rainfall or melting of snow cover rapidly due to a sudden rise in temperature that is enough to induce a large snow-melt to cause flows that end in life and/or property destruction downstream. Traditionally, simultaneous extremes are rarely studied, only point extremes are dealt with rigor. Mainly due to the complexity involved at such large spatial and short temporal scales and also the curse of dimensionality because the number of such events is a small subset of the observations anyhow and analyzing them in higher dimensions is not optimal and involves making assumptions on the behavior of the variable in space-time that are difficult to validate. Most applications solve the prediction of extremes by using Flood Frequency Analysis (FFA). Their focus is on fitting a distribution to the extreme events, say one per year or all flows above a set threshold. The advantage being that flows of even higher return periods can be estimated at a single point in space. The disadvantage is that it gives no information about the processes involved or an indication of circumstances under which these flows might occur. Conditional distribution of events can be taken, say summer and winter events, but that reduces the already small pool of values to an even smaller one leading to much more uncertainty and over-fitting.

Understanding the spatio-temporal aspects of extremes completely with the state-of-the-art knowledge is still not possible. This thesis will not solve this problem entirely but will provide more tools to identify the issues with the currently used techniques and offer some methods to solve them i.e., better simulations. The opposite of extreme floods i.e., droughts are also hydrological extremes. These are not dealt with explicitly here. However, the tools provided here apply to drought simulation as well. For droughts, the focus is on the length of no/little precipitation or very low flow durations.

It is assumed throughout this thesis, unlike FFA where, only existing extreme events to predict other extremes is not enough. Furthermore, a stronger assumption is made that all the important properties mentioned so far of multiple variables in space and time have to be preserved by simulated series. In other words, the aim is to generate series that are statistical equivalents of a given reference. Such an approach allows for the flexibility to apply the methods presented here to be used almost anywhere without any loss of generality

and not only to hydrological variables. Furthermore, once dependence in space and time is determined, flood magnitudes using FFA can be used at a location to estimate variable states under consideration at other sites. *This chapter focuses on measures that may be used to identify and describe the behavior of time series in general with simultaneous extremes considered implicitly.*

The multivariate dependence of precipitation and discharge series are considered mainly as demonstration variables. The other important aspect of extremes is the state of the catchment such as land-use and geology but these are set aside as they are much more consistent in time than any other variable and also because the relationship between all variables and discharge is the result of the catchment state, which can be obtained from the dependence structure of all these variables or the state variables of a rainfall-runoff model to a large extent.

Analyzing the combined behavior of precipitation and discharge at the same location shows that they do not correspond as they should. Mainly, because it does not rain all the time but the rivers do flow, almost, all the time, albeit with changing magnitude of flow. Ultimately, high flows in streams are caused by some form of precipitation. The delay in discharge series compared to precipitation is because precipitation has to interact with several variables on the way before its effect appears as discharge in streams. Describing these processes is not trivial. Efforts have been made in the past to have models that represent reality but all suffer from one major drawback or another or in many cases several drawbacks. For precipitation, it takes a finite amount of time till a rise in stream flow is observed after a rainfall or snow-melt event. The main modifiers here being temperature and the catchment state. For example, suppose it is above freezing and it rains. Depending on how wet the soil of the catchment is, which is related to time of the year (amount of vegetation that intercepts rain) and how much rain had occurred previously, different hydrographs namely the rising and recession limbs and the peak values can be observed. Another example could be that of snow-melt which is mainly controlled by the atmosphere over a time scale that is always longer than that of a typical rainfall-runoff event.

Another crucial difference that complicates the combined modeling of precipitation and discharge is that precipitation happens discontinuously in space and time but discharge is recorded only at one point in space and continuously in time. To make things simpler, point precipitation at multiple locations is also converted to a spatially-averaged point time series similar to that of discharge by interpolating it in space at multiple points for any given time step and then averaging it for the entire catchment that contributes to the point where discharge is being recorded. Such a method introduces a large uncertainty in the total amount of precipitation. Interpolation schemes aim to interpolate at unknown locations based on observations only. Hence, the volume that is obtained is a function of these observations which may or may not be a sample big enough to capture the variability, and volume, fully at the catchment scale. Given that the distribution of precipitation is exponential in space at any time step, the chances that a gauge actually captures the biggest value are practically zero. This was discussed at length in Bárdossy and Anwar (2023).

Larger catchments behave much more smoothly as compared to smaller ones because they act as much larger low-pass filters. Smaller catchments on the other hand can present trickier situations. For example, sewage flows represent a larger proportion of the low-flow

season. Also, it is possible that water from other catchments enters in to the stream through underground cracks in rocks as discharge with no direct precipitation over the catchment.

On the other hand, temperature behaves much more smoothly in space-time and does not show the discontinuous behavior similar to precipitation. Same applies to potential evapotranspiration which is mostly a function of vegetation and temperature (however, wind can have a considerable impact at consistently high speeds) that could be discontinuous but behaves much more consistently on a catchment scale. Temperature gradient inversions in atmosphere can be a problem for snow-melt simulation at certain locations.

All the aforementioned challenges would be somewhat simpler to handle, if the data were observed at finer spatio-temporal scales. The lower temporal resolution discharge presents itself as a problem when precipitation concentration times are smaller than the frequency of observed discharge. This mainly happens for smaller catchments where the times-of-concentration can be in minutes/hours while the recorded discharges are on daily basis.

To address the problems mentioned above, several tools and methods are described in literature that are in use and produce acceptable results. Some useful tools were presented in previous chapters to analyze dependence and simulate time series. These are combined in the coming subsections to assess realizations/simulations that are close to a given reference in terms of dependence.

7.1 Desired properties of measures describing higher-order dependence

Before defining measures that describe dependence in a truly multivariate sense, some characteristics that all these measures should have are proposed. Here, *the term higher-order dependence refers to the combined behavior of variables in more than 2 dimensions* i.e., unlike the previously described measures that only considered pairs, these should be able to take more than two variables. First and foremost, these measures should reflect the change in dependence as variables are included in or excluded from any existing arrangement. For example, consider three variables and their accompanying Pearson's correlations (ρ_p). In total, there are three correlation values. Now suppose, the third variable is exchanged with a new one. As a result the new correlations change but not all of them. One of them is independent of the new variable. In other words, the change in dependence is not reflected fully. It is expected from measures describing/quantifying higher-order dependence that they reflect the change in it everywhere as the variables change.

Secondly, these measures should reflect the expected properties of the multivariate Gaussian copula when large samples are drawn from it and used as input. For example, the directional asymmetry (AD) for the Gaussian copula in any number of dimensions is zero on the mean. While, the order asymmetry (AO) can fluctuate much more widely around zero, but it should show a very consistent behavior when looking from various directions. Note that this may not be the case for variables with truncated behavior such as precipitation where the zeros represent a major portion of the distribution.

Further tests can be made by using d random variables with N values such that all variables are copies of each other i.e., $x_1^i = \ldots = x_d^i$. Using this as an input, the measures should produce the same value for order asymmetries (these can be non-zero for mixed distributions) and have zero directional asymmetries.

7.2 Copula asymmetries in higher dimensions

In Chapter 3, copula order and directional asymmetries (AO and AD respectively) were presented. These were for the bivariate case. The same concept can be extended to d dimensions. The main focus being on the fact that each point's dependence should reflect a d-dimensional dependence and not only in a bivariate sense as pairwise correlations. This is achieved by resorting to *Geometry*. The idea is as follows. Bivariate asymmetries were calculated based on distances of points from a line in a given direction in two dimensions in Chapter 3. The same can be employed in d dimensions i.e., the shortest possible distance of a d-dimensional point to a d-dimensional line in a d-dimensional hypercube and the vector that forms it. From hereon, order asymmetry is referred to as AO^d instead of AO where, d represents the dimensions of the copula. d can be any positive number greater than one. The same applies to the directional asymmetry. Furthermore, asymmetries will now be presented as d-dimensional vectors (they cannot be scalars for obvious reasons) and will be referred to as d-asymmetries as a whole. They will be mainly used when more than two variables are considered.

7.2.1 Order asymmetry in d dimensions

The bivariate case of raw AO is considered that is then generalized to higher dimensions so that the two approaches can be related. Consider a point u with the coordinates (u_1, u_2) in a copula and the vector $\vec{AB}$ (the line $u_2 = 1 - u_1$). Let the point at which the shortest line passing through u and intersecting $\vec{AB}$ be R. Then, the distance between u and R is given by,

$$distance(R, u) = \frac{|u_1 + u_2 - 1|}{\sqrt{2}} \tag{7.1}$$

And the coordinates of R are,

$$x_R = \frac{u_1 - u_2 + 1}{2} \quad \text{and} \quad y_R = \frac{-u_1 + u_2 + 1}{2} \tag{7.2}$$

Then, raw AO^2 for N such points $[u_1^N, u_2^N]$ is defined as,

$$AO^{raw,2} = \left[\frac{1}{N} \sum_{i=0}^{N-1} \left(u_1^i - x_i\right)^3, \frac{1}{N} \sum_{i=0}^{N-1} \left(u_2^i - y_i\right)^3 \right] \tag{7.3}$$

Before the d-dimensional case is presented, it is useful to discuss the relationship between $AO^{raw,2}$ and AO_{raw}. For the case of AO_{raw} in Sec. 3.4 the distance of a point was calculated from the line $u_2 = 1 - u_1$ (using Eq. 7.1 and without the denominator) then, assigned a negative sign when the u_1 was less then x_R or a positive sign when the u_1 was greater then x_R. Afterwards, this signed distance was raised to the power of three. The exponent of three only serves to amplify the weight of larger distances and to keep the sign of the direction. Finally, AO_{raw} was the mean of all these signed distances. For the case of $AO^{raw,2}$, the magnitude of the components of the vectors are raised to the power three first, of the line joining R and $\vec{AB}$ and then the mean at each component is calculated. In other words, the coordinate-wise distance rather then the bivariate distance is considered. It is done so to accommodate the higher dimensional cases in which the convenient positive or negative signs cannot be utilized.

In three dimensions, there are three possible directions that are analogous to the lines that were used in the two-dimensional case, they become four in the four-dimensional case, five in the five-dimensional case and so on. Using linear algebra, the general case of order asymmetries is presented. Suppose a direction vector $\vec{AB}$ passing through two points A and B and a point u having the coordinates $(u_1, \ldots, u_d)$. $\vec{AB}$ can only have certain directions. It cannot be on any of the $(d-1)$-dimensional faces and must pass through two opposite corners of the d-dimensional hypercube. u can be anywhere in the hypercube. Also, let the vector pointing from A to u be $\vec{Au}$. Let the line that passes through u while intersecting $\vec{AB}$ perpendicularly be $\vec{R}$. Then, the coordinates of the intersection point (R) on $\vec{AB}$ can be written as,

$$R = \vec{AB}\frac{\vec{Au} \cdot \vec{AB}}{\vec{AB} \cdot \vec{AB}} + \vec{A} \tag{7.4}$$

Then, the theoretical forms of d-dimensional j-th raw order asymmetry ($AO_j^{raw,d}$) may be written as,

$$AO_j^{raw,d} = \left[E[(U_1 - R_1)^3], \ldots, E[(U_d - R_d)^3]\right]_j \tag{7.5}$$

or

$$AO_j^{raw,d} = \left[\int_{[0,1]^d} \left(u_1 - r_1\right)^3 du_d \ldots du_1, \ldots, \int_{[0,1]^d} \left(u_d - r_d\right)^3 du_d \ldots du_1\right]_j \tag{7.6}$$

where, $R_1, \ldots, R_d$ and $r_1, \ldots, r_d$ refer to the components of R, depending on the employed form. More importantly, the empirical form of d-dimensional j-th raw order asymmetry ($AO_j^{raw,d}$) for N points $[u_1^N, \ldots, u_d^N]$ can be written as,

$$AO_j^{raw,d} = \left[\frac{1}{N}\sum_{i=0}^{N-1}\left(u_1^i - r_1^i\right)^3, \ldots, \frac{1}{N}\sum_{i=0}^{N-1}\left(u_d^i - r_d^i\right)^3\right]_j \tag{7.7}$$

where, j represents the j-th vector $\vec{AB}$. As mentioned before, the vector $\vec{AB}$ can only have d directions. The idea being similar to that in AO_{raw} i.e., while one component goes low, the others go high. In case of $AO^{raw,d}$, while one component goes low, all other go high. For example, in the three-dimensional case points A and B will have the following coordinates along with the direction of the corresponding direction vector $\vec{AB}$,

j	A	B	$\vec{AB}$
1	[0, 1, 1]	[1, 0, 0]	[+1, -1, -1]
2	[1, 0, 1]	[0, 1, 0]	[-1, +1, -1]
3	[1, 1, 0]	[0, 0, 1]	[-1, -1, +1]

In four dimensions, there are a total of four such direction vectors possible,

j	A	B	$\vec{AB}$
1	[0, 1, 1, 1]	[1, 0, 0, 0]	[+1, -1, -1, -1]
2	[1, 0, 1, 1]	[0, 1, 0, 0]	[-1, +1, -1, -1]
3	[1, 1, 0, 1]	[0, 0, 1, 0]	[-1, -1, +1, -1]
4	[1, 1, 1, 0]	[0, 0, 0, 1]	[-1, -1, -1, +1]

For the d-dimensional case, a row and a column are added to get the corresponding direction vectors $\vec{AB}$ while following the same procedure as that in the above mentioned cases. A further restriction that applies is that of the intersection set of the points in A and B should be empty. Due to this reason, for $AO^{raw,2}$, there is only one possible vector $\vec{AB}$.

$AO^{raw,d}$ cannot be normalized as the ρ_d has no defined lower bounds. This was mentioned already in Sec. 3.7. Hence, d-dimensional j-th order asymmetry (AO^d_j) can be written as,

$$AO^d_j = AO^{raw,d}_j \tag{7.8}$$

The drawback of Eq. 7.8 is that it is only comparable for two cases when both of them have the same ρ_d.

7.2.2 Directional asymmetry in d dimensions

The case of the d-directional asymmetry is very similar to that of the order asymmetries. Firstly, AD becomes AD^d. There is only one possible direction that AD^d can have i.e., all components go from low to high. More specifically, all points project on the line $u_1 = u_2 = \ldots = u_d$. For example, coordinates of point A and B are $[0, ..., 0]^d$ and $[1, ..., 1]^d$ respectively while the direction of $\vec{AB}$ is $[1, ..., 1]^d$ in d-dimensions. The rest of the calculations stay similar to those shown in Sec. 7.2.1 for AO^d. d-directional asymmetry can also be modified to show relative strength of dependence when comparing two or more multivariate series. Namely, by considering the absolute distance of points from $\vec{AB}$. This is left out as an outlook of this study.

7.2.3 Comparing old and new copula asymmetries

Using the equations described above and the ones described in Chapter 3 raw order and directional asymmetries are computed to demonstrate how the two forms vary in their final values but convey similar information when using the same input data. Given in Fig. 7.1 is an empirical copula of five points P0, P1, P2, P3 and P4 that is used for demonstration. Their coordinates in the copula are given in the columns $C1$ and $C2$ of Table 7.1.

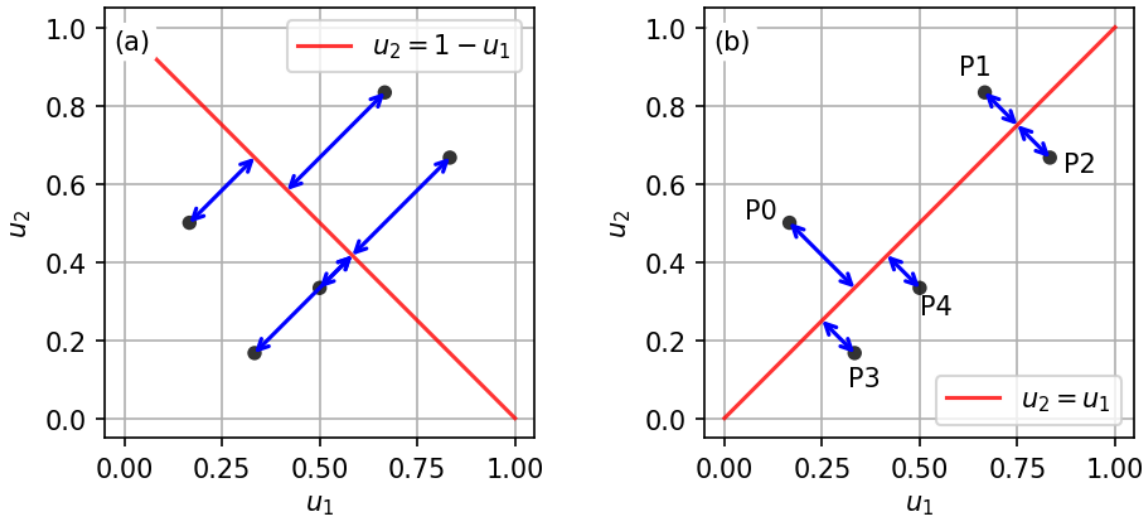


Figure 7.1: Exemplary empirical copula of five points for demonstrating the difference between the old and new asymmetries. (a) shows the distance of points for the case of the order asymmetry while (b) shows the distance of points for the case of the directional asymmetry.

Table 7.2 shows calculations for AO_{raw} and AD_{raw} (the older versions from Chapter 3). The column $C0$ shows labels of points that correspond to the points appearing in Fig. 7.1. $C1$ shows the result of the inner term of the summation in Eq. 3.9 i.e., $(u_1 + u_2 - 1.0)^3$. The mean of all the terms in the final row, is AO_{raw} of these points. Here, a positive AO_{raw} shows that the dependence of high-highs is larger compared to low-lows. A high-high refers to any point in copula that is above the line $u_2 = 1 - u_1$. Similarly, $C2$ shows the corresponding calculations for AD_{raw}. In this case, a negative value shows that the strength of low-highs is larger than high-lows. A low-high point is any point in copula that is above the line $u_2 = u_1$.

Calculations for $AO^{raw,2}$ are given in the Table 7.3. It can be observed here that more columns i.e., more calculations, are needed as compared to the their older counter parts. Considering the case of $AO^{raw,2}$, columns $C1$ and $C2$ are the coordinates of the points R (which may or may not be unique for all points in a copula) calculated using Eq. 7.4. Here, $\vec{AB}$ is $[+1, -1]$, which is formed by the point A and B with coordinates $[0, 1]$ and $[1, 0]$ respectively. $\vec{AB}$ is the same as the line $u_1 + u_2 - 1.0 = 0$. $C3$ and $C4$ show the corresponding inner terms of the summations in Eq. 7.7 by taking the difference of the respective components of the points in the empirical copula and the points R and then raising them to the power of three. The final row shows their mean value i.e., $AO^{raw,2}$. The resulting vector

of $AO^{raw,2}$ is $[+2.08E-3, +2.08E-3]$. Note that both the coordinates have the same sign and values. The positive sign signifies that the high-highs have a larger strength compared to the low-highs because the vector is pointing towards the upper right corner, with respect to $\vec{AB}$, of the copula. This would be the same as a positive AO_{raw}, which is in fact the case as shown in Table 7.2. The artifact, of both the coordinates of the vector being the same, appears only in the two-dimensional case. To reiterate, note that in two dimensions, only one vector for the order asymmetry is possible unlike three or more dimensions.

Similarly, calculations for $AD^{raw,2}$ are presented in Table 7.4. Here, the only difference compared to $AO^{raw,2}$ is $\vec{AB}$, which is $[1, 1]$ formed by the line joining the points A and B whose coordinates are $[0, 0]$ and $[1, 1]$ respectively. This also changes the coordinates of the points R and consequently all the following calculations. Finally, $AD^{raw,2}$ is yielded by taking the mean of all values in the columns $C3$ and $C4$ i.e., $[-6.94E-4, +6.94E-4]$. Once again, both the coordinates have the same magnitude but opposite signs. In this case $AD^{raw,2}$ is pointing towards the upper left corner, with respect to $\vec{AB}$, of the copula which signifies that u_2 reacted faster than u_1, which the AD_{raw} also revealed in Table 7.2.

Finally, it can be summarized that the absolute values of the old and new asymmetries are not equal, which is not surprising as the steps taken to obtain them are different. However, they convey the same information while considering their signs and directions.

Table 7.1: Coordinates of points in the empirical copula shown in Fig. 7.1 as a table. These are used for exemplary computations of AO_{raw}, AD_{raw}, $AO^{raw,2}$ and $AD^{raw,2}$. Refer to Section 7.2.3 for description.

C0	**C1**	**C2**
Point	u_1	u_2
P0	0.17	0.50
P1	0.67	0.83
P2	0.83	0.67
P3	0.33	0.17
P4	0.50	0.33

Table 7.2: Exemplary calculations for AO_{raw} and AD_{raw}. Refer to Section 7.2.3 for description.

C0	**C1**	**C2**
Point	$(u_1 + u_2 - 1.0)^3$	$(u_1 - u_2)^3$
P0	-3.70E-2	-3.70E-2
P1	+1.25E-1	-4.63E-3
P2	+1.25E-1	+4.63E-3
P3	-1.25E-1	+4.63E-3
P4	-4.63E-3	+4.63E-3
Mean	**+1.67E-2**	**-5.56E-3**

Table 7.3: Exemplary calculations for $AO^{raw,2}$. Refer to Section 7.2.3 for description.

C0	**C1**	**C2**	**C3**	**C4**
Point	r_1	r_2	$(u_1 - r_1)^3$	$(u_2 - r_2)^3$
P0	0.33	0.67	-4.63E-3	-4.63E-3
P1	0.42	0.58	+1.56E-2	+1.56E-2
P2	0.58	0.42	+1.56E-2	+1.56E-2
P3	0.58	0.42	-1.56E-2	-1.56E-2
P4	0.58	0.42	-5.79E-4	-5.79E-4
Mean			**+2.08E-3**	**+2.08E-3**

Table 7.4: Exemplary calculations for $AD^{raw,2}$. Refer to Section 7.2.3 for description.

C0	**C1**	**C2**	**C3**	**C4**
Point	r_1	r_2	$(u_1 - r_1)^3$	$(u_2 - r_2)^3$
P0	0.33	0.33	-4.63E-3	+4.63E-3
P1	0.75	0.75	-5.79E-4	+5.79E-4
P2	0.75	0.75	+5.79E-4	-5.79E-4
P3	0.25	0.25	+5.79E-4	-5.79E-4
P4	0.42	0.42	+5.79E-4	-5.79E-4
Mean			**-6.94E-4**	**+6.94E-4**

7.2.4 Interpretation of d-asymmetries

Fig. 7.2 demonstrate d-asymmetries for 5 discharge time series at various locations in the study area. In the background are series simulated with Gaussian dependence (using the Amplitude Adjusted Fourier Transform (AAFT) algorithm) only to show that the observed

dependence is significantly different. Formal comparisons using various methods are shown in the chapter of Results. d-asymmetries have to be interpreted slightly differently as compared to the older versions described in Chapter 3, in that each component provides information about the behavior in its own dimension. For example, consider the three-dimensional case of order d-asymmetries. For variables having a Gaussian dependence, all the values will swing around zero. In practice, the j-th component of AO_j^d will show much more variability as compared to the rest of the components as it has the opposite direction in terms of the major axis and all the points have the farthest distance from it, assuming the variables' dependence is along the major axis. Slightly larger distances will affect the j-th component the most when they are raised to the power of three. This has an important bearing on how AO_j^ds are interpreted. As the points gather more and more on one side of the vector $\vec{AB}$, a more pronounced deviation away from zero can be observed. The sign of the component will show whether low-lows are more likely or the high-highs. Components apart from the j-th show less deviations comparatively for variables that have their dependence distributed mostly along the major axis. If the case be that the non-j-th components show higher deviations than the j-th then it could be stated with confidence that points in the d-dimensional copula are not along the major axis. This can happen when some variables are inversely related to the others in time. The directional d-asymmetry shows the relative difference in timing of all stations' time series with respect to each other. The component having the largest positive value should be seen as the one that acts the earliest (or fastest) while the one having the largest negative value should be seen as the one that reacts the latest.

It is advised to use d-asymmetries with some healthy skepticism. They represent the mean of the cube of the signed distances in each dimension, which is very sensitive. The mean value can be a result of practically infinite combinations. It could be that only one point that lies away from the direction vector while other lie near to it, produce the same value. The same drawback applies to the older versions.

7.2.5 Usage of d-asymmetries

Use of d-asymmetries is the same as that of the bivariate asymmetries. These serve as diagnostic tools rather than objective functions. It is not advised to use them as objective functions due to the previously mentioned reasons regarding their sensitivity when the points are too few. The comparison of these for the reference and simulations should demonstrate quickly whether a bias in direction of dependence exists or not. Furthermore, as the lower bounds on ρ_d are undefined, d-asymmetries cannot be normalized similar to their original two-dimensional versions. For this reason, only their forms can be compared for any two cases but not the magnitudes when ρ_ds are different.

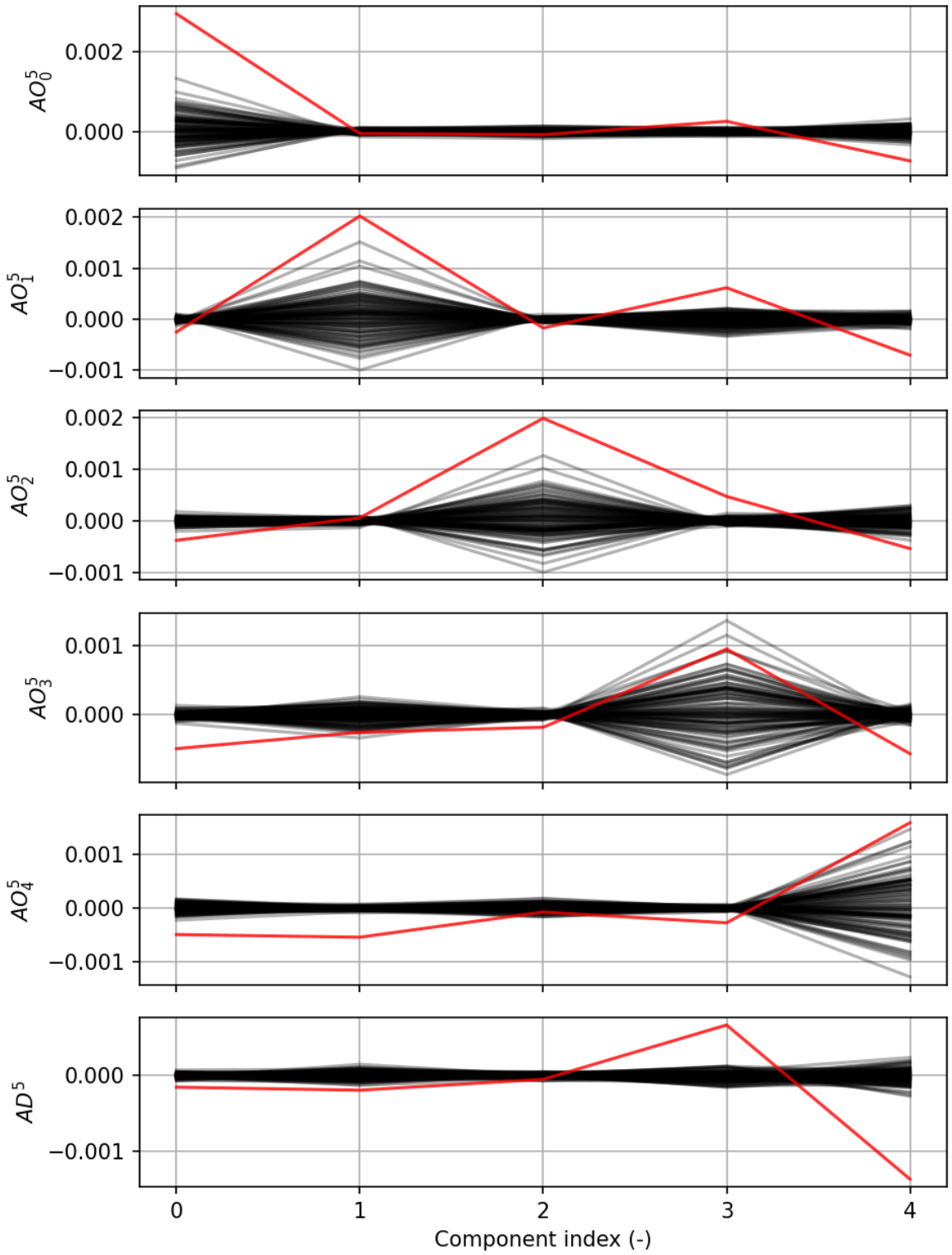


Figure 7.2: d-asymmetries for five discharge stations with daily records for the period of 1961-2015 in the Neckar catchment. In red are observed series while in black are series with Gaussian dependence (AAFT).

7.3 The Fourier transform, maximum correlation and hydrological extremes

One of the main use of the Fourier transform (FT) that was left out in Chapter 5 is that of the maximum possible correlation among multiple variables. In other words, the correlation among any number of variables can be quantified if the timing of events are not considered while keeping their auto-correlation functions intact. This property can be seen as something akin to the worst (or best) possible case that represents all bad (or good) things in all time series happening at the same time at all locations. The value of this correlation ranges from zero to one. The notion of a maximal correlation exists and was given by Gebelein (1941). This can be imagined as the correlation of the sorted values of two series but note that the auto-correlation functions of any two series are not the same for real world cases, hence, the maximal correlation is always less than or equal to the correlation of the sorted series. There are methods of computing the maximal correlation other than the one presented here. These are difficult to program but the one using the FT (developed by the supervisor of the author) is much more straight forward and provides additional information in terms of contributions from individual frequencies to the final value. Following are some formal details.

Suppose a time series of multiple variables (X_M) having M variables and N time steps. Borrowing terminologies developed in Chapter 5, the magnitude spectra (MAG) of all them can be written as,

$$MAG_X^m(k) = |FT_X^m(k)| \quad \text{for } k = 0, \ldots, \lceil N \div 2 \rceil \text{ and } m = 0, \ldots, M-1 \tag{7.9}$$

where, m represents the index of the m-th variable in X_M. Then, the maximum possible cross-power ($MXPOW$) between these M variables is given by,

$$MXPOW(X_0, \ldots, X_{M-1}) = \sum_{k=1}^{\lceil N \div 2 \rceil} \left(\prod_{m=0}^{M-1} MAG_X^m(k) \right) \tag{7.10}$$

where, $\lceil\rceil$ is the ceiling operation. It rounds any value with digits after the decimal to the next integer e.g., 1.01 becomes 2 while 1.0 remain 1.0. Notice that only half the frequencies are used and that the first frequency index which is the sum of the series is not used to compute the power. $MXPOW$ can be normalized to yield the maximum possible cross-correlation ($MXCORR$),

$$MXCORR(X_0, \ldots, X_{M-1}) = \frac{\sum_{k=1}^{\lceil N \div 2 \rceil} \left(\prod_{m=0}^{M-1} MAG_X^m(k) \right)}{\sqrt[M]{\prod_{m=0}^{M-1} \left(\sum_{k=1}^{\lceil N \div 2 \rceil} \left(MAG_X^m(k) \right)^M \right)}} \tag{7.11}$$

Please note that if too many values are taken, numerical overflow may easily result as they have finite precision in the memory of a computer.

7.3.1 Interpretation of the maximum correlation

Eq. 7.10 and 7.11 have some drawbacks. Suppose, two triplet series X_M and Y_M both with N steps. Let it be the case, that they both produce the same power and correlation values. This does not necessarily mean that they have the same correlation structure because the same product-sum can be produced by many combinations. Even for the contributions at a single frequency by all variables, there are three degrees of freedom. Infinitely many possible magnitudes could produce the same product. The solution for this is that the simulated auto-correlation functions i.e., magnitude spectra of all series should be the same or very close to those of the reference. All the generators presented here can do this to a very good extent. The curious reader might wonder that why the terms relating to the maximum possible power and correlation are introduced and not just the multivariate power and correlations. The answer to this question is again related to the degrees of freedom i.e., for more than two variables the ability to assign a sign (or direction) to the multivariate correlation is lost. The d-dimensional Spearman's correlation coefficients ρ_d and $\tilde{\rho}_d$ that were presented in Sec. 3.7, suffer from a similar problem in that they have no defined lower bounds and the least correlation tends to zero as the number of dimensions grow larger.

The question that now arises is that how to interpret the value of $MXCORR$? What does say a value of 0.75 mean? $MXCORR$ is akin to the upper Fréchet-Hoeffding bounds for a copula from Chapter 3 with a crucial difference. The upper Fréchet-Hoeffding bounds are for variables that are fully correlated but $MXCORR$ will yield a value that is likely to be lower than 1.0 for a set of variables. Suppose the following scenario of three discharge time series. Two series have a perfect $MXCORR$ of 1.0 but the third one has a $MXCORR$ of 0.5 with both of them. It could be that the third station has no power in half of its spectrum. The resulting $MXCORR$ value for the three of them will now be $2/3$. This means that the $MXCORR$ value is mainly controlled by series that have the absolute correlation closer to zeros with all of them. In reality, $MXCORR$ is brought down incrementally by all of the considered series where some will contribute more than the others to the decorrelation. Consider another example where 10 discharge time series are considered with one location always having the same value. This means that it has no variance and consequently $MXCORR$ becomes zero because the magnitude at all the frequencies is zero.

Finally, given that so many calculations are required to compute $MXCORR$, it is possible that the resulting spectrum has no meaning i.e., it has no significance. This can be easily tested for all combinations by comparing them against the $MXCORR$ of many randomized series. If the $MXCORR$ of the reference is consistently higher than that of the ensemble of the randomized series, then it can be stated with confidence that its value is significant.

7.3.2 Use of the maximum correlation

Using $MXCORR$ is not so straight forward to comprehend for a group of 10 time series as compared to only a pair of them due to the many possibilities of combinations. It could be that only 5 out of the 10 time series are correlated enough to bring their highest values together and not the others. While only relying on the $MXCORR$ of the 10, this cannot be ascertained. The more sensible method of achieving is that all the possible combinations ranging from all possible pairs to a multuplet of 10 have to be evaluated by only considering the ones that have correlations above a set threshold.

The total number of combinations that are possible for a group of 10 series are 1013. For 100 time series it is a, mind boggling, 1,267,650,600,228,229,437,251,415,720,939 $MXCORR$ values (an order of 30). It will be difficult to fit these many spectra in computer memory to begin with. To handle such a large number of possibilities, some assumptions can be made. From Geostatistics, it is known that values far away from each other are less related to the values closer to each other. For discharge, this may not be the case as a headwater gauge has a relationship with the down stream gauges, even if they are far away. Although, their correlation in the high frequencies will become smaller as the distance increases because of the low pass filtering applied by the river and the in-between catchment area. Hence, groups can be identified that are close to each other while ignoring the ones that are too far away. Next, the groups of any given size can be sorted based on their $MXCORR$ values. Maps of these combinations can then provide the information about groups of points that are likely to bring extremes together. This makes much more sense for precipitation than for discharge series. Another way of making groups could be by utilizing the pairwise $MXCORR$s. Suppose, the variables A, B, C have $MXCORR$s of 0.8, 0.85, 0.9 for the pairs AB, AC and BC respectively. Now, the trivariate $MXCORR$ of ABC can be calculated as well, its value will always be less that of $MXCORR$ AB because C has both the $MXCORR$s less than 1. As the value of $MXCORR$ is bound between 0 and 1, the result will either be no change or a decrease but never an increase in the final value by adding more variables. For practical cases, as the number of variables grows, $MXCORR$ tends to zero.

Another question that arises now is that of the $MXCORR$ value related to simultaneous extremes. A value of 0.6 is better than 0.5 but what does this objectively mean? Only those series that have higher correlation in the high frequencies are more likely to bring their extremes together. It could be that most of the correlation comes from the annual and seasonal cycles. These holds less significance as these contribute very little to the high values, and especially if they are abrupt.

Fortunately, it is not that important to know exactly what the value of $MXCORR$ should be for simultaneous extremes. This can be directly observed in the observed series anyhow. The important matter is that the simulated time series should have similar and unbiased $MXCORR$ values. If they are consistently too high, then more cases (most likely) of simultaneous extremes and vice versa for the consistently lower $MXCORR$ values are simulated.

It should be noted that, $MXCORR$ provides the information about the possibility of having extremes together but not guarantee it. Consider a scenario with two discharge time series. One is of a river that is mainly snowfed and another one that is rainfed. Assuming that

the mean flows are the same (it does not matter though as the first term is discarded in the Fourier transform) and the main precipitation season is in winter for both. For the snowfed river, peak flows will only occur in the Summer while for the rainfed they will only happen during the winter. In terms of the Fourier transform, they both will have very similar annual cycles but will be shifted by about 180^o. Computing the $MXCORR$ will show that their maximum possible correlation is rather high. But this would be misleading, as they never experience high flows at the same time. Furthermore, as was shown earlier, both of them will have a relatively small correlation in the higher frequencies resulting in very little probability of simultaneous extremes. More stress has to be put on the contribution of the high frequencies while making sure that the phases are mostly synchronized.

Consider another example where the likelihood of simultaneous extremes may be very large. Suppose two adjacent catchments in an area having the same climate. The annual and seasonal cycles are well synchronized along with much higher probability that any large-scale precipitation will happen over both of them resulting in higher chances of high flows together. More precisely, sudden peaks in flows require many high-frequency components to contribute at the given time step (i.e., come into phase). Which means that for the synchronization of peaks at multiple locations, they must have a higher correlation in the high-frequency range as well. This is what the simulations should also reflect.

All of the cases stated above are for observed cases and should come as no surprise. No fancy measures are needed to reach this conclusion. The reason why $MXCORR$ is used is to make sure that the simulations also follows a similar pattern in dynamics. It allows for the evaluation whether the properties of the simulations actually follow the relationship in space and time that was also observed. This tool can be used to test existing generators. One could also use this as an objective function during optimization. This would guarantee their usability for worst-case scenarios but is however not recommended due to the possibility of over-fitting. Another very important reason is that the copula asymmetries are a function of ρ_s. If reference and simulations are to be compared for asymmetries then it is imperative that their correlations match first. Without them matching, a comparison is meaningless.

To identify which pairs of catchments bring floods together, reference series where it is known that the extreme events are synchronized e.g., stations directly up- or down-stream of each other. To ascertain if two locations bring extreme events together, their empirical cross-copula can be evaluated. If the densities are skewed towards the top-right corner of the copula, i.e., towards $[1, 1]$, than this means that they are very likely to bring floods together. In terms of numbers, their order and directional cross-asymmetries (AO and AD respectively) can be calculated. A high positive AO and a very small absolute AD mean that the high-highs are more synchronized while a high negative AO and a very small absolute AD mean that their low-lows are more synchronized. If absolute AD is high and AO is small then that means that both low-lows and high-highs are not very likely to be synchronized. If both AO and AD are low then that could mean two things. First, if the Spearman's correlation (ρ_s) is high then the dependence of low-lows and high-highs is equal and also high. Second, if the ρ_s is low then the dependence of low-lows and high-highs is small. The use of the terms *small, high* and *likely* is on purpose. That is because their definitions are relative to some user defined reference.

The notion of the maximum correlation can be extended to the objective functions in the spectral domain defined in Chapter 6 for the multi-site case. These are introduced shortly.

7.4 d-dimensional cumulative power spectrum

Auto- and cross-$CMPOW$ for the bi-variate case were presented in Chapter 6. By changing the input variables in Eq. 7.11, the maximum possible correlation of that variable can be computed. Note that the magnitude spectrum (MAG) rather then the power spectrum (POW) is used. The details are as follows for a generic case of the variable X_M where the variable X has M columns/series.

Suppose a variable X, composed of N real values, whose forward real Fourier transform (FT) is given by,

$$FT_X(k) = \sum_{i=0}^{N-1} x(i) exp\left(\frac{-2\pi jik}{N}\right) \quad \text{for } k = 0, \ldots, N \tag{7.12}$$

where, k is the frequency, $x(i)$ is the ith value of X, and j is $\sqrt{-1}$ with each term in equation being a complex number. The magnitude spectrum (MAG) of the series X can be written as,

$$MAG_X(k) = |FT_X(k)| \quad \text{for } k = 1, \ldots, A \tag{7.13}$$

$$A = \begin{cases} \frac{N}{2}, & \text{if } N \text{ is even} \\ \frac{N+1}{2} - 1, & \text{if } N \text{ is odd} \end{cases} \tag{7.14}$$

The d-dimensional cumulative power spectrum ($CMPOW^d$), can then be defined as follows,

$$CMPOW_X^d(k) = \frac{\sum_{k=1}^{k}\left(\prod_{m=0}^{M-1} MAG_X^m(k)\right)}{\sqrt[M]{\prod_{m=0}^{M-1}\left(\sum_{k=1}^{A}\left(MAG_X^m(k)\right)^M\right)}} \quad \text{for } k = 1, \ldots, A \tag{7.15}$$

7.5 d-dimensional marginal cumulative periodogram

By using the time series of the marginals as input in Eq. 7.15 the maximum correlation among any number of variables can be calculated as a periodogram. The time series used as input can be multuplets of precipitation from multiple locations or a combination of various variables at the same location. Care must be taken though. Suppose, temperature

and precipitation at multiple locations are provided. Calculating the $CMPOW_X^d$ for all of these makes little sense as precipitation and temperature have negligible correlation in general as temperature has a very strong annual cycle magnitude while for precipitation it is much smaller and most of the power is evenly distributed in the high frequencies. It will yield a very small value because any variable that has a small correlation will bring down $MXCORR$ with it, even if the others have a significant dependence. What could be done is in addition to the $CMPOW_X^d$ of all variables combined, it can also be calculated for precipitation only for all locations along with the same for temperature. Automatic detection of variables belonging to the same group can be done by an unsupervised classification algorithm where the number of groups have to be at least the number of the variables used. This yields three correlations or spectra which in turn provides more information albeit at the price of more computational time. Matching or comparing these three spectra for the reference and simulation gives allows for a better judgment for any significant bias in dependence. Maximum correlation is denoted by $MXCORR_p$, where the subscript p stands for the Pearson's correlation. Similarly, $CMPOW_p^d$ is the d-dimensional marginal cumulative periodogram.

Fig. 7.3 and 7.4 show two cases where the former has series that are close to each other in space while the latter has series that are farther from each other. In the background are simulations of the same variables using Gaussian dependence (AAFT) only. The final values of the spectra are the $MXCORR$s of the series. An immediate difference can be realized. For the nearby points' case, the spectra are close to each other with a slight difference while the $MXCORR$ values match the observed case very well. The latter farther points' case shows a significant difference. Finally, considering a spectrum as a diagnostic measure shows that it is superior to using a single value as it provides a more complete picture of the temporal scales of bias.

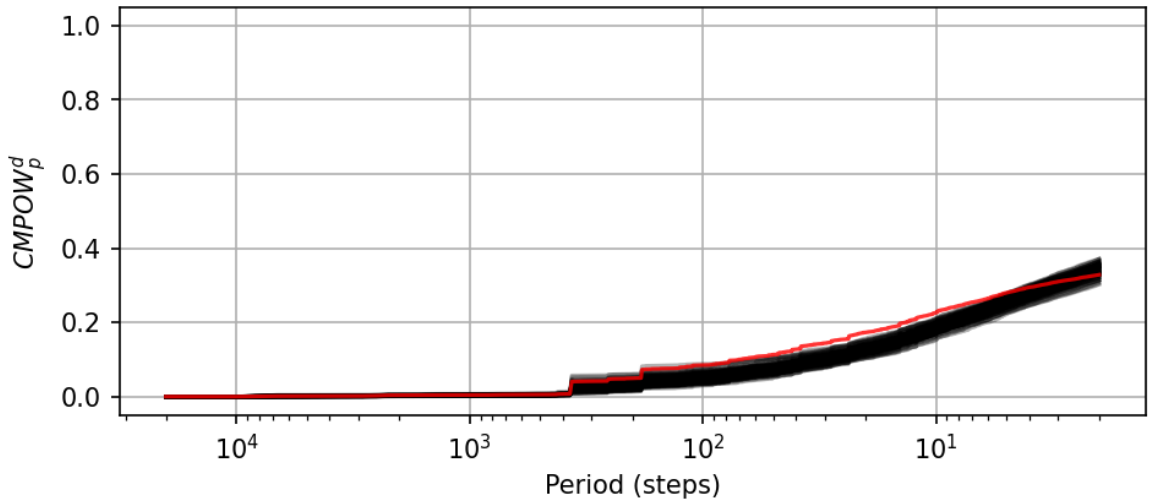


Figure 7.3: $CMPOW_p^d$ for five precipitation stations with daily records for the period of 1961-2015 in BW that are near to each other. In red are observed series while in black are series with Gaussian dependence (AAFT).

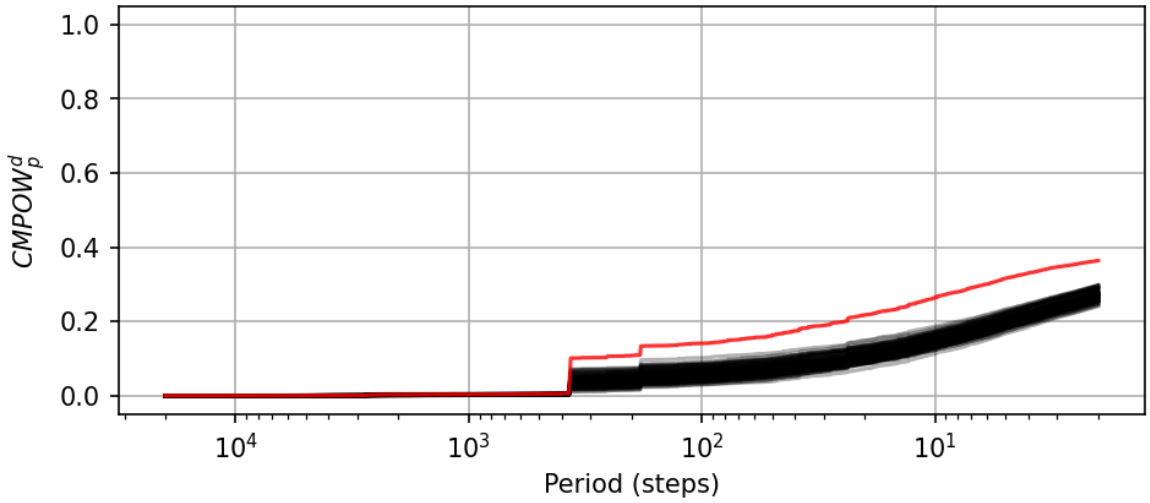


Figure 7.4: $CMPOW_p^d$ for five precipitation stations with daily records for the period of 1961-2015 in BW that are far away from each other. In red are observed series while in black are series with Gaussian dependence (AAFT).

7.6 d-dimensional grade cumulative periodogram

Similar to $CMPOW_p^d$, using the time series of grades yields the maximum correlation of grades in d-dimensions. It is denoted by $MXCORR_s$, where the subscript s stands for the Spearman's correlation and $CMPOW_s^d$ is the d-dimensional grade cumulative periodogram.

Fig. 7.5 and 7.6 show examples of the $CMPOW_s^d$ for the same settings as those for $CMPOW_p^d$. An even larger problem can be observed. The background simulations were performed using the Amplitude Adjusted Fourier Transform (AAFT) algorithm where the magnitude spectrum of the marginals was used and the grades were not considered. The resulting copula of the simulations has properties drastically different than the reference, as evidenced by the figures. The strength of the resulting $MXCORR$ is reduced by one half for the grades. This also becomes an issue while comparing d-asymmetries of reference and simulations. An interesting aspect to notice is that the simulations have significantly larger magnitudes of the annual cycle, which are not present for the reference. Also, the simulations are missing one of the long-term signal strength which is present for the reference. This could be an indicator of non-linearity i.e., long-term trend in this case, as was pointed out by Nakamura et al. (2006).

7.7 d-dimensional copula order asymmetry cumulative periodogram

In Sec. 6.3, only the pairwise power spectra of asymmetries were shown. The d-dimensional version is presented here by utilizing the new formulas developed in Sec. 7.2.1. d-

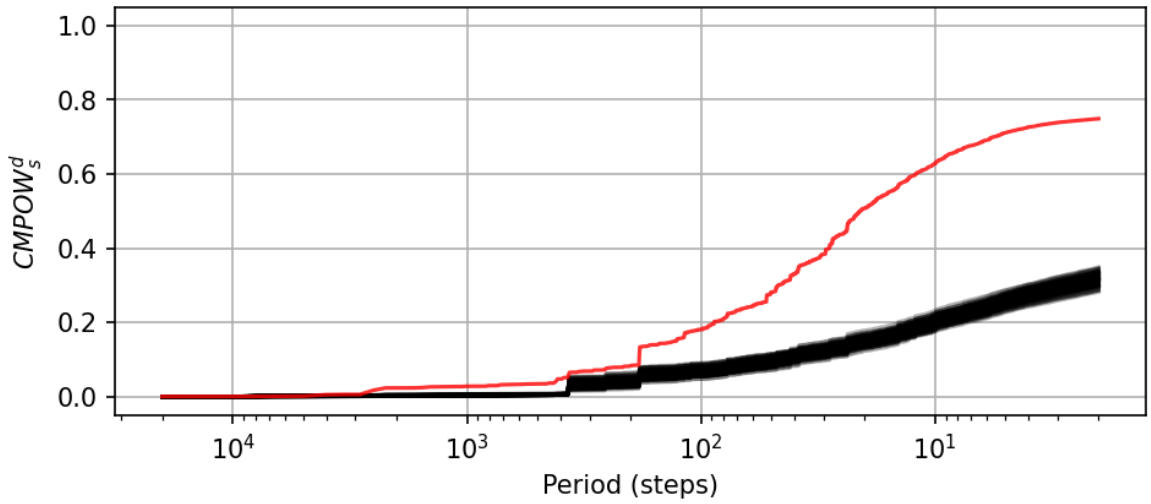


Figure 7.5: $CMPOW_s^d$ for five precipitation stations with daily records for the period of 1961-2015 in BW that are near to each other. In red are observed series while in black are series with Gaussian dependence (AAFT).

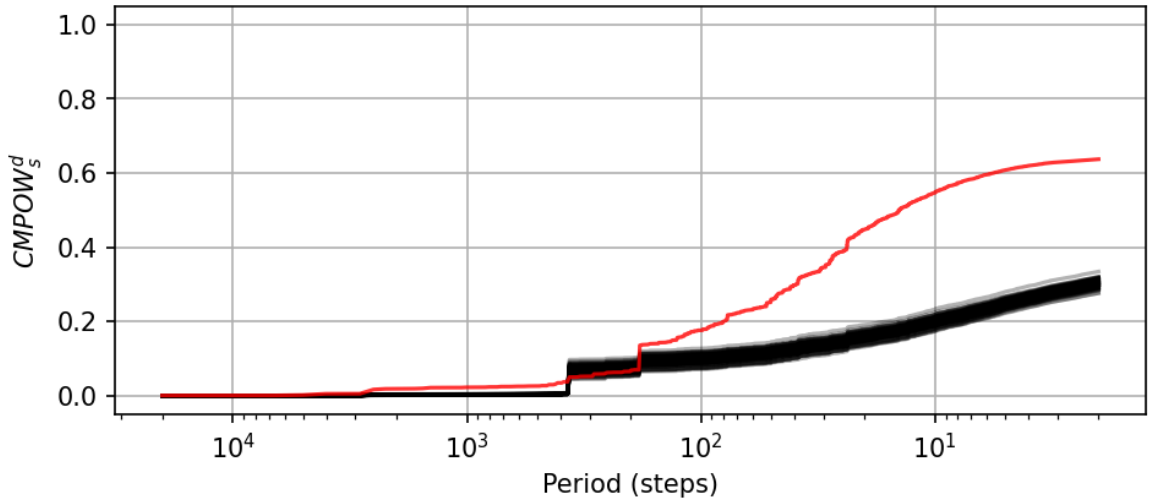


Figure 7.6: $CMPOW_s^d$ for five precipitation stations with daily records for the period of 1961-2015 in BW that are far away from each other. In red are observed series while in black are series with Gaussian dependence (AAFT).

dimensional vectors are used instead of a single value per time step. Hence, instead of taking the mean for any component all the values for each time step are used and these are treated as a multivariate time series. Also, here the notion of lag- or lead-steps does not exist as the asymmetries are only computed for the cross-case. Lagged or leaded time series are not discussed as they were already handled in Chapter 6 for various cases. Although, this could also be done for the auto-case which may be an even a more elegant solution, albeit at the cost of no normalization. Finally, when the values at each component for any order asymmetry in d dimensions are used as input in Eq. 7.15, the d-dimensional copula order asymmetry cumulative periodogram is obtained and is denoted by $CMPOW^d_{AO}$. Note that there are a total of d number of $CMPOW^d_{AO}$s except for the two-dimensional case where it is only one.

An important question to pose here is that what do these time series and their periodogram mean? This can be explained by tracing the path of points in each component in time. As a weather variable is bound to the atmosphere, it has to follow roughly the same path every year and must behave in a physical manner. For a simulated time series, it could very well be that they do not follow the same behavior due to any number of reasons. This will result in a mismatch of the two spectra, which would signify that the dynamics simulated by the generator are not realistic.

Fig. 7.7 and 7.8 demonstrate d-asymmetries for the same setting as that for $CMPOW^d_p$ and $CMPOW^d_s$. It was shown that the $MXCORR$s of the simulated series are not similar to those of the reference, so, a direct comparison does not make sense for the absolute values. Nevertheless, the shape could be compared. Looking at both of them, it can be observed that, on average, the maximum value attained by the spectra are greater for the nearby points as compared to the points that are farther apart.

7.8 d-dimensional copula directional asymmetry cumulative periodogram

Similarly, when taking the time series of components while computing the directional d-asymmetry described in Sec. 7.2.2, the d-dimensional copula directional asymmetry cumulative periodogram can be obtained that is denoted by $CMPOW^d_{AD}$. It is different then the order asymmetries because it contains the distances from the major axis only. Fig. 7.7 and 7.8 show examples.

7.9 Worst-case scenario multi-site time series

It was previously mentioned that the $MXCORR$ represents the best- or worst-case scenario depending on context. For clarification some details are provided here that show how such series may be constructed. As was shown that $MXCORR$ is constructed by using magnitude spectra of multiple time series, which is nothing more than computing the correlation of two or more series without considering their phase differences while keeping their auto-correlation functions. To create a multi-site time series, the same phase spectra for all series

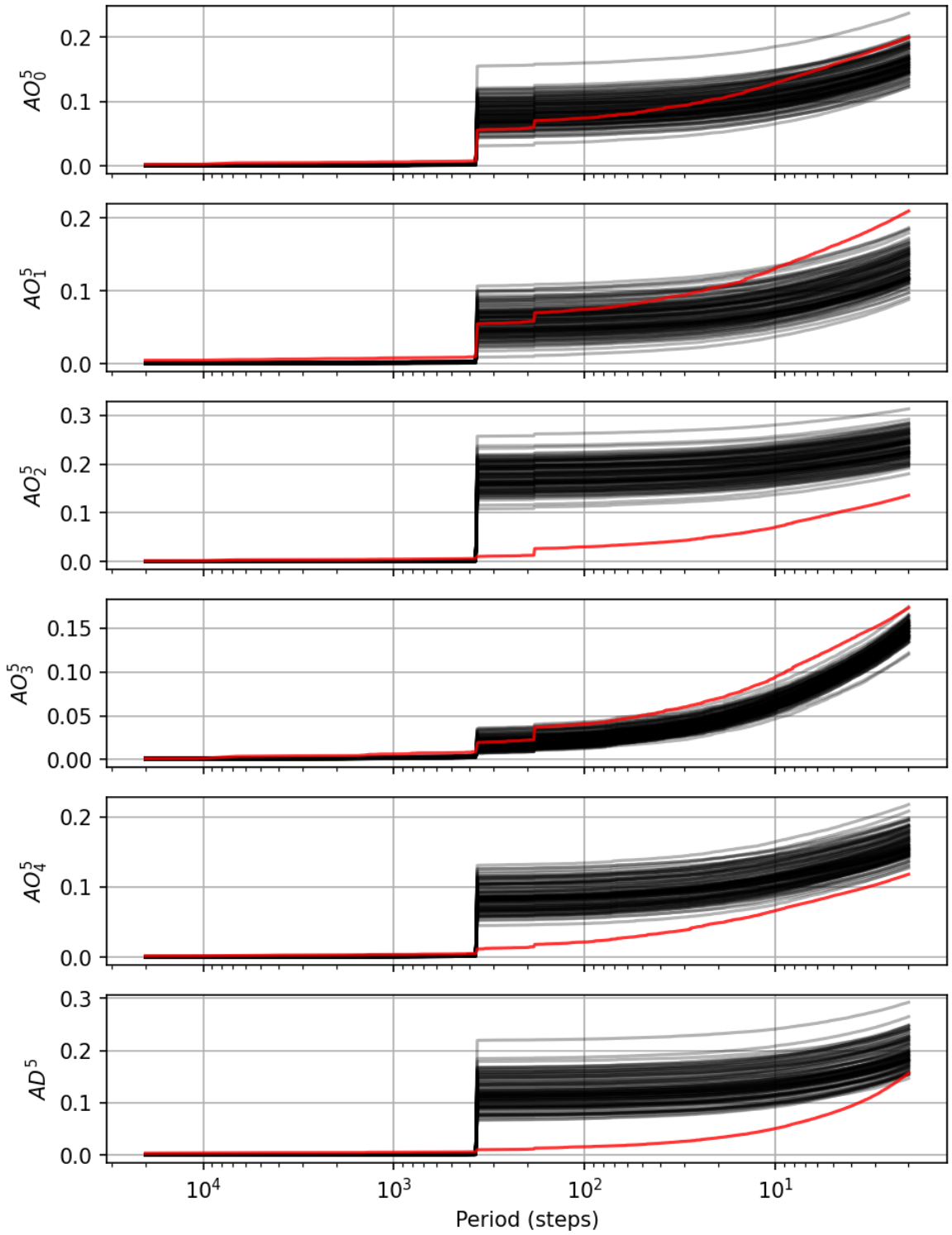


Figure 7.7: $CMPOW_{AO}^{d}$ (top five) and $CMPOW_{AD}^{d}$ (bottom) for five precipitation stations with daily records for the period of 1961-2015 in BW that are near to each other. In red are observed series while in black are series with Gaussian dependence (AAFT).

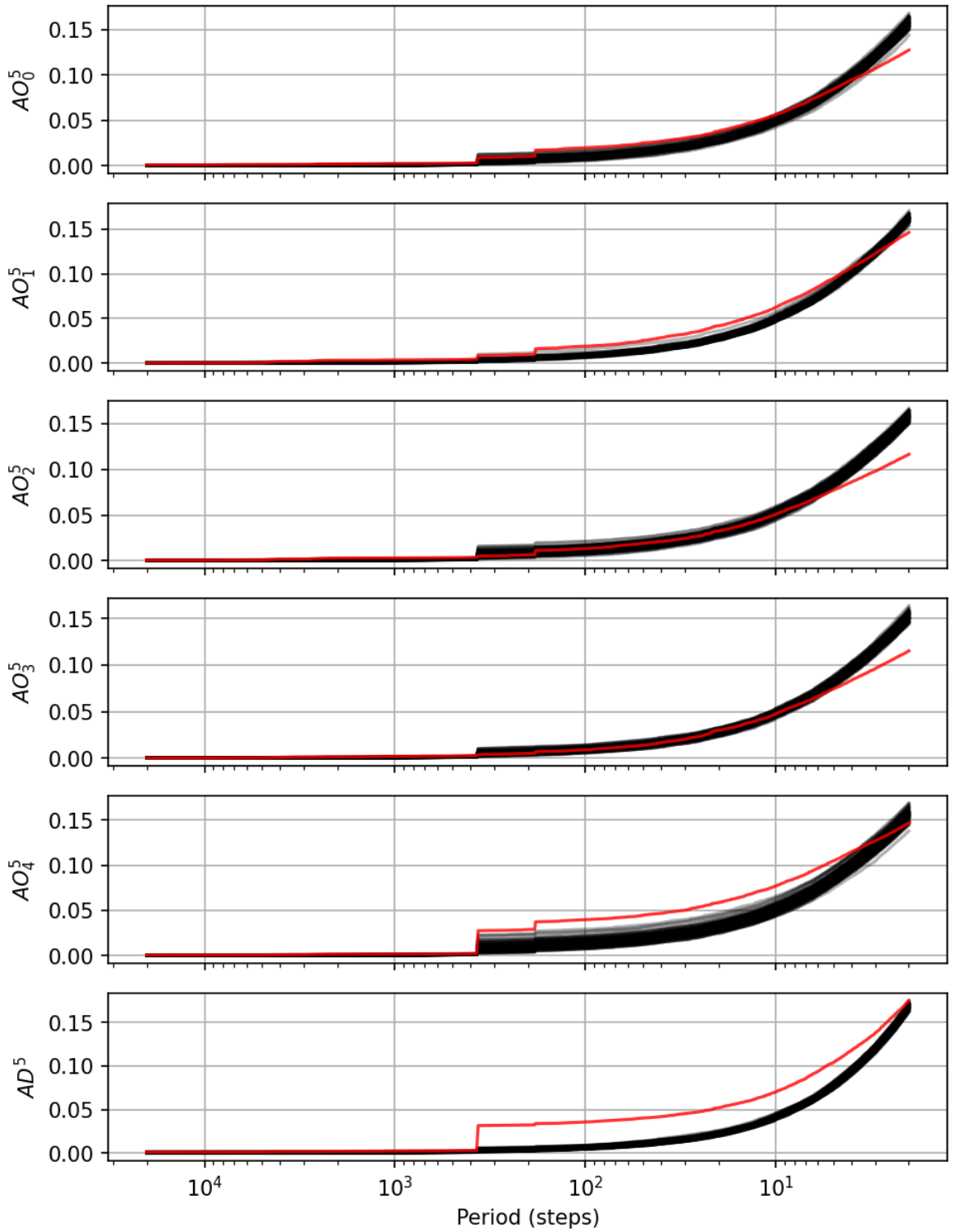


Figure 7.8: $CMPOW_{AO}^{d}$ (top five) and $CMPOW_{AD}^{d}$ (bottom) for five precipitation stations with daily records for the period of 1961-2015 in BW that are far away from each other. In red are observed series while in black are series with Gaussian dependence (AAFT).

can be taken and used as input for the AAFT or any suitable algorithm to create new series for each site. It may very well be that all the phase spectra are not equal but the chances are very good that all the series have now fully synchronized simultaneous extreme values. Note that this is not the same as using sorted time series as the sorted time series have auto-correlation functions different than those of the reference. This case can also be compared additionally to the reference and all the simulations. Ideally, the maximum spatial-sum of these series should be the largest among all the series. It should be noted that the final time series properties are somewhat dependent on the first shuffle. Hence, to get the worst-case scenario, many such shuffles can be used and the one that yields the auto-correlation functions closest to the reference can be selected as the final series. Interestingly, series that would produce the extremes floods together are also very likely to have regions in time where many low values take place consecutively. If the highs are grouped together then the lows have to be as well. For simulated series, a test can be written to evaluate how the lengths of low-flow durations compare to those of the reference. However, simultaneous extremes tend to happen for much shorter times than droughts due the exponential-like distribution of precipitation and consequently discharge, other frequencies may be more responsible for such behavior. These could be many small frequencies that synchronize in such a manner that all produce a combined low, and due to their slow rise and fall times, persist in time to produce droughts.

7.10 d-dimensional correlation based on mean phase

So far, the phase spectrum was left out of the discussion which is the other important half. As the magnitude spectrum shows the strength of the signal at a given frequency, the phase determines the location of the signal start in time. Hidden in the phase spectrum is the asymmetric behavior i.e., non-Gaussian dependence. To define the d-dimensional correlation, the cross-correlation in the frequency domain for pair of series is presented first. The cross-power spectrum between two series X and Y is given by,

$$POW_{XY}(k) = MAG_X(k) \cdot MAG_Y(k) \cdot \cos\Big(\angle FT_X(k) - \angle FT_Y(k)\Big) \quad \text{for } k = 1, \ldots, A$$

The summation of all $POW_{XY}(k)$ values yields the total cross-power (assuming a mean of zero). By normalizing it properly, the correlation in the frequency domain (up to a given frequency) is obtained by the following equation,

$$CORR_{XY}(k) = \frac{\sum_{k=1}^{k} POW_{XY}(k)}{\sqrt{\sum_{k=1}^{A}\Big(MAG_X(k) \cdot MAG_Y(k)\Big)}} \quad \text{for } k = 1, \ldots, A$$

The individual contribution at or up to each frequency for any two series using the above equation can be calculated. For example, the contribution of the annual cycle to the total

power or the contribution of frequencies between any two given thresholds. Another important aspect can be the assessment of long term trends in the data. These, if they exist, can be very clearly seen by the relative contribution of frequencies whose periods are longer than one year time for environmental variables (Nakamura et al., 2006). At any given frequency, the contribution is related to the difference of the phases, the lesser the difference the more the contribution. For a difference of π radians, a negative contribution is produced. Two positively related series will have, in general, phases that have little differences while for negatively related series, the phases will have opposite directions. For series with no linear dependence, the phases will differ by half π radians or mixed in such a way that the final power/correlation becomes zero or close to it.

Now to answer the question, *What is the correlation among M number of series?* As was stated earlier, the direction of correlation (or dependence) is not defined for dimensions greater than two. Hence, only the positive direction and the overall strength of correlation can be determined. Given M number of series, the number of groups of two without replacement are given by,

$$\binom{M}{2} = {}^{M}C_2 = \frac{M!}{2(M-2)!}$$

For three series, three distinct pairwise correlations are possible. Suppose, two series have a perfect positive correlation while the third one has a perfect negative correlation with both of them. The mean of these is -0.33. There are infinitely many ways to produce a mean correlation with the same values. It is difficult to get a mean of zero as long as any two are related to each other, regardless of the direction. It was shown that the contribution at each frequency to the power between two series is dependent on their phase difference. Interestingly, the cosine of the difference also swings between -1 and $+1$, same as ρ_p. Hence, the mean of the cosines of the phase difference for M (always greater than two) series can be calculated. This is termed as the *mean phase correlation* (MNPHS) which is computed as follows,

$$MNPHS(k) = \left| \frac{1}{{}^{M}C_2} \sum_{i} \cos\left(\angle FT_X(k) - \angle FT_Y(k) \right)_i \right| \tag{7.16}$$

where, i represents the index of each possible combination pair X and Y without replacement for M variables. As mentioned before, taking the absolute is necessary because even with a negative sign the contribution from some series is still positive. Similarly, the mean power can be defined by combining $CMPOW^d$ and $MNPHS$ to produce the mean power periodogram (MNPOW). It can be written as follows,

$$MNPOW(k) = MNPHS(k) \cdot \prod_{m=0}^{M-1} MAG_X^m(k) \tag{7.17}$$

Finally, the mean d-dimensional cumulative mean correlation periodogram (CMCORR) can be written as follows,

$$CMCORR(k) = \frac{\sum_{k=1}^{k} MNPOW(k)}{\sqrt[M]{\prod_{m=0}^{M-1}\left(\sum_{k=1}^{A}\left(MAG_X^m(k)\right)^M\right)}} \quad \text{for } k = 1, \ldots, A \tag{7.18}$$

It should be noted, that this is not a true d-dimensional correlation (if one is even possible) because it still depends on the phase differences that are computed for pairs. Still, this can be used to show the overall behavior of M series. Similar to previous sections, depending on inputs the marginal, grade, order-asymmetry and directional-asymmetry cumulative spectra can be calculated. Contrary to the maximum correlation, these would show the actual behavior of the input series in time. For the various cases of inputs, these are denoted by $CMCORR_p^d$, $CMCORR_s^d$, $CMCORR_{AO}^d$ and $CMCORR_{AD}^d$ and are the d-dimensional cumulative marginal, d-dimensional cumulative grade, d-dimensional cumulative order asymmetry and d-dimensional cumulative directional asymmetry correlation periodograms respectively.

7.10.1 Interpretation and usage of the mean correlation

$CMCORR^d$ does have the properties that required from a dependence measure with one deficiency i.e., it swings between 0 and 1 i.e., no direction. For the case when all phases are out of sync, a value of zero (or close to it) is obtained. When all are in sync, a one i.e., $MXCORR$ is produced. One use of $CMCORR^d$ is that of determining the contribution of high frequencies to the overall dependence. For a point time series such as discharge, to create a sudden peak, many phases have to synchronize such that their magnitudes combined produce the said peak. There is no way around this. Long term frequencies such as the seasonal and annual cycles have less to do with it as they rise and fall too slowly, they do affect the regions in time that are likely to have high values though. In order for the simultaneous extremes to take place, the short term frequencies must also synchronize for all the locations.

The above explanation for simultaneous extremes is good for the simple case where only one peak occurs in the entire time series. In reality it could be much more different as there are many such peaks scattered all over time in a somewhat random fashion. Then, does the explanation of the high frequencies' synchronization hold? The answer is *partly*. As there are many high frequencies as compared to the short ones, in order for a peak to occur only a subset of these have to synchronize and not all of them necessarily. This also means that some frequencies with somewhat longer periods may also come into play. For this reason, the very short ones cannot always be considered, say a day or a week but rather up to a month. This will also become evident in the chapter of Results.

One could argue that the behavior of phase synchronization can be better represented by

$MNPHS$. This could be the case in exceptional circumstances but consider that the phase synchronization is only half the story, the actual contribution comes from the magnitude. Hence, it could be that frequencies close to each other having different magnitudes, produce very different results. In general, series having more contribution in the short frequencies are more likely to have simultaneous extremes as compared to the ones that have lesser contribution. This can also be used to sort groups of series that have very similar properties in time.

One final point to keep in mind is that of the difference between taking the mean of correlations and the mean of cosines of phase differences. For the case of three series with two perfectly correlated while one having a negative correlation to both, the mean ρ_p is -0.33. In case of the cosines, say two phases are fully in sync and the other has the opposite direction, a mean of 0.33 is produced, which makes more sense as having two correlated series have a greater weight. More importantly, when any of the phases differ by half π, a null contribution at that frequency is produced.

7.10.2 Combined interpretation and usage of maximum and mean phase correlation

The maximum and mean correlation spectra provide very useful information. The former tells shows the potential effect that multiple series have, in a combined manner, while the later shows how it actually is. The difference of these two holds the answer to the mismatch in timing at each frequency. For example, if the difference between the two is large (with a high value of the maximum correlation) then it can be concluded that the dynamics at various points are very similar but potentially shifted in time i.e., asymmetric. Such information is important when working with fine temporal resolution data such as hourly or sub-hourly scales. The simulations can then be tested to check whether they also exhibit such a relation or not. For simultaneous extremes, the focus is more on the synchronization of the high frequency components, as they are the responsible ones. It should be kept in mind, while comparing, that the gradients of these spectra (reference and simulation) have to be similar and should not have any constant offset that might be there due to the mismatch in low frequency waves. A consistent divergence in the high frequency range signifies that the simultaneous extremes of reference and simulation are not similar. The simultaneous extremes are more frequent for the case when the $CMCORR^d$ diverges to a value greater than that for the reference while it is vice-versa for the case when it goes below.

7.11 Distribution of the spatial-sums

It is not disputed that the extreme floods in rivers are a direct consequence of large precipitation over an area for some given time interval. Distributions of the sums of all values at a given time step can be compared for the simulations with the reference. These should show a bias immediately, if one exists. Such distributions will be shown in the chapter of Results. Care should be taken while interpreting these. Large floods are results of continuous precipitation for long times. As the distributions do not hold any temporal structure, it should

not be seen as a decisive factor. However, on average the distributions of the simulations should be similar to those of the reference i.e., no significant bias. Another idea could be to look at the cumulative time series of the 1st-order differences or the distribution of the 1st-order differences. The magnitudes of the peaks and valleys in these series should be similar for the reference and simulations. The drawback of this is that it can be only done for the single-site case. Such distributions will be shown in the chapter of Results for the reference and simulations. Additionally, two more distributions can be plotted along side. One of the worst-case scenario and the other of the sorted time series' spatial-sums. These should also put things in perspective as to how the correlation function in space is like.

7.12 Conditional simultaneous upper-tail probabilities

Similar to the distribution of spatial-sums, another type of aggregate test on them can be defined to assess whether the simulations follow a behavior similar to that of the reference. Suppose, the case of 5 observed discharge time series. For these, the probabilities of cases where a given combination of 4 series have tail values all above the return period of 10 years or more. Similarly, the probability of cases where only 3 out of the four had all upper-tail values of the return period of 10 years or more can be determined. Comparing these for all the possible cases of combinations of 4 out of the 5 series will produce a set of scatter points where the values on the horizontal axis are the probabilities of at least three of the 4 having simultaneous 10 year or higher events while on the vertical axis are corresponding probabilities where all 4 have 10 year events. Such scatters will be shown in the chapter of Results for the reference and simulations.

7.13 Testing 3D interpolations/simulations for correct dependence structure

Interestingly, all the d-dimensional measures described till now can also be used to evaluate variables that are interpolated in space-time. Consider the example of temperature. A regular grid of 50x50 $units^2$ has to be interpolated at 100 time steps. Out of the 2500 cells, 10 are control points that are used as input for interpolating at the remaining 2490 points. The method of Ordinary Kriging is used to interpolate all the points for all the time steps. Now the question is how the higher-order dependence of points can be evaluated? The tools presented here can be used in the following manner to answer the question. For the control points, draw time series of combinations of 4 and compute any (or all) d-dimensional properties. Now, draw random points from the grid that are not the control points in groups of 4, and as many as possible. Care should be taken that the distances between the points have the same distribution of distances as those of the control points. Compute the same d-dimensional properties. Finally, compare the two for each measure, say the distributions of $MXCORR$ for the control and interpolation points. Ideally, these two distributions should not show any significant difference. Similarly, any simulation method can be tested in the same manner, only the number of fields that are interpolated are more than one which necessitates a comparison of many more distributions.

8 Persistence of Spatio-temporal Properties

Before the results of the various approaches to generate multivariate time series and tools to diagnose correct behavior are presented, the persistence of some of the properties in space and time are discussed and evaluated for observed series. The term *persistence* refers to the *constantness* of properties in space-time in this thesis. The aim here is to show how these evolve as distinct time periods are considered, how length of the selected periods affects the persistence and how the dependence among points changes as the distance between them increases or decreases. What is being explained here is a phenomenon that is commonly used when dealing with a system formed by interaction of multiples variables between/among each other. The concept is called *Ergodicity*, which in the context of this thesis refers to the fact that if a time series is long enough, then the properties of the underlying process can be deduced from it. For this thesis, the properties would be those that were defined/discussed in the previous chapters e.g., Pearson's correlation (ρ_p), Spearman's correlation (ρ_s), order asymmetry (AO), directional asymmetry (AD), copula entropy (H) etc. Another purpose of this chapter is to demonstrate how changes in time series' behavior are linked to the shape of various measures in space-time. The measures point out that some aspect of the time series have changed but they cannot explain why exactly.

It is assumed that for a property to be considered as an objective function in a time series generator, it should be stable enough in time and space. If it is not, then that implies that simulated time series are not very representative of the generating process behind the original one and that they are an over-fit. It makes sense to impose such constraints on considered properties for simulation because time series generators are unaware of the generating process. They only mimic a subset of its properties. In general, climate variables exhibit persistent behavior over long periods of time or have little trend inside them as they are influenced by processes that take place on a global scale. They can also respond in a drastic manner to, say, an asteroid impact. Ideally, the simulated time series should show the same type of variability in space-time as that of the reference.

It is should be noted that simulated time series using the methods developed in this thesis have properties that are similar to those of the reference time series. As the earth's climate is permanently evolving, the problem that needs to be addressed next is that of including information in the simulation process that takes such effects into account. As far as true Ergodicity is concerned, it cannot be disputed that time series should be generated by accounting for the possible effects of climate change in the simulation process. Nevertheless, this does not render the new methods useless as it will be demonstrated that the variability of various measures of the produced time series show similar patterns as those of the observed to a large extent. Furthermore, the evaluation of the dependence structure of time series using the newly developed metrics will show the evolution of properties in time if

any exists. It should also be noted that effects of climate change can be easily evaluated and incorporated into the generation of precipitation statistically as it is only related to the atmosphere but doing the same for discharge is very difficult because changes in discharge are due to changes in precipitation, river bed and land use or land cover etc. It is unlikely that the information about the former are recorded for such long periods over study areas.

Finally, the following settings are used to demonstrate the persistence. The considered time period of the data is from 1961 to 2015 at a daily scale. River discharge and point precipitation observed data in Neckar River catchment are used as the demonstration variables. Various properties are computed for time periods of 1 and 26 years. In some cases, 5 and 10 year periods are also chosen for demonstration of the evolving behavior in time. Color codes are used to identify a given year or a window of years wherever deemed necessary. Hourly time series are not analyzed as they were too short in their lengths compared to the daily series.

8.1 Auto-properties

Considering the daily discharge of the sub-catchment Enz in the study area for various time windows, Fig. 8.1, 8.2 and 8.3 show the auto-properties of the time series for the time periods of 1, 10 and 26 years respectively. The y-axes limits are kept same for both to show the relative reduction in variance of properties over the longer period of time. ρ_s and H show a relatively higher stability over time even at the 1 year time window. In comparison, AO, AD and to some extent ρ_p show a much larger variability. An interesting thing to note here about AO is that it varies much more for smaller lead steps as compared to the other properties. More precisely, it shows that at smaller lead steps the relative densities in the copulas of high-highs and low-lows can fluctuate to a very large degree from year to year but as the number of lead steps grow larger, it becomes more negative on average up to a certain limit (about a month here). This in turn means that low flows periods in this time period persist much longer than high flow periods, which makes sense as high values do not persist as long as low values in time at a given point. Furthermore, AD starts with negative values consistently even for the 1 year window case. Signaling that low-highs occur more often than high-lows. As the number of lead steps grows larger, this distinction disappears to a certain extent but the values of AD do fluctuate around -0.2. H shows a much more consistent behavior for the larger lead steps comparatively, especially looking at the 1 year window case. Its slope plateaus around the 10 lead step mark, after which it changes rather slowly but never touches 1 i.e., the case of full independence. This is an important aspect of discharge here i.e., the memory of the sub-catchment is quiet long i.e., it can hold a large amount of water and release it very slowly. Both ρ_p and ρ_s do not reach a value of zero even in one month but H does reach an almost constant value. This could be an indication that after a lead steps of about 10 days, no meaningful information can be extracted from the data. Another very important aspect that should be kept in view is that the value of H is copula bin size dependent. A finer resolution copula will convey different information than a coarser one. Concerning the evolution of dependence in time, Fig. 8.2 (10 year windows) shows that there is indeed some evolution in time when moving from one decade to the

next. For example, ρ_s shows a decrease of strength in the smaller lags (from 1 day to two weeks). ρ_p seems to be very similar for all the decades apart from 1961-1970 which has unusually high strength. AO shows no consistent pattern but a definite transition for AD can be observed in that it is decreasing as the time goes by. Similar to ρ_s, H also shows a decrease in dependence as the time continues to pass. Fig. 8.3 shows a binary view of the relative dependence strength between the past (1961-1986) and recent (1987-2012) years. Here again, we see a relative decrease in the strength of all the metrics in the recent years. Exact cause(s) of this behavior needs investigation which is out of the scope of this thesis.

For lead steps of 1, 7, 14, 28 days Fig. 8.4 and 8.5 show the Cumulative power spectrum of auto-order asymmetry ($CMPOW_{AO\_AUTO}$). Again, a considerable variability for the 1 year window case can be observed. The contribution of the annual cycle of $AO(t)$ varies to a large degree. The strengths of the rest of the signals do not show the same behavior. This makes sense as the low flow seasons are time dependent and repeat every year at, more or less, the same time. For the 26 years window case, the two curves are very similar except for the strength of the long term cycles (greater than 1 year). This could be a signal of a trend in one of the time series where it is stronger. Ideally, the strength of these long term signals should be zero or very little at least. However, they cannot be exactly zero as that would mean that the weather repeats itself every year. Their increasing relative strength points to non-stationarity, which is generally undesired when working with time series but it cannot be avoided in observed time series. Such an information draws attention to the fact that either the climate or the catchment state (say land-use) are changing in time. This means that the rainfall-runoff modelers may have to take into account those model parameters that control evapotranspiration and update them accordingly. Another reason for these long term signals could be that the timing of low flows have shifted in time either forwards or backwards. For the past time period (1961-1986) it can be seen that that strength of the cycle at the period of ca. 1 decade is much larger than that of the recent period (1987-2012) and the strength of the annual and the 5-year cycle of the recent time period is higher than that of the past. This signifies that potential synchronization of the shorter frequencies have become larger in the recent time which in turn means that the periods of droughts and floods have shorter durations but occur relatively frequently. To assess whether this behavior is significant in space-time, the same test could be done for nearby rivers. Finally, Fig. 8.6 and 8.7 show the Cumulative power spectrum of auto-directional asymmetry ($CMPOW_{AD\_AUTO}$). These show a completely different behavior than their counterpart. Here, a larger variability contributed by the higher frequencies can be observed. This makes sense as AD is more sensitive to high-lows and low-highs, which are rather abrupt in nature compared to the smoother low-lows and high-highs. For the lead step of 1, the contributions of high frequencies can be observed to be significantly larger compared to the case of say 28 lead steps where the greater contribution to variability is by the frequencies having periods of 10 to 100 time steps. Unlike $CMPOW_{AO\_AUTO}$, $CMPOW_{AD\_AUTO}$ shows a change in shape as the number of lead steps increases. A very small annual cycle can be seen as the number of lead steps increases and almost no long term frequencies when observing the 26 year window curves. Another important matter to note is that recent times (1987-2012) show relatively more strength in the very high frequencies (10 days and lesser) and less in the longer ones (10 to 100 days). More strength in higher frequencies is linked to fast rising and recess-

ing peaks. This could be due to climate change as the number of snowy days decrease, more precipitation reaches the streams instead of being slowly released via snow melt in winter i.e., the time series are less smooth in recent times or at least they have more potential to be so. This can be further investigated in the future.

From all the figures, it can be concluded that as the length of the time period used to compute a given property increases, so do the properties become more similar for the various windows. Most of the measures become very similar for the longest period of 26 years but do not match exactly. This makes sense as the catchments and climate are undergoing constant change and the weather never repeats itself exactly. The same can be said for the state of a catchment.

8.2 Cross-properties

This thesis is about the spatial aspects of extremes, hence a more detailed view is shown as compared to the auto-case. For daily discharge, a total of five headwater sub-catchments are used to demonstrate the spatial variability. These gauges are Kirchentellinsfurt (Neckar), Vaihingen (Enz), Untergriesheim (Jagst), Stein (Kocher) and Plochingen (Fils) (note the order). Fig. 8.8, 8.9 and 8.10 show the d-asymmetries for the time windows of 1, 10 and 26 years respectively. d-dimensional order asymmetry (AO_j^d) show no significant difference in their relative behavior because the j-th component of each AO_j^d shows a very similar spread. Only the decade of 1971-1980 in the Fig. 8.9 shows behavior that is out of place as compared to the rest. The cause(s) for this needs further investigation. It should be kept in view that in higher dimensions, the asymmetries do not lend themselves to easy interpretation. Only their relative behavior to each other can be evaluated. d-dimensional directional asymmetry (AD^d) provides information about the relative timing of the low-highs and high-lows in various time series. In other words, which catchment reacts the most abruptly or the fastest and which the slowest. Plochingen (Fils) (at index 4) has the overall largest positive values followed by Stein (Kocher) (at index 3), Kirchentellinsfurt (Neckar) (at index 0), Untergriesheim (Jagst) (at index 2) and then Vaihingen (Enz) (at index 1). Relating this to the map of the catchment (Fig. 4.2), it could be stated that discharge drivers (i.e., precipitation) enter from the South-East of the catchment which leads to such a shift in timing of the events. Furthermore, Plochingen (Fils) has an area of about 700 km^2, which is considerably smaller than the rest. This small size also gives it the lead in the reaction times to precipitation events.

Fig. 8.11 shows the d-dimensional marginal cumulative periodogram ($CMPOW_p^d$) while Fig. 8.12 shows the d-dimensional grade cumulative periodogram ($CMPOW_s^d$) for various time windows. Both the multivariate periodograms show that the combined variability is mostly dominated by the annual cycle in the long term and approach equal values as the time windows grow in size. For the smaller windows, there appears to be years where the peaks of the stations were potentially likely to be synchronized. Fig. 8.13 and 8.14 show the d-dimensional copula order asymmetry cumulative periodogram ($CMPOW_{AO}^d$) and d-dimensional copula directional asymmetry cumulative periodogram ($CMPOW_{AD}^d$). Here, a much larger difference between the spectra can be observed. $CMPOW_{AO}^d$s show practically

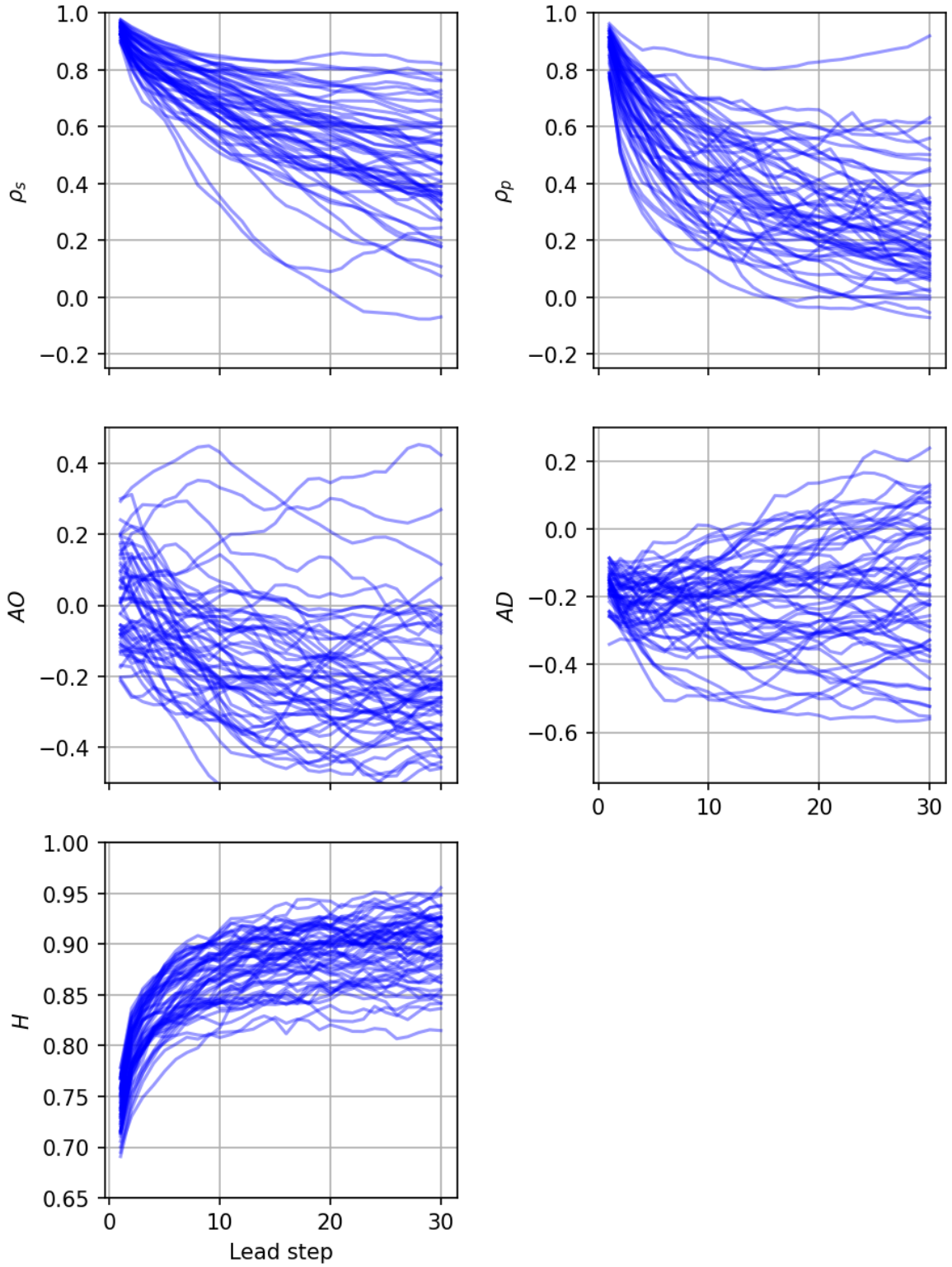


Figure 8.1: Various auto-properties of daily discharge for every 1 year window.

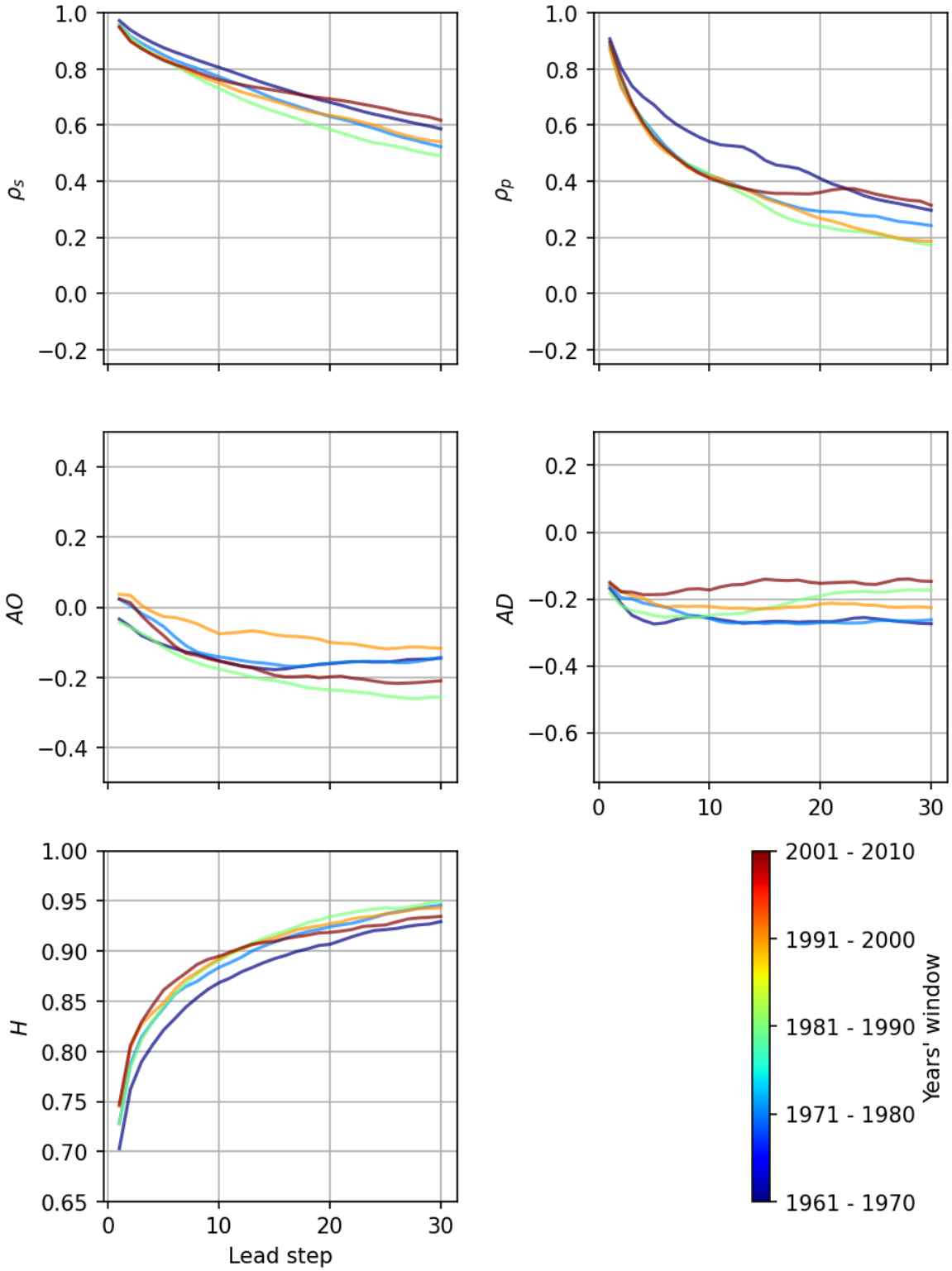


Figure 8.2: Various auto-properties of daily discharge for every 10 years window.

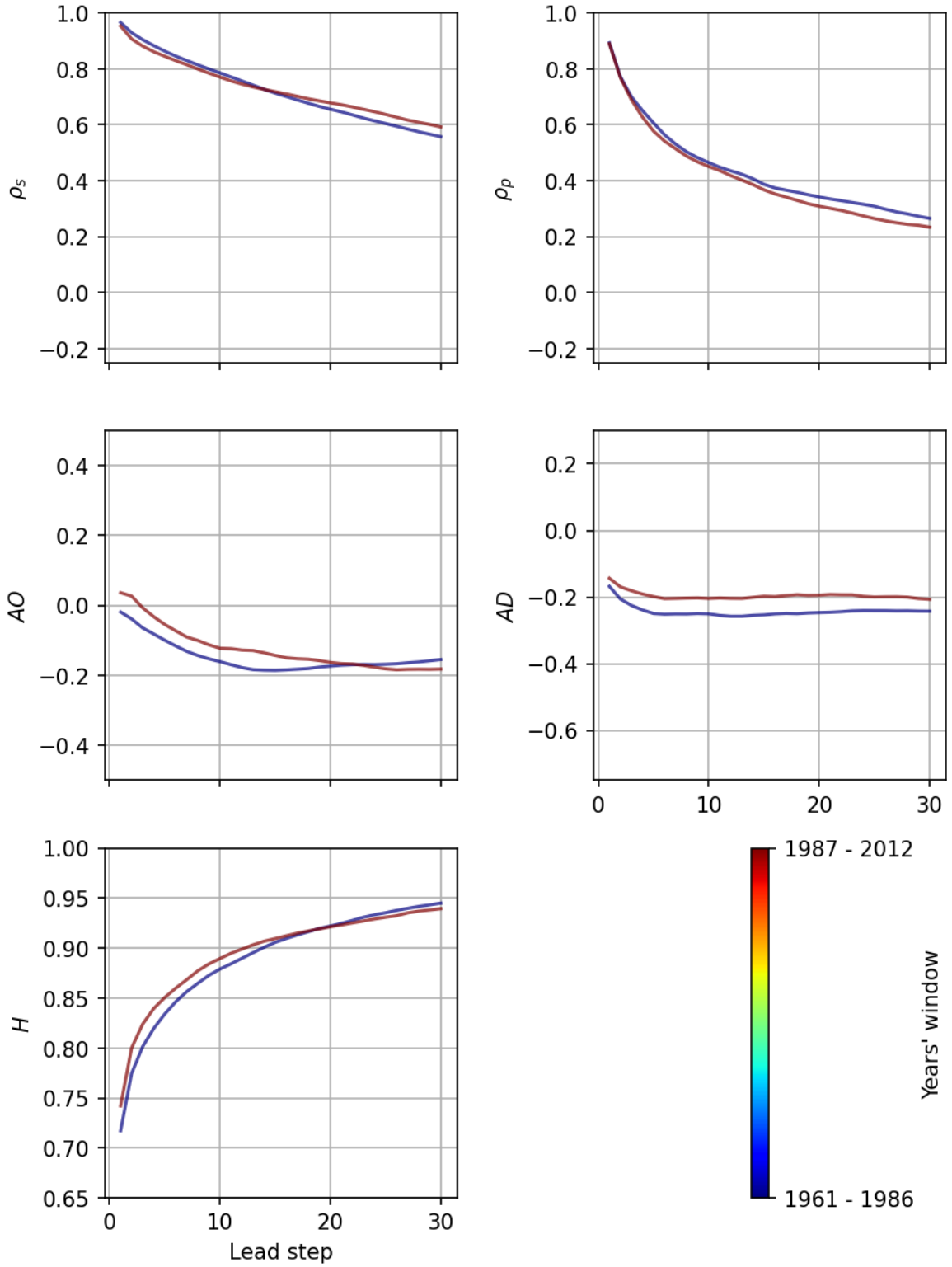


Figure 8.3: Various auto-properties of daily discharge for every 26 years window.

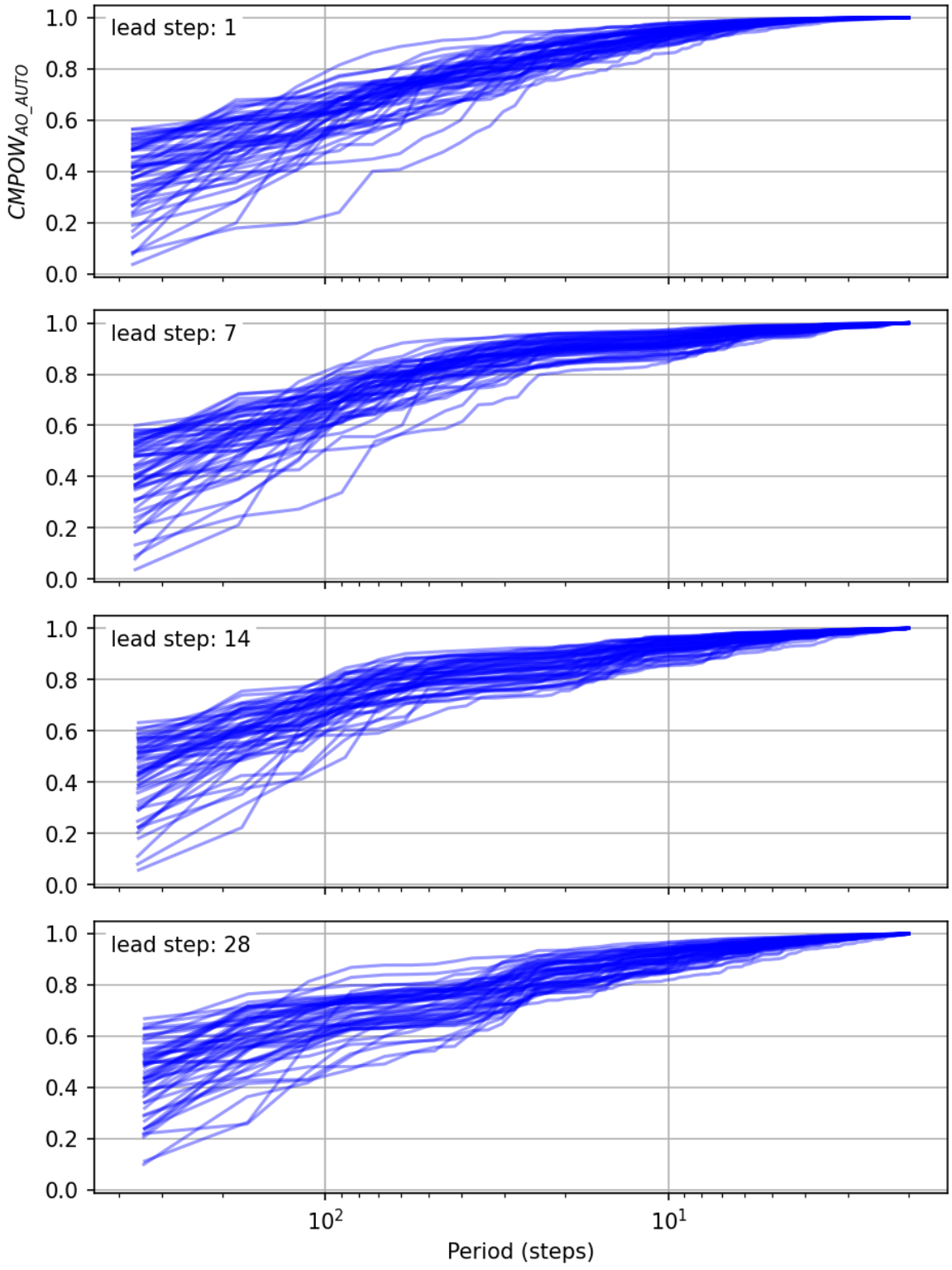


Figure 8.4: Various $CMPOW_{AO\_AUTO}$ of daily discharge for every 1 year window.

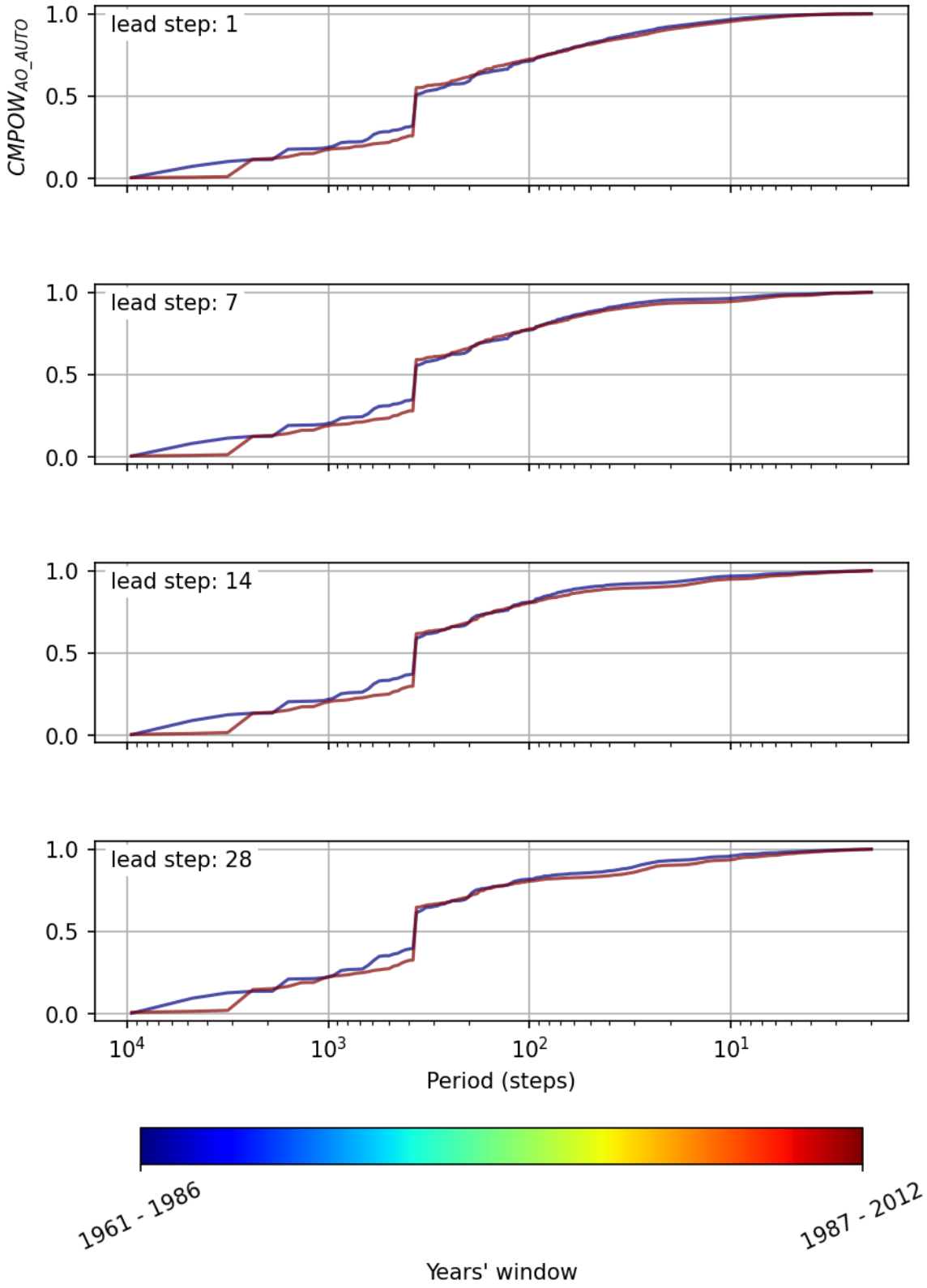


Figure 8.5: Various $CMPOW_{AO\_AUTO}$ of daily discharge for every 26 years window.

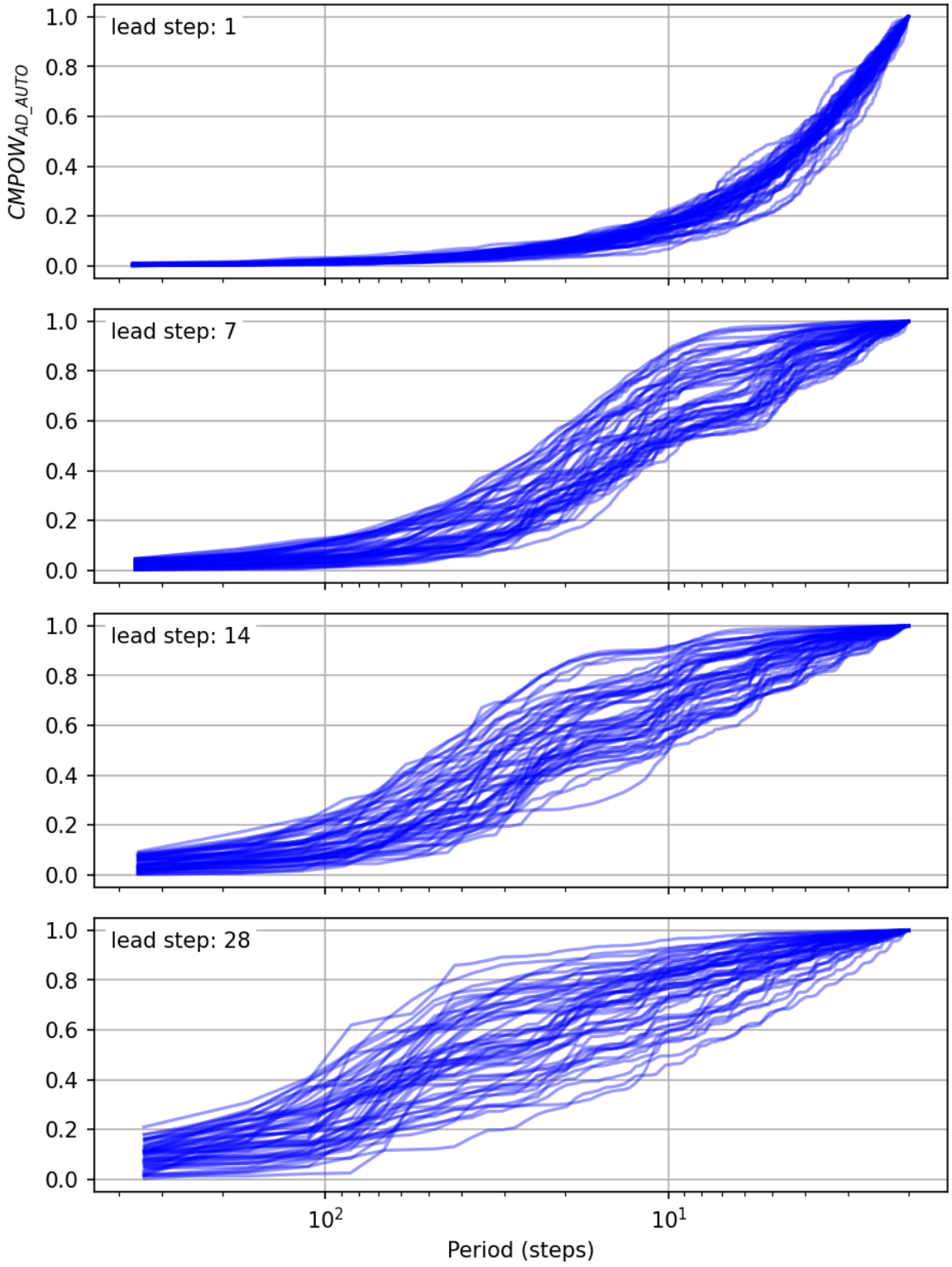


Figure 8.6: Various $CMPOW_{AD\_AUTO}$ of daily discharge for every 1 year window.

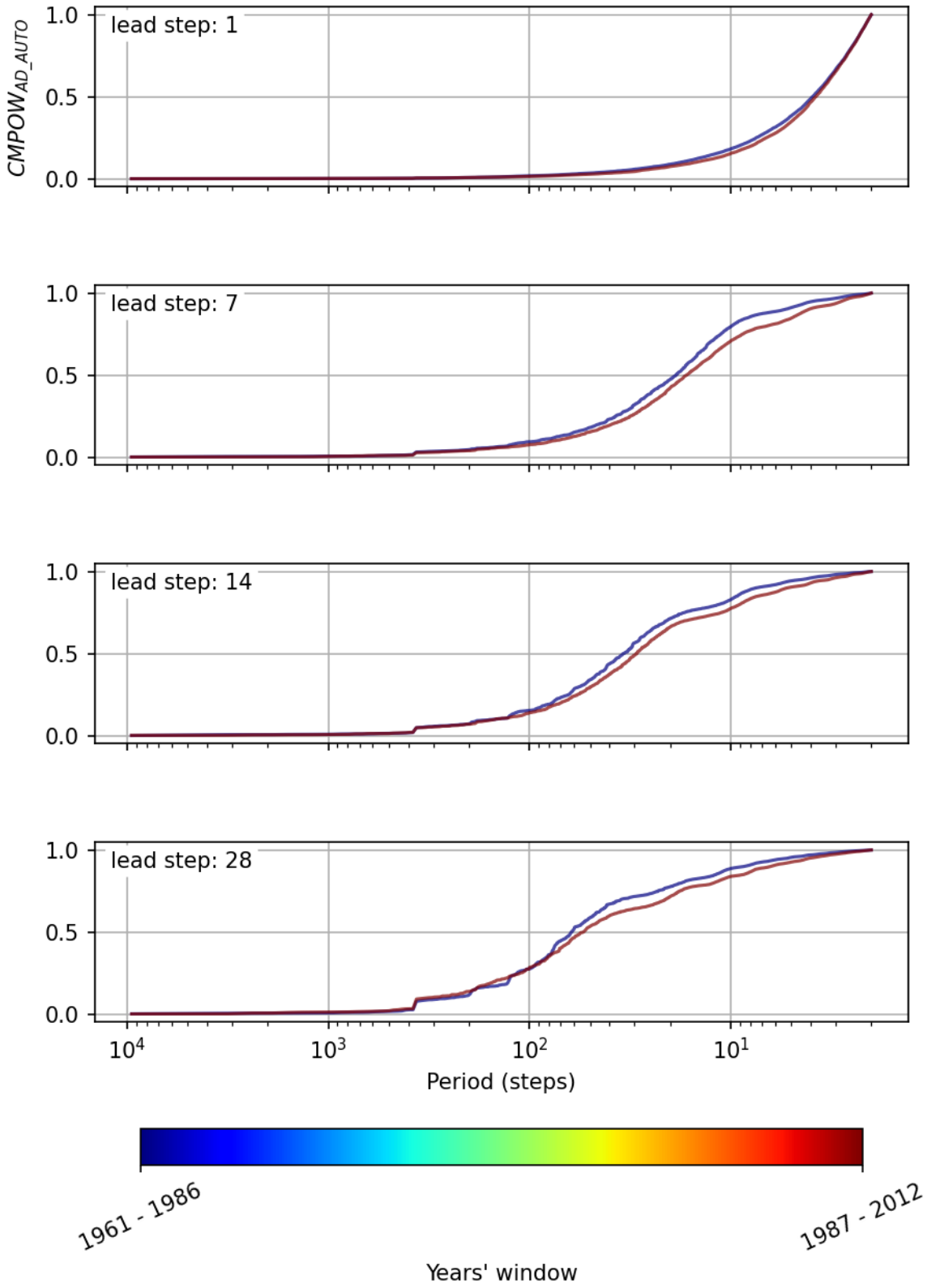


Figure 8.7: Various $CMPOW_{AD\_AUTO}$ of daily discharge for every 26 years window.

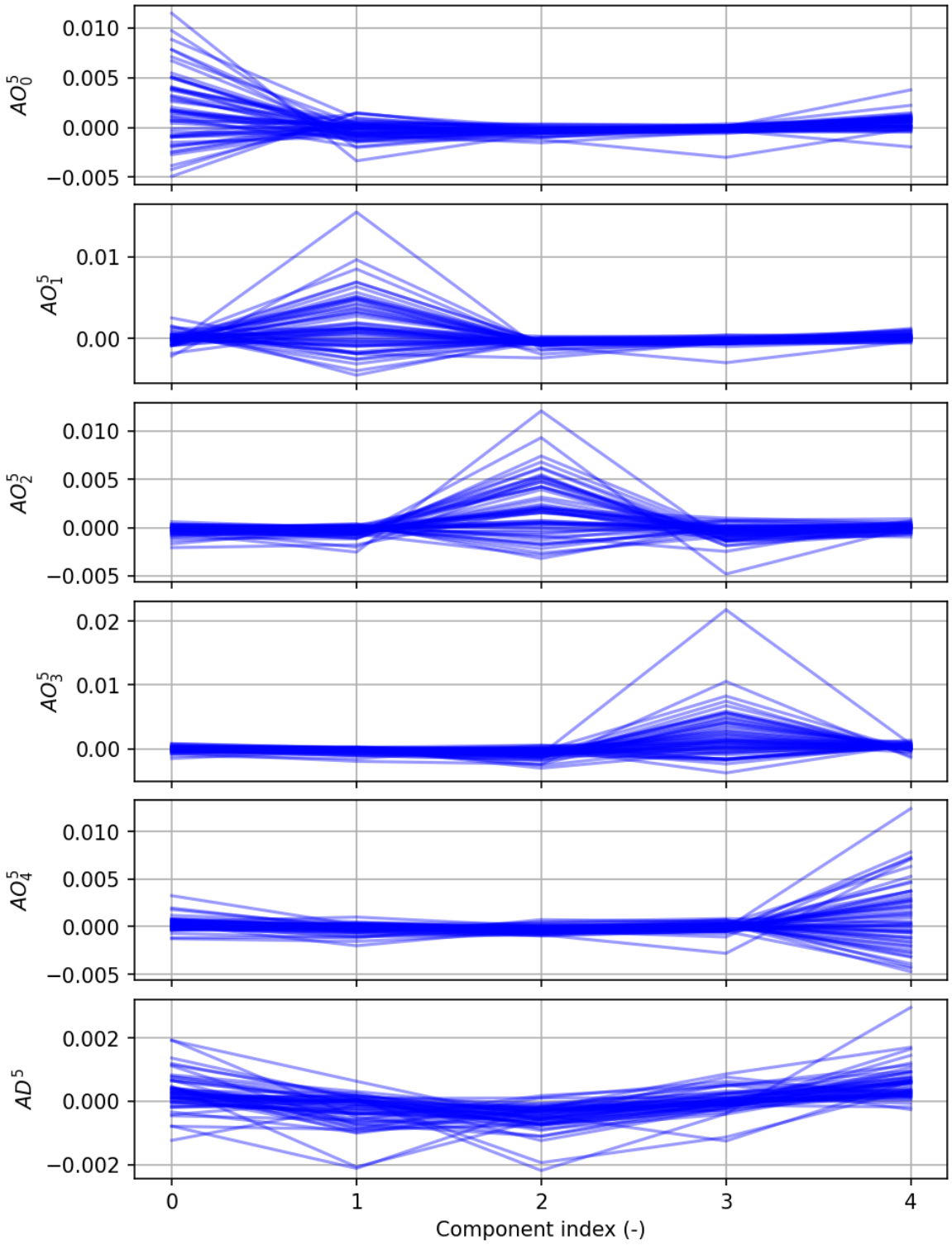


Figure 8.8: d-asymmetries of daily discharge for every 1 year window.

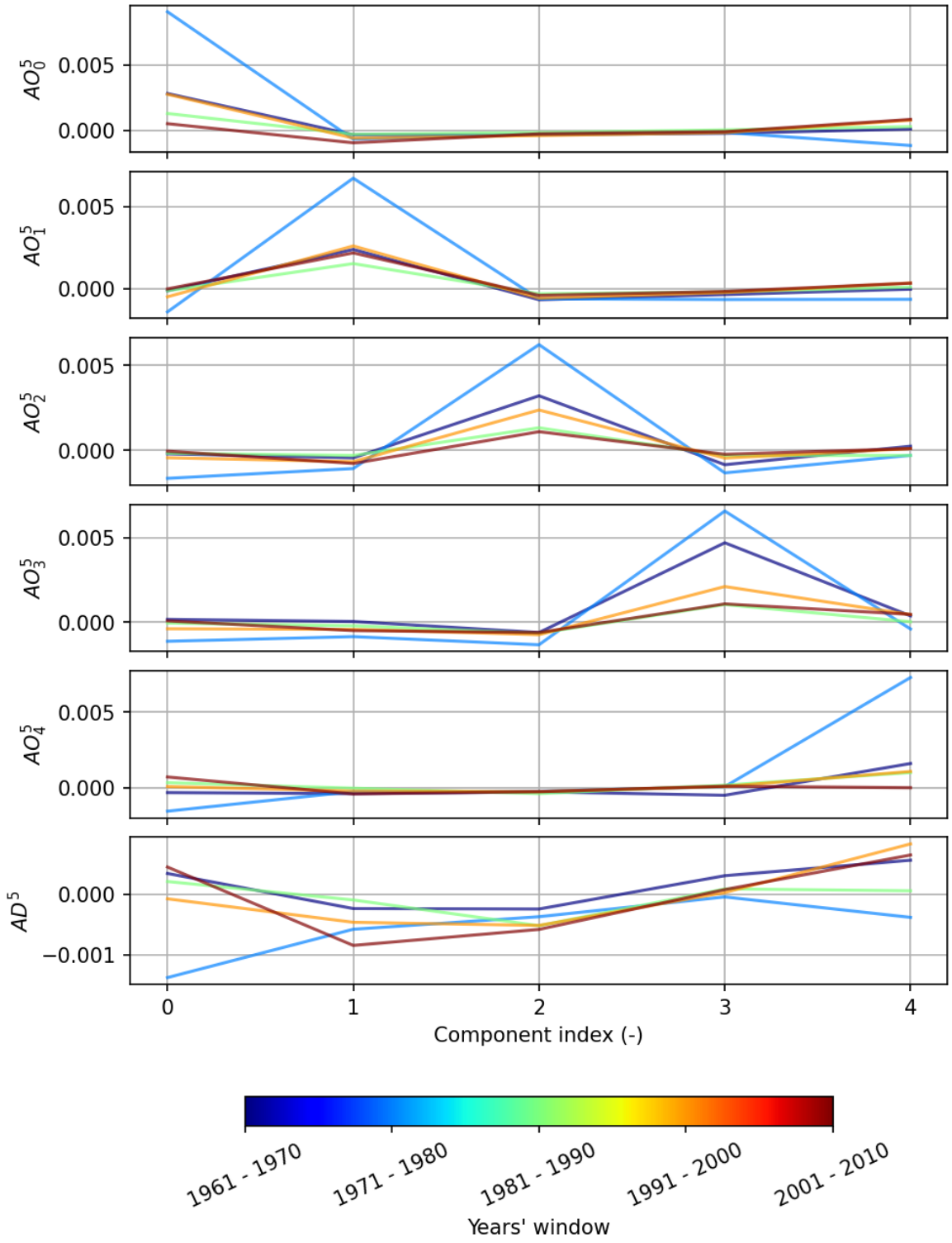


Figure 8.9: d-asymmetries of daily discharge for every 10 years window.

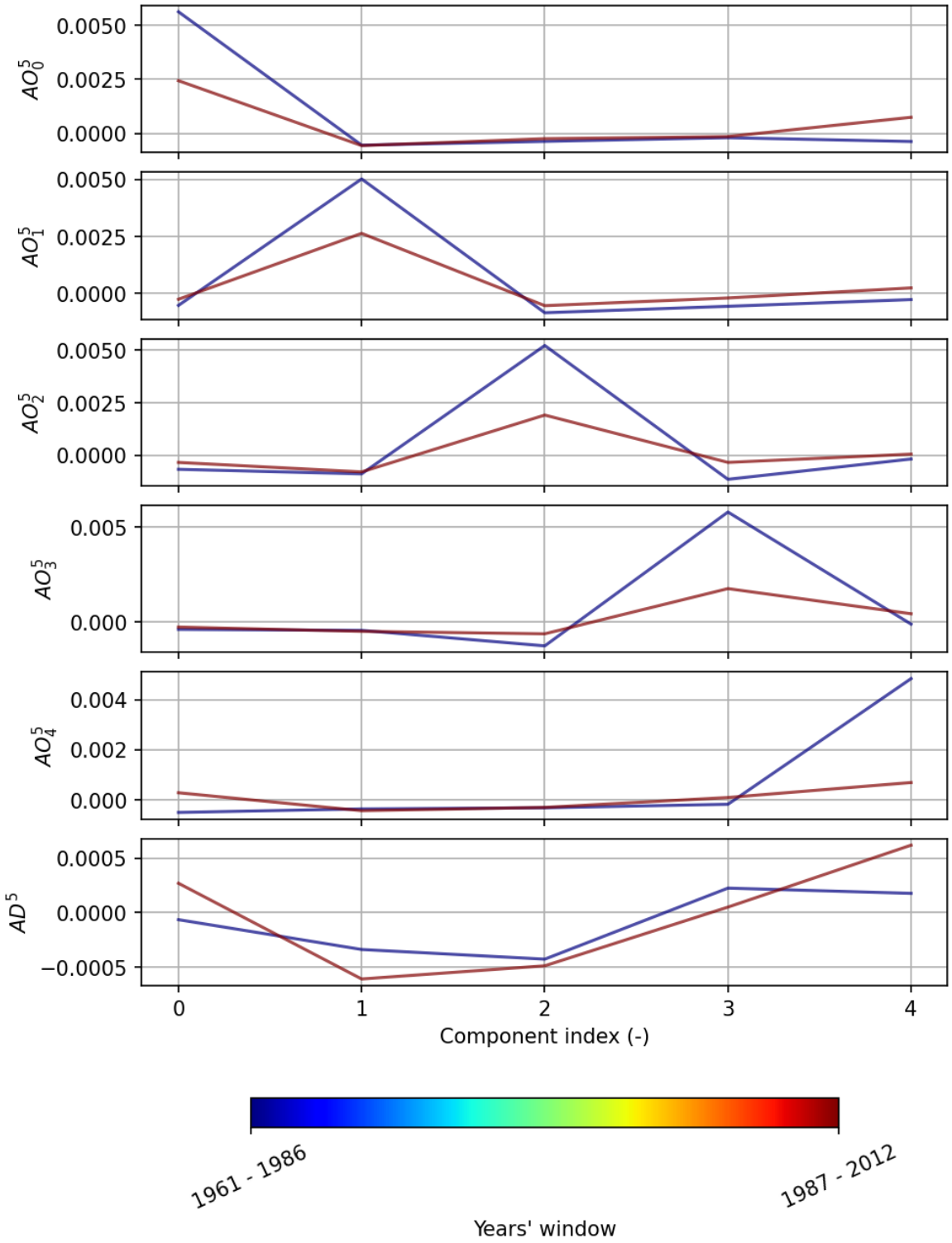


Figure 8.10: d-asymmetries of daily discharge for every 26 years window.

no contribution in the frequencies with periods shorter than 1 year time. Also, there seems to be a trend in time series as there exists a strength in frequencies longer than 1 year time in some windows. Interestingly, all of them exhibit very similar behavior. Considering the $CMPOW^d_{AD}$, especially, for the longest windows, the directional asymmetry spectra show considerable difference. Firstly, for the two 26 year series, the past time period (1961-1986) has a considerably large power in the longest frequency while the recent (1987-2015) does not. Secondly, the recent has a significant contribution in the annual cycle. Apart from that, some contributions from the high frequencies can be observed as well. This could be an indication of the evolving catchments or more probably that the direction of prevailing winds that bring rain has changed. In the recent case, there exists the synchronization of peaks in the annual cycle mainly, while in the other a very large shift in the timings of the peaks due to the very large frequency can be seen.

Fig. 8.15 and 8.16 show the d-dimensional cumulative marginal correlation periodogram ($CMCORR^d_p$) and d-dimensional cumulative grade correlation periodogram ($CMCORR^d_s$) respectively for the same time periods but only considering the frequencies whose periods are less than one year time. Again, very similar patterns compared to previous cases can be observed. Most of the contribution is coming from the longer frequencies, while the shorter ones seem to have no power. $CMCORR^d_s$ appears to be much more stable than $CMCORR^d_p$. This mismatch is directly related to amounts of rainfall where the relative order stayed the same but magnitudes changed over time. Furthermore, a very interesting pattern can be observed. Namely, that the power in the shorter cycles increased for the more recent times in terms of the combined behavior for both $CMCORR^d_p$ and $CMCORR^d_s$. This can be clearly seen in the 10 and 26 year time windows. It could be an indication that lengths of the hydrological years fluctuate more for the present than for the past.

Fig. 8.17, 8.18 and 8.19 show the d-dimensional cumulative order asymmetry correlation periodogram ($CMCORR^d_{AO}$) and d-dimensional cumulative directional asymmetry correlation periodogram ($CMCORR^d_{AD}$) for the time windows of 1, 10 and 26 years respectively. These also show considerable variability compared to $CMPOW^d_{AO}$ and $CMPOW^d_{AD}$. It is curious to note that the contribution of frequencies shorter than the annual cycle when considering the phases is much more pronounced here. For the case of $CMPOW^d_{AO}$s, the half-year cycle shows a significant contribution while it is missing for $CMPOW^d_{AD}$ when looking at the 10 and 26 year time windows. Also, the overall value is much smaller because the contribution of frequencies up to and including the annual cycle is not shown. Similar to $CMPOW^d_{AO}$ and $CMPOW^d_{AD}$, it can be stated with confidence that the overall timings of values relative to each other are largely synchronized on larger time periods. In terms of climate change, a consistent pattern cannot be seen on the decadal scale but only for the 26 year windows. There, it can be observed that the strength of the recent times (1987-2012) compared to the past (1961-1986) has increased for the $CMCORR^d_{AO}$. This signifies that for each spectrum, the distance of points in the d-dimensional space from the respective line through the hypercube has more correlation in time as compared to the past which is similar to what was observed previously for the case of $CMCORR^d_s$. However, $CMCORR^d_{AD}$ shows that although the final value that both the recent and past curves obtain is very close, the relative contributions in the long and short frequencies is not the same. For the recent times, it the contributions in the short frequencies are larger. This signifies that the com-

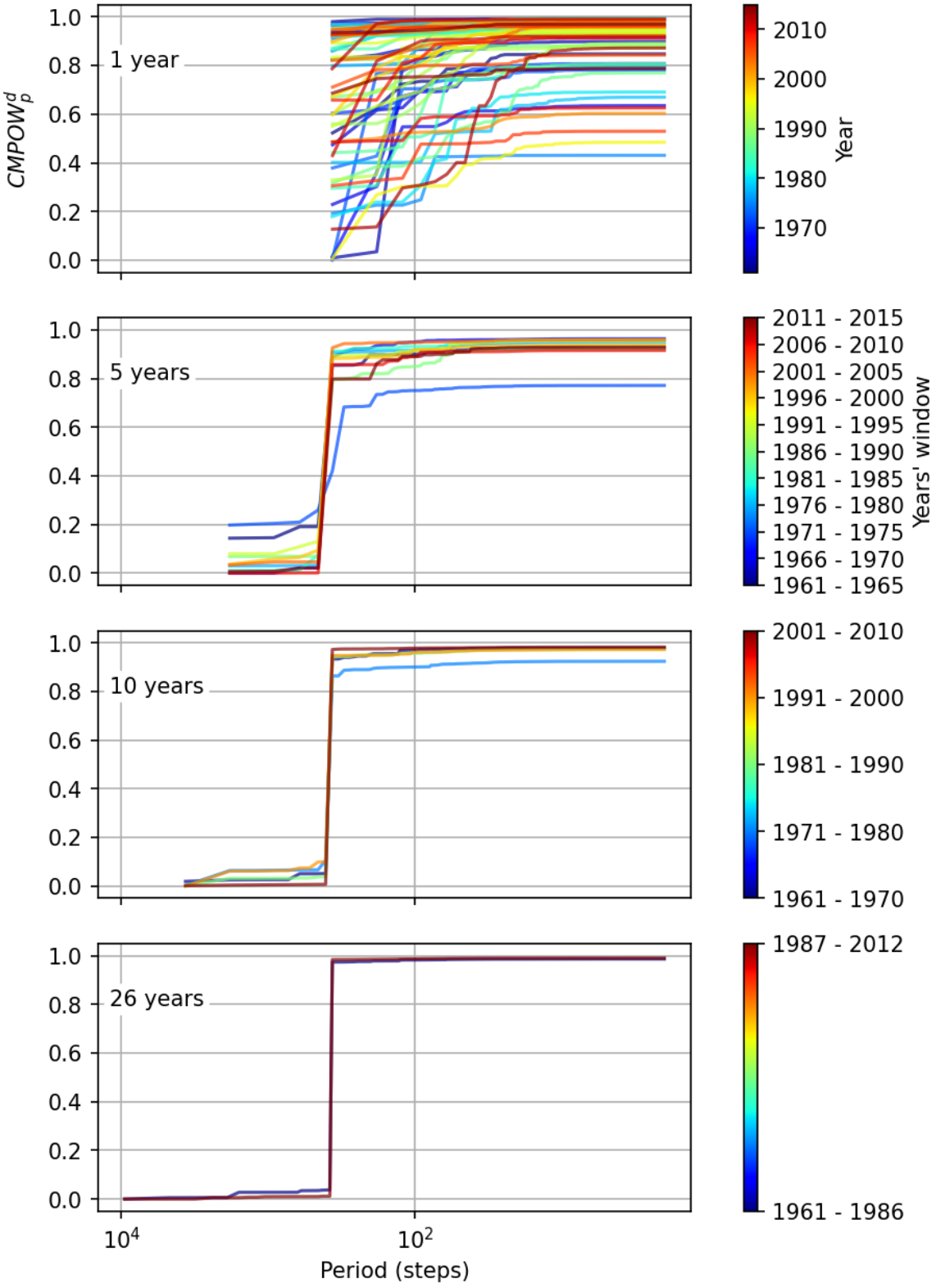


Figure 8.11: $CMPOW_p^d$ of daily discharge for various time windows.

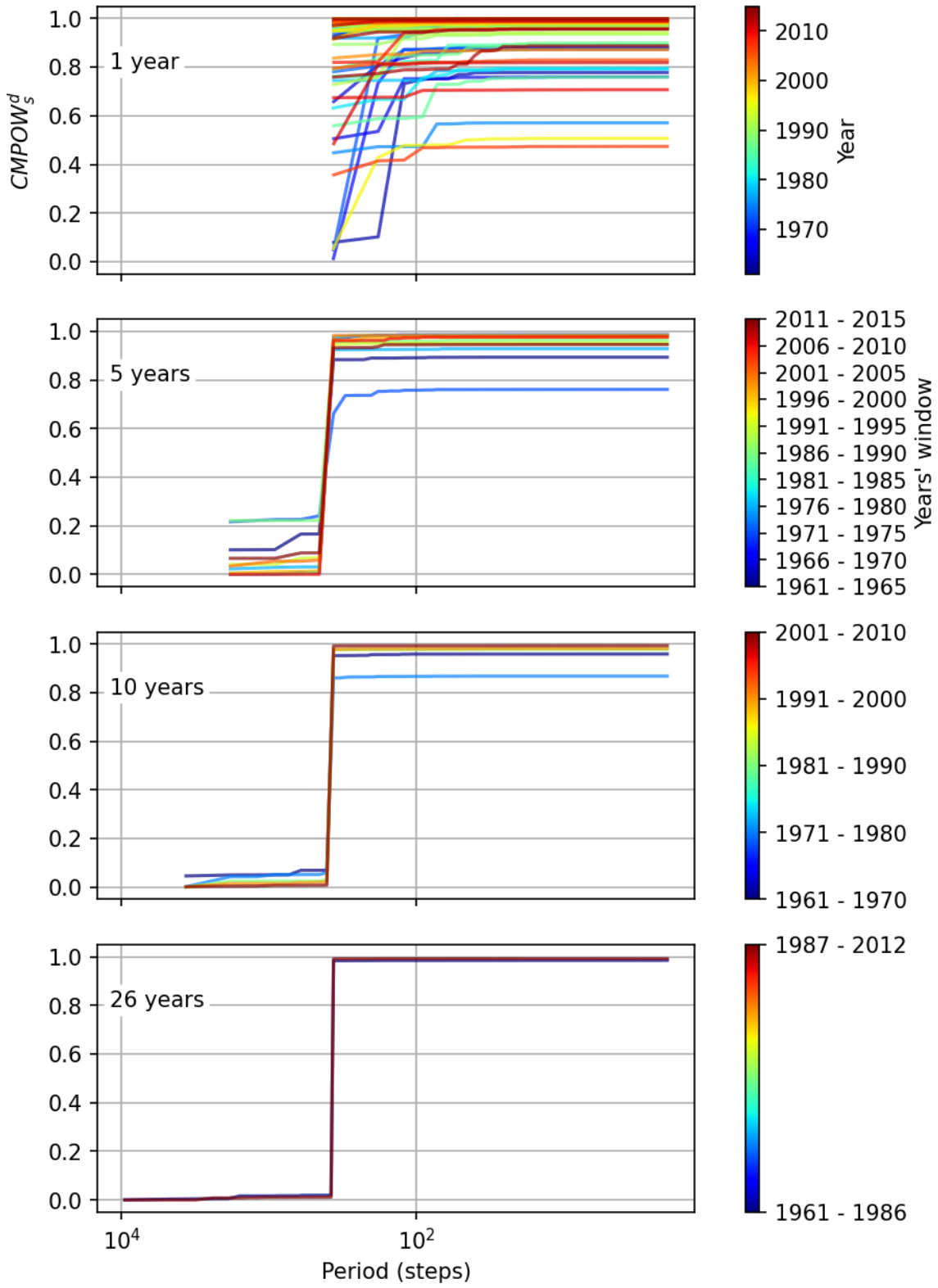


Figure 8.12: $CMPOW_s^d$ of daily discharge for various time windows.

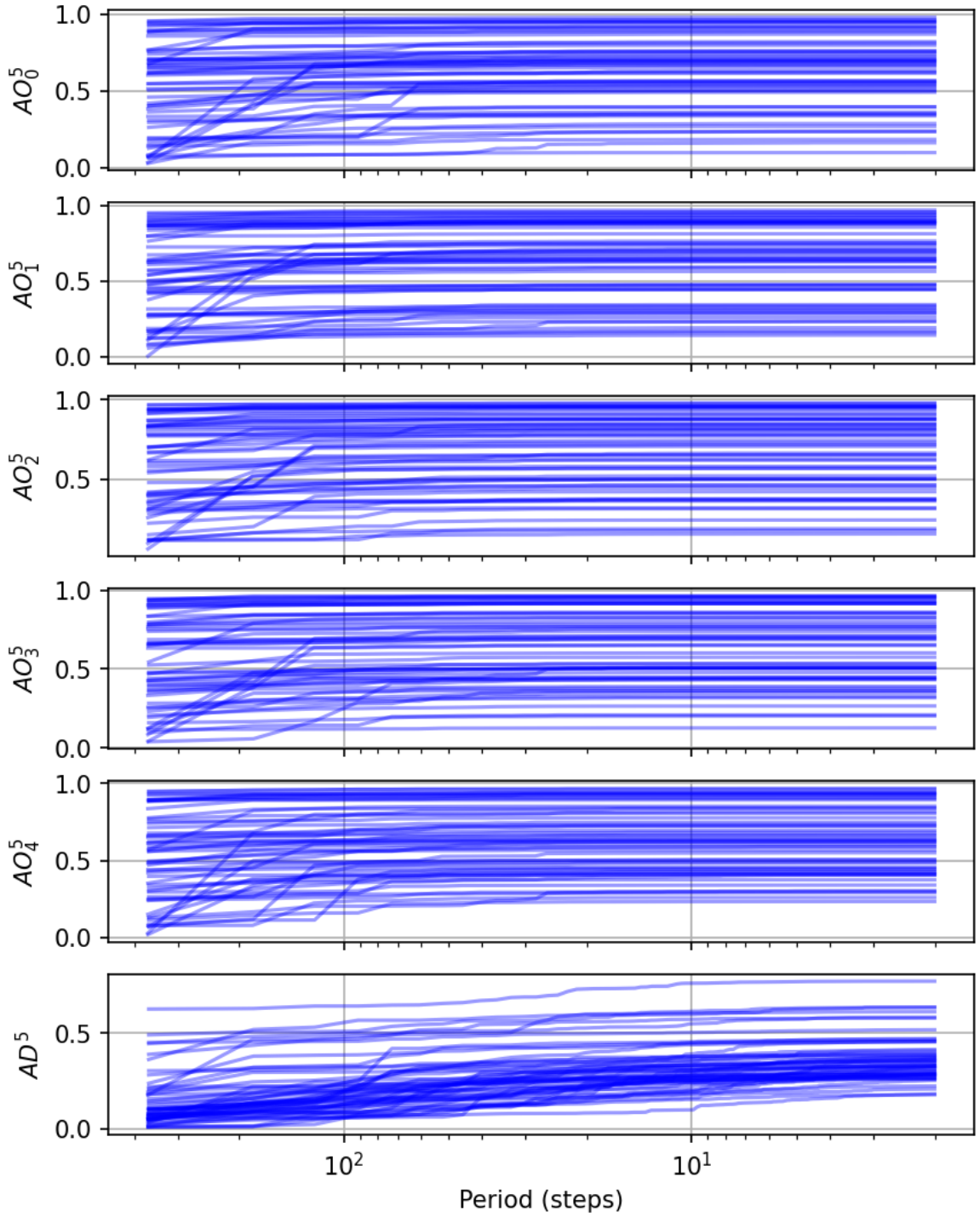


Figure 8.13: $CMPOW_{AO}^{d}$ (top five) and $CMPOW_{AD}^{d}$ (bottom) of daily discharge for every 1 year window.

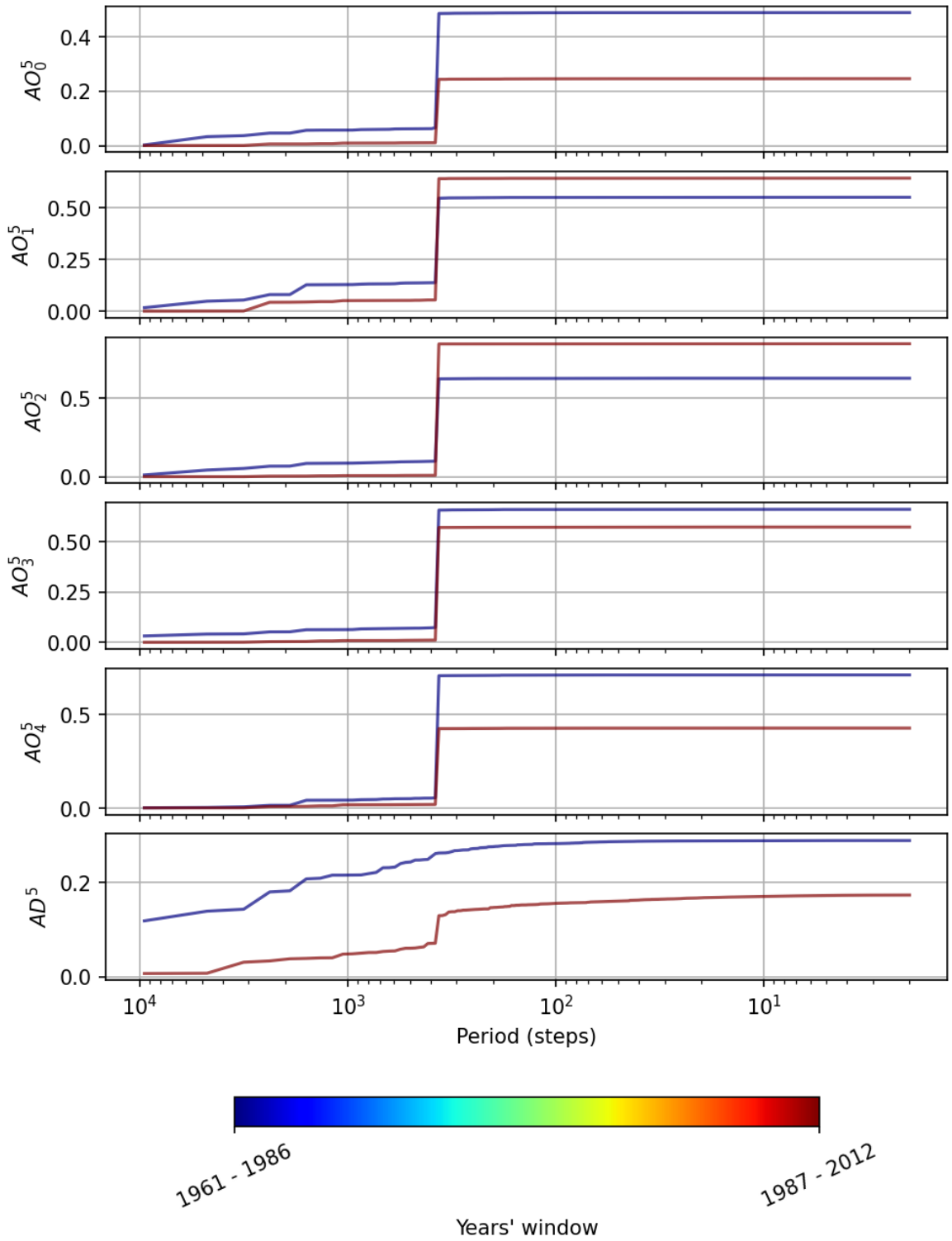


Figure 8.14: $CMPOW_{AO}^{d}$ (top five) and $CMPOW_{AD}^{d}$ (bottom) of daily discharge for every 26 years window.

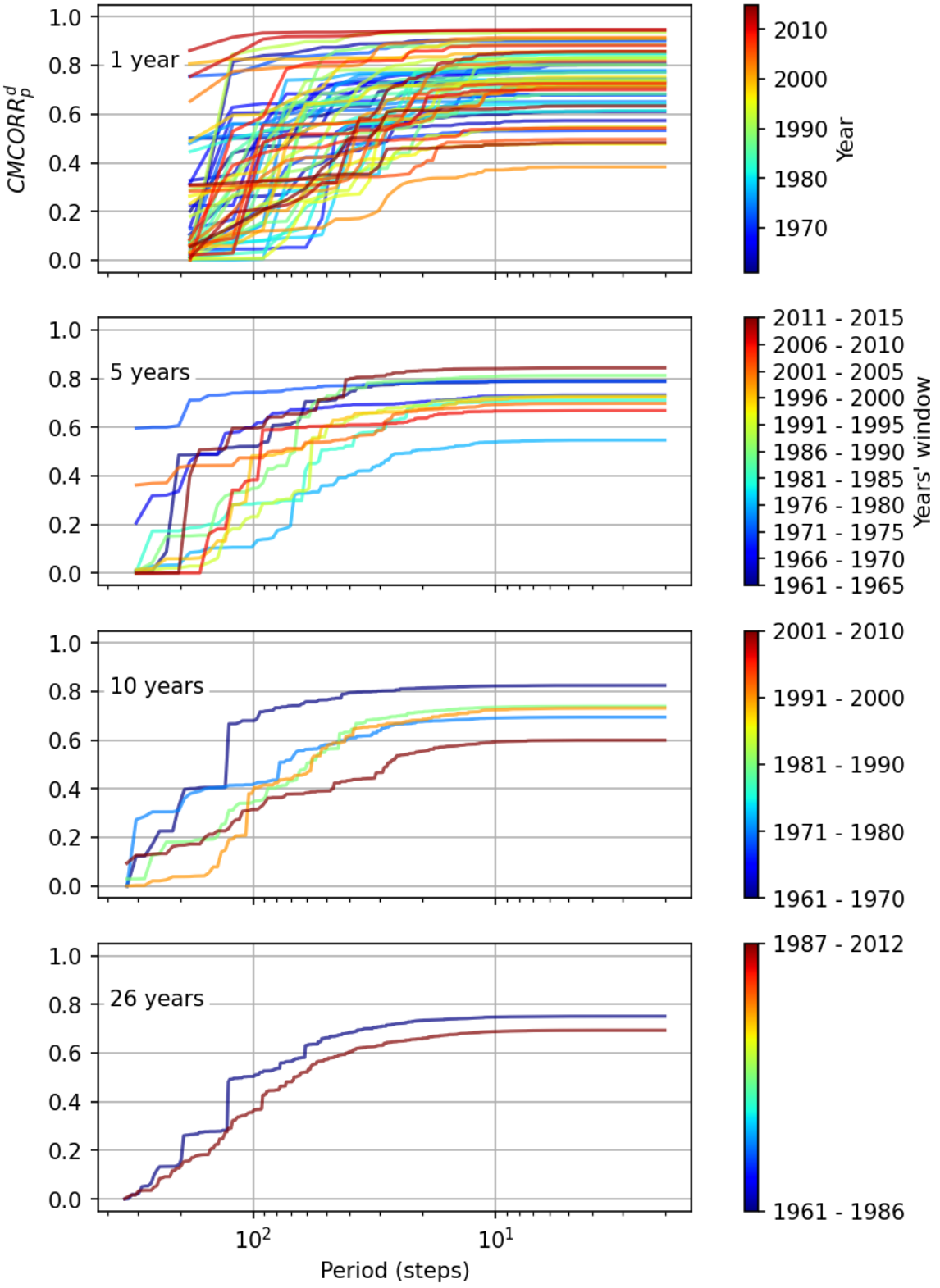


Figure 8.15: $CMCORR_p^d$ of daily discharge for various time windows. Note that only the contribution of frequencies of periods less than one year time are considered.

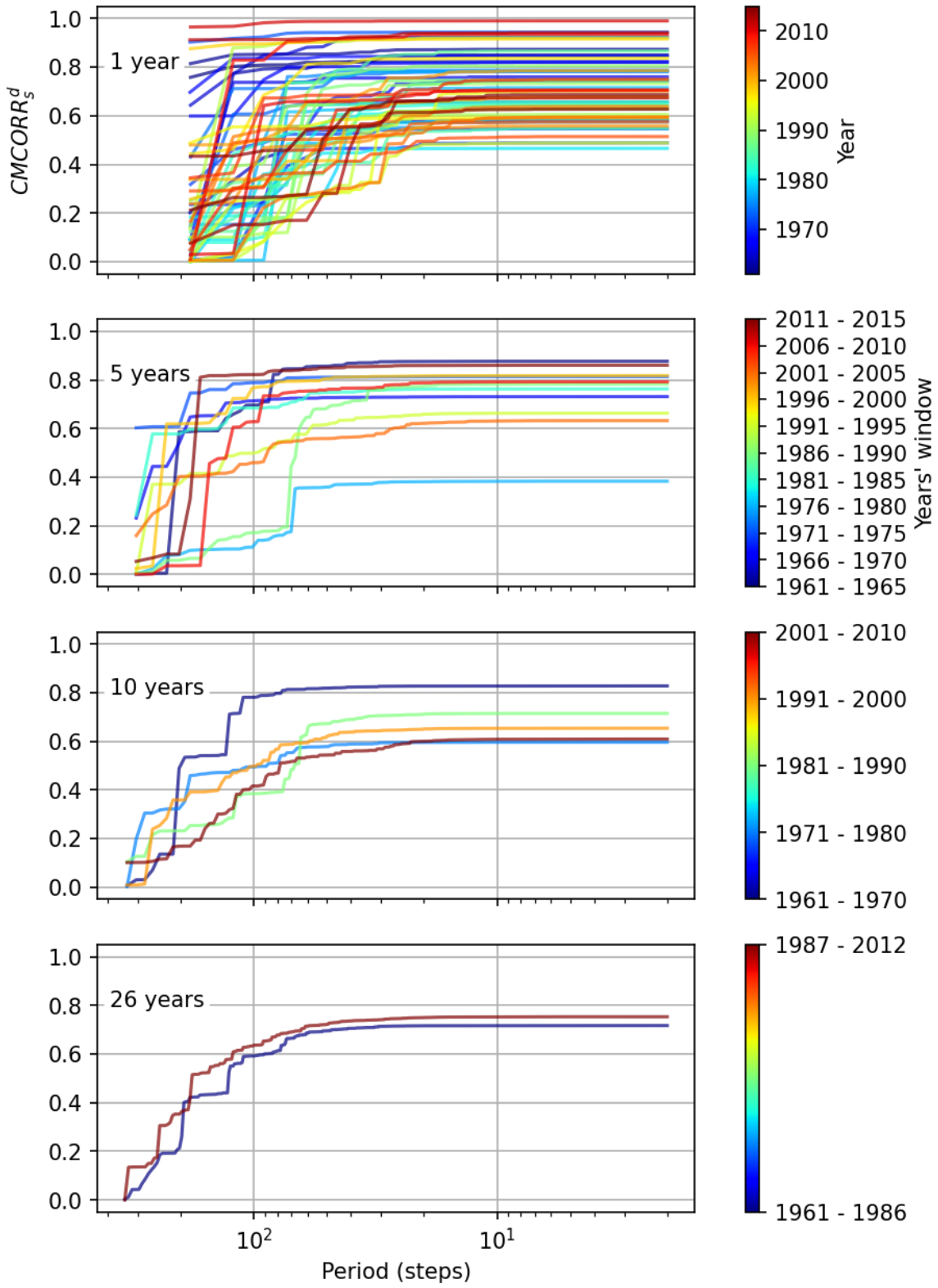


Figure 8.16: $CMCORR_s^d$ of daily discharge for various time windows. Note that only the contribution of frequencies of periods less than one year time are considered.

bined behavior of the time series is noisier in the recent period than that in the past. Once again, the cause of this could be due the decreasing number of snowy day as the average temperature of the study area is increasing over time. It would be interesting to investigate the exact cause(s) for this behavior.

Coming to matter of main interest, Fig. 8.20 shows the distribution of spatial sums for various windows of time. It is interesting to note that the sums of the highest values show a considerable difference even for the longer periods. This highlights the difficult problem of extrapolation. Generally, as the extremes become larger, they lose correlation and tend to a constant value, but here the slope is still positive. Meaning that they still have considerable dependence and the even larger values may take place in the future. No consistent increasing or decreasing pattern can be observed for the time windows of 1, 5 and 10 years. Only the 26 year windows show a significant difference. The recent period (1987-2012) compared to the past (1961-1986) had more combined discharge but only in the upper tail i.e., probabilities less than 1 in a 100 days. This could also be related to less snowy days somehow and requires further investigation.

In order to see the difference of dependence of distinct sets of data, daily precipitation from the study area is used as a demonstration variable. Discharge cannot be used in this instance as the gauges are too near to each other and also that discharge is an integrated output of an area whereas precipitation is only at points. To do so, three sets having five precipitation time series each are used. Two sets for stations that are near to each other (A and B) while the other is of stations that are far away from each other (C). Each point in the sets have a mean distance 39, 32 and 161 km with its neighbors respectively. The description of these gauges was given in Chapter 4 Section 4.1 and Fig. 4.3. The hypothesis is that the nearer the stations, the higher their overall correlation is. Due to the large incidence of zeros, their overall form is different compared to that of discharge. Fig. 8.21 shows the $CMPOW_p^d$ of the three multivariate series for the same length in time of 55 years. All three behave differently in that the spectra have contributions coming from different sections of the frequencies. B starts with low values in the long frequencies but over takes both A and C by having consistently higher magnitudes in higher frequencies. This signifies that B has large events that may synchronize in time but these events may take place at any time of the year. A and C show a different behavior in that there exists the potential that the events that take place in time have a relationship with the annual cycle. Fig. 8.22 shows $CMPOW_s^d$ for the same setting. Here, a different behavior can be observed in that B now shows an annual cycle in the copula domain and all show a significant half-year cycle as well. Moreover, the contribution of high frequencies is much less compared to $CMPOW_p^d$. This points to the fact the in the marginal domain, abrupt large values occur more often which means that the precipitation has a long upper tail that distorts the power spectrum at each location much more compared to $CMPOW_s^d$ where the effect of tails is not so pronounced. Having $CMPOW_p^d$ and $CMPOW_s^d$ is not enough of a constraint to decide whether a given set of variables behave consistently in space as it does not take the relative timing into account in a realistic manner, only the potential maximum synchronization can be extracted.

Fig. 8.23 and 8.24 show the $CMCORR_p^d$ and $CMCORR_s^d$ respectively for the same time period but only considering the frequencies whose periods are less than one year time. As

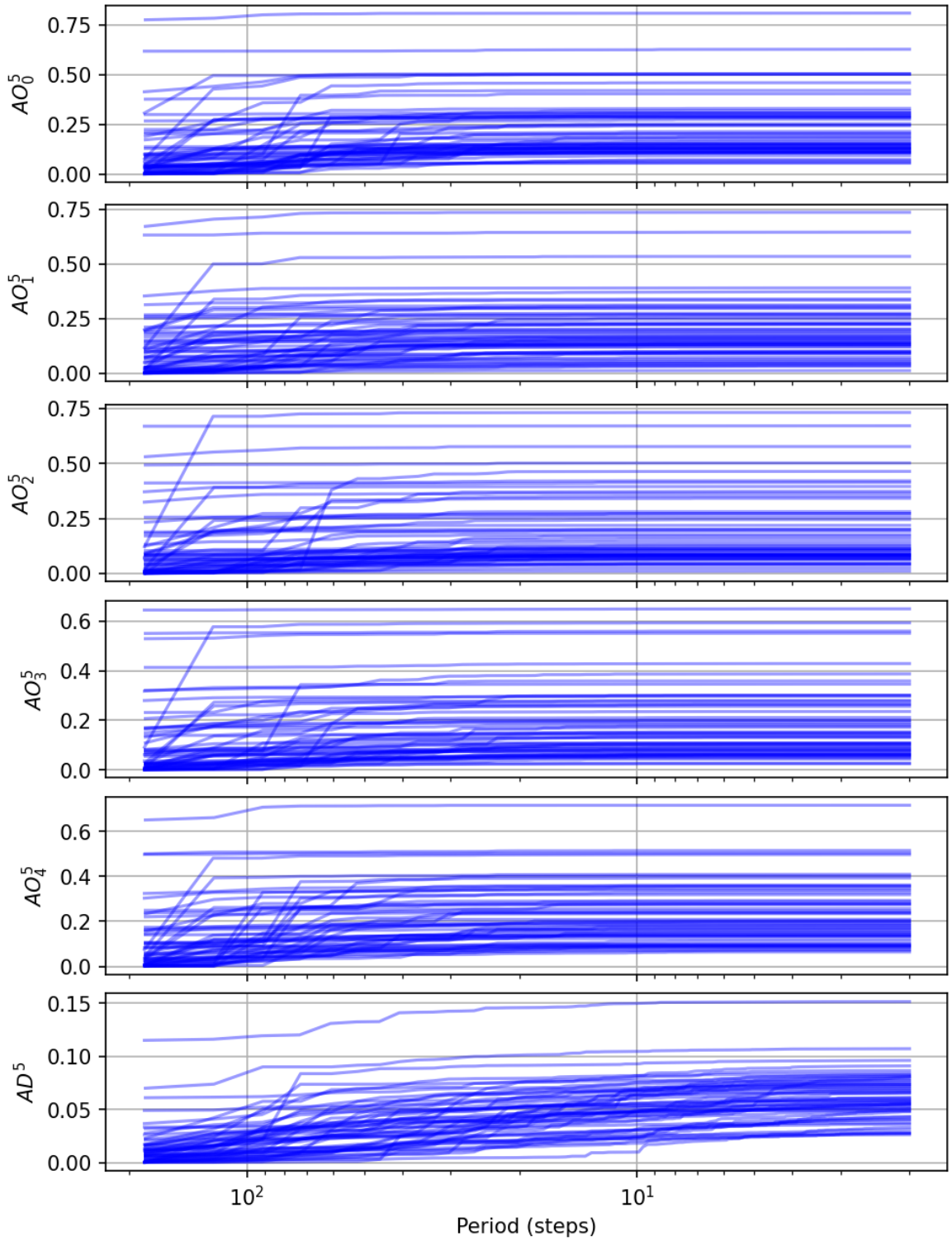


Figure 8.17: $CMCORR_{AO}^{d}$ (top five) and $CMCORR_{AD}^{d}$ (bottom) of daily discharge for every 1 year window.

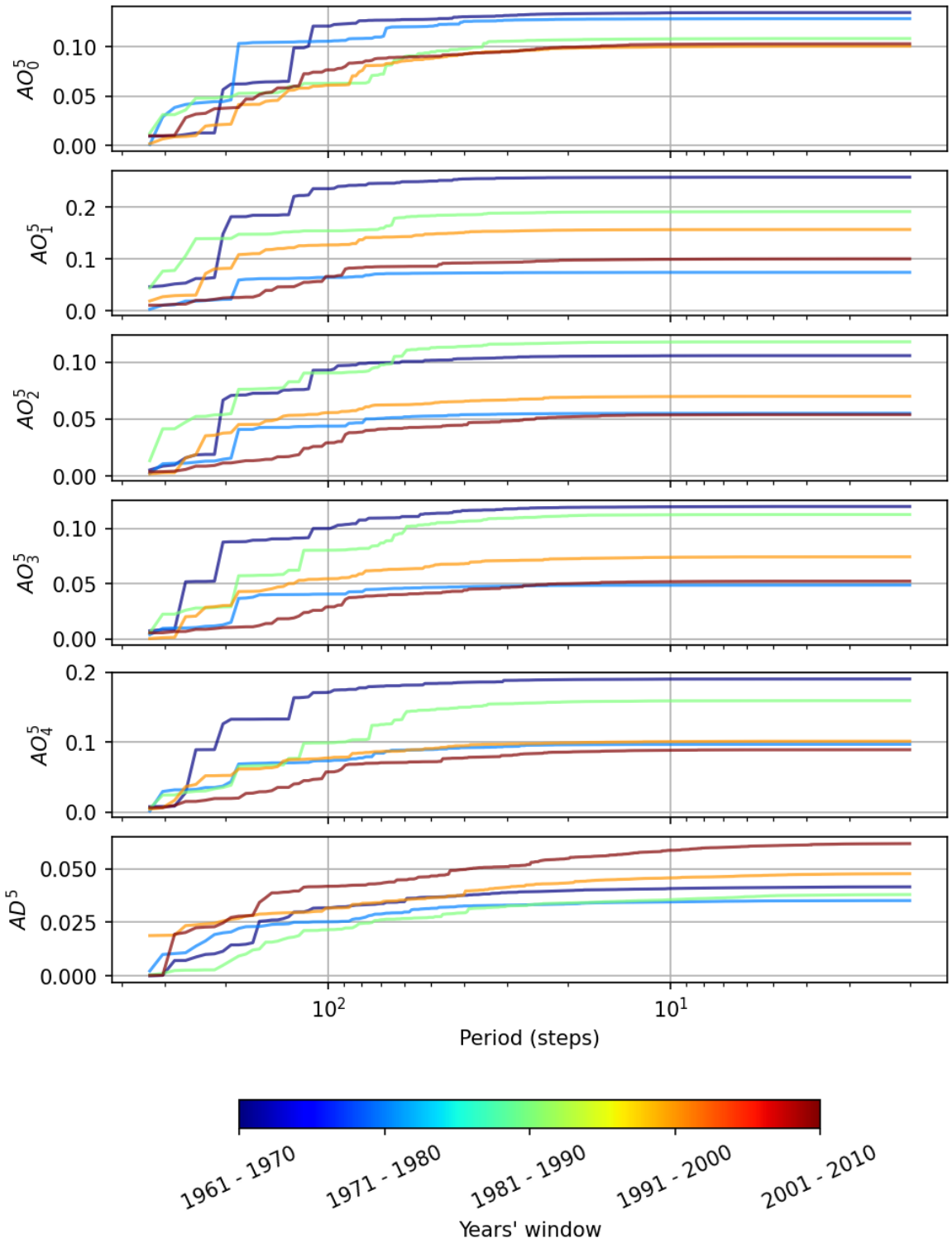


Figure 8.18: $CMCORR^d_{AO}$ (top five) and $CMCORR^d_{AD}$ (bottom) of daily discharge for every 10 years window.

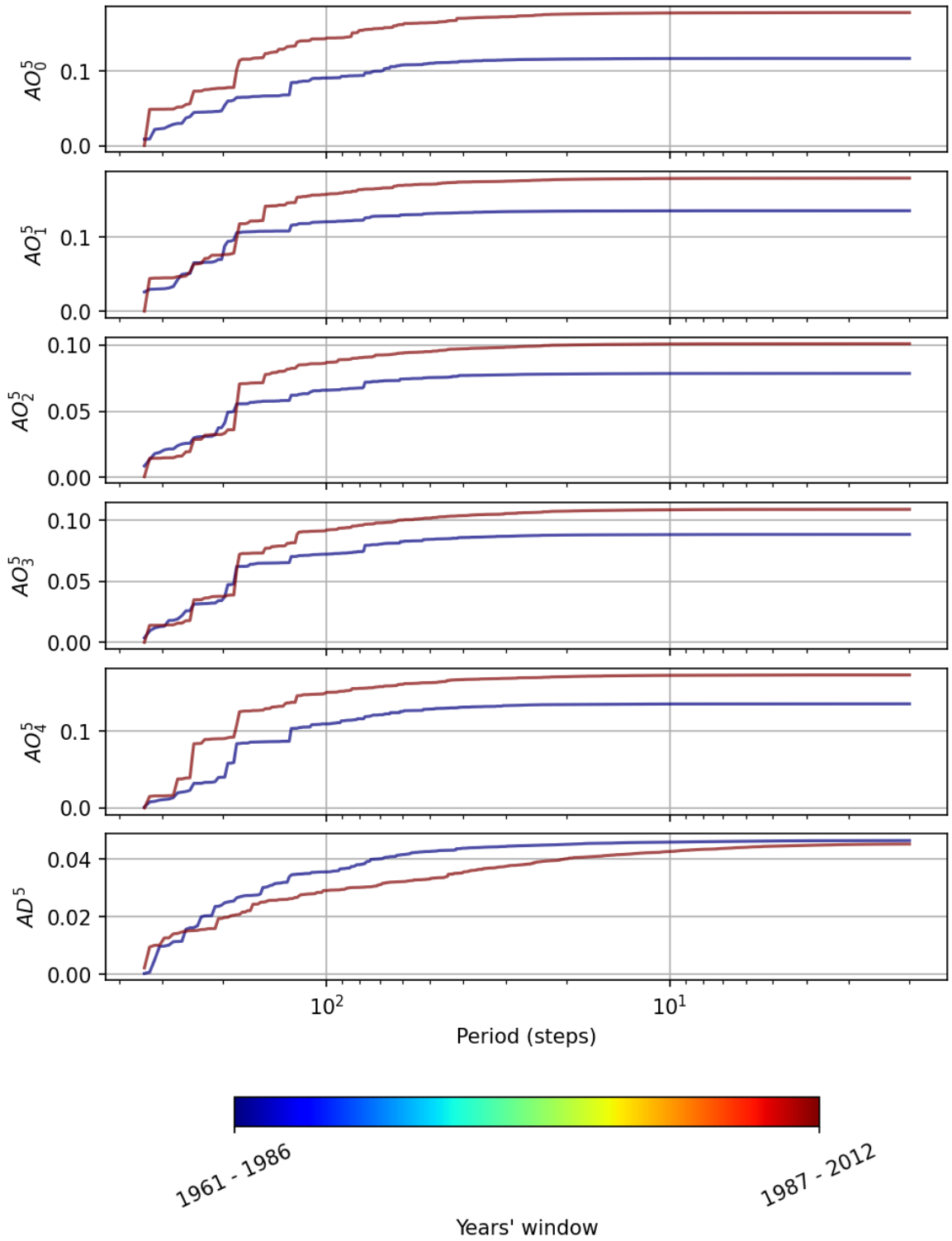


Figure 8.19: $CMCORR^d_{AO}$ (top five) and $CMCORR^d_{AD}$ (bottom) of daily discharge for every 26 years window.

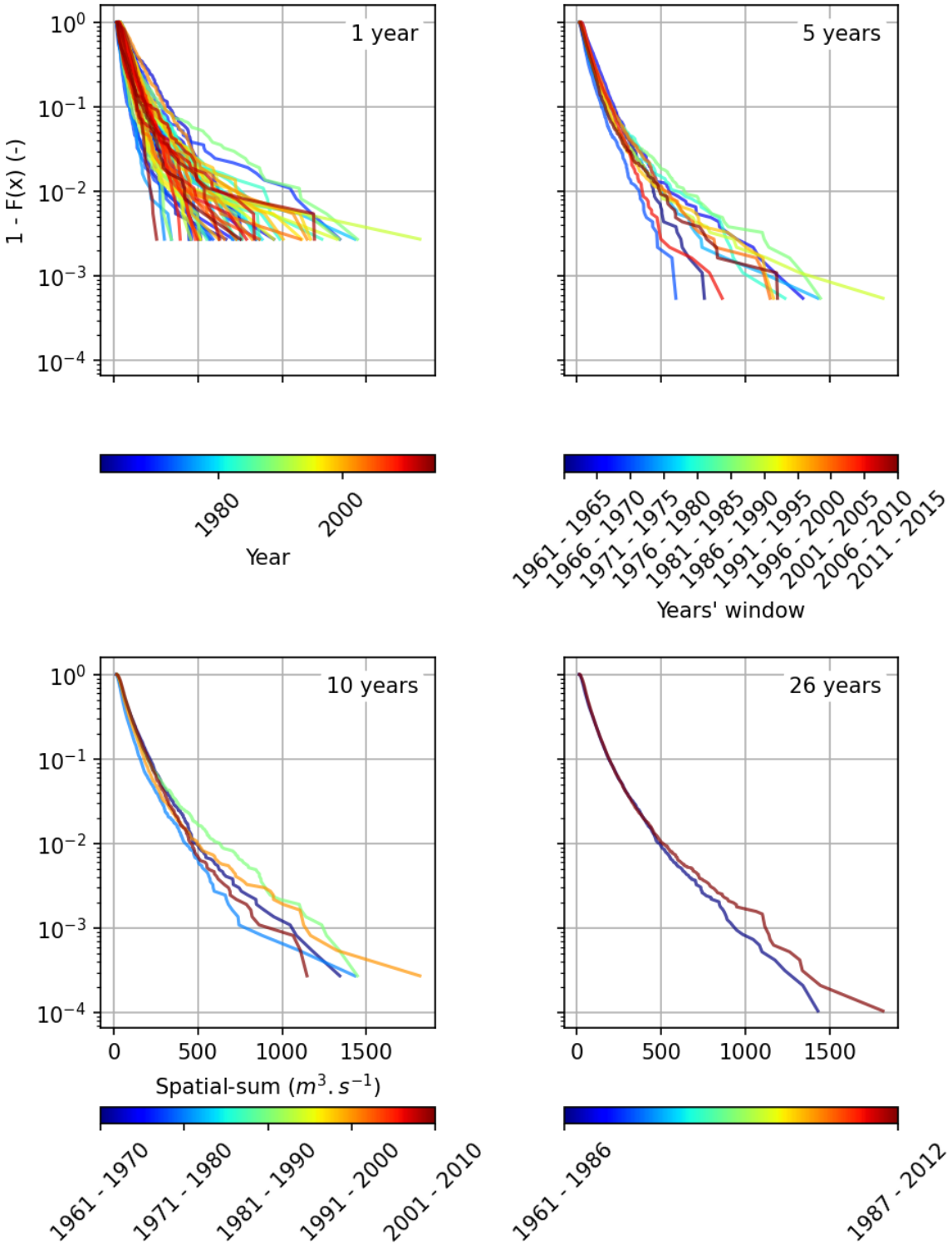


Figure 8.20: Distributions of spatial-sums of daily discharge for various windows.

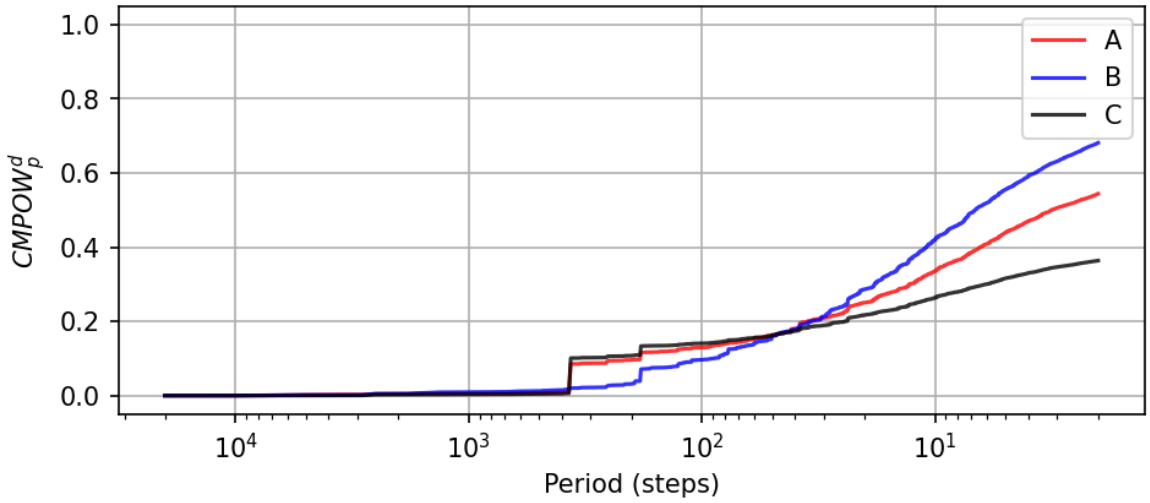


Figure 8.21: Comparison of $CMPOW_p^d$ of daily precipitation for three sets of stations.

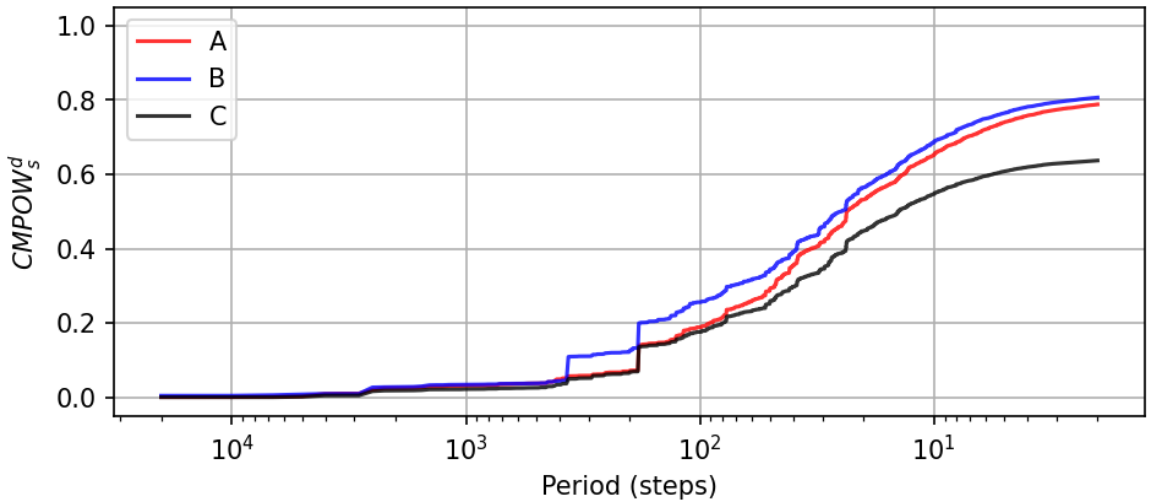


Figure 8.22: Comparison of $CMPOW_s^d$ of daily precipitation for three sets of stations.

the spectra of A and B grow close to each other while C spectra stay much more below them. This is inline with the assumption that points that are farther from each other should be less correlated than the nearer ones. Finally, Fig. 8.25 shows the daily spatial-sum distribution of the three sets. Similar to discharge, the nearer stations have bigger maxima compared to the farthest one. More importantly, the largest sums do not decorrelate and keep on increasing. Even for the stations that are 161 km apart from each other on average. This signifies that, same as discharge, in the future even larger sums are potentially possible. This is an important aspect that the simulated time series should also reproduce. This aspect will be evaluated for both discharge and precipitation in the chapter of Results.

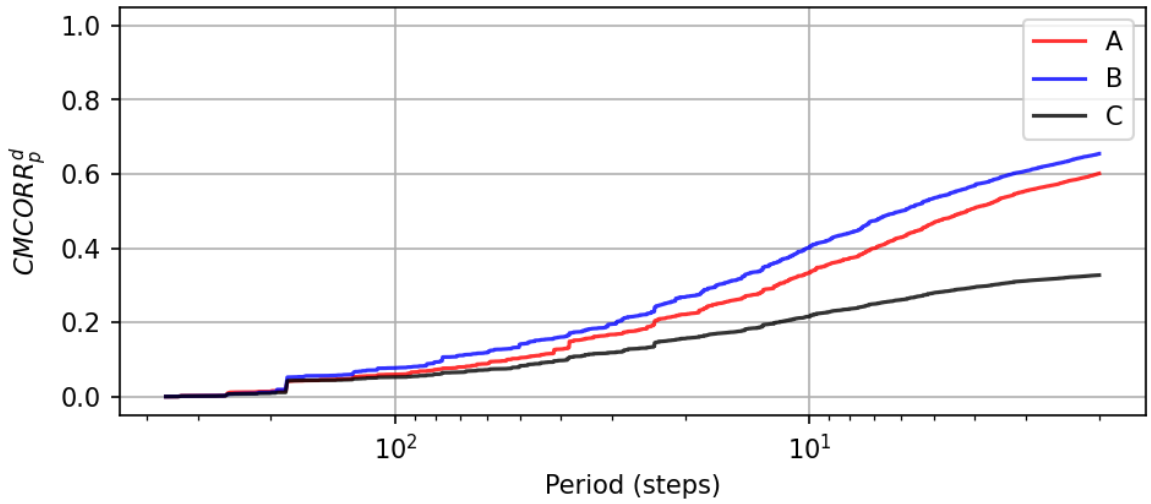


Figure 8.23: Comparison of $CMCORR_p^d$ of daily precipitation for three sets of stations. Note that only the contribution of frequencies of periods less than one year time are considered.

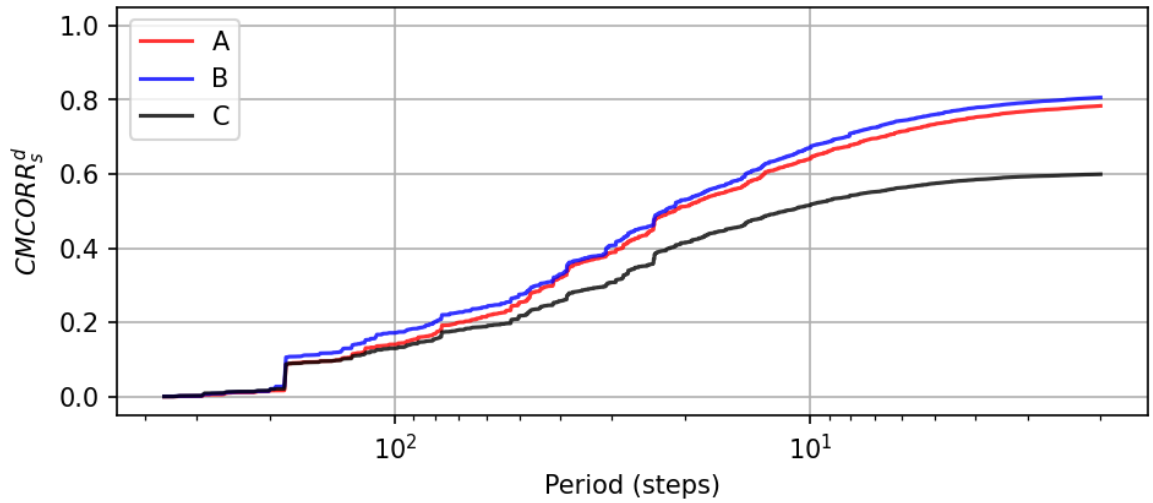


Figure 8.24: Comparison of $CMCORR_s^d$ of daily precipitation for three sets of stations. Note that only the contribution of frequencies of periods less than one year time are considered.

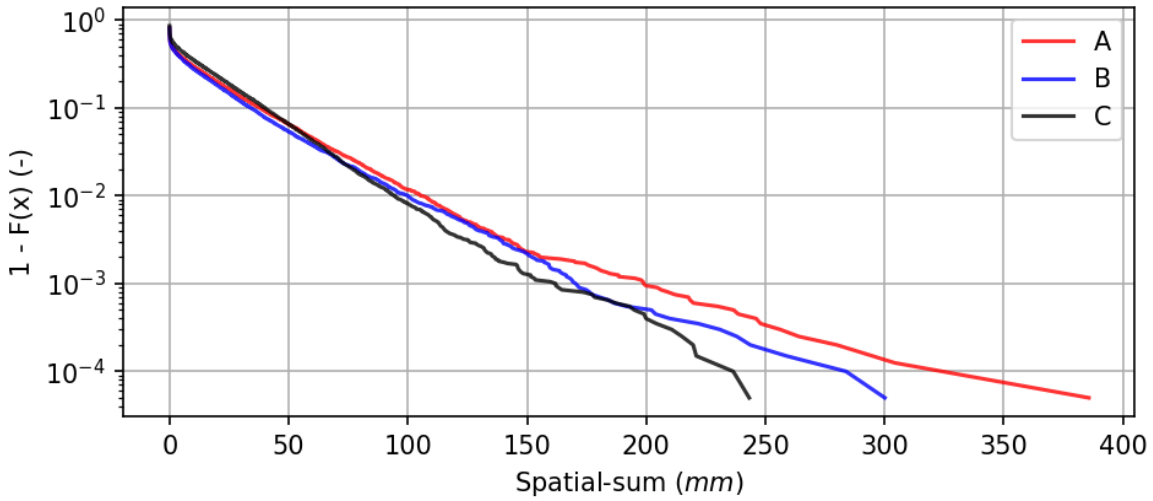


Figure 8.25: Comparision of the distributions of spatial-sums of daily precipitation for three sets of stations.

9 Results

So far, the variables precipitation and discharge, methods to simulate them as point time series and measures to evaluate their spatio-temporal behavior were described. Here, the results of all existing and new methods are shown, discussed and evaluated. Multi-site daily discharge is shown first and then daily precipitation. The goal is to find properties that are important to be reproduced by the simulated time series in order for them to be deemed usable in terms of representing extreme values.

Afterwards, results of a somewhat more practical case are presented. Precipitation, temperature and potential evapotranspiration are simulated simultaneously that later serve as inputs for the HBV. The properties of resulting model discharge are then evaluated. Additionally, this also serves as a validation for the *plausiblness* of the simulated model inputs. The hypothesis being that, if the simulated series used as model inputs produce output discharge that is similar to that by using the observed series as inputs then it can be stated with confidence that these series do incorporate realistic properties and can be, therefore, used to test scenarios of extreme cases.

Finally, the results of the hourly precipitation simulations are presented. This case of finer temporal resolution was investigated additionally, apart from the daily, to evaluate whether there was really a difference in the capabilities of the algorithms when presented with inputs that are predominantly composed of zeros.

9.1 Selection of time series for simulation

In Sec. 4.1 the precipitation and discharge gauges that are available in the study area were shown. Given the large number of them, not all could be simulated. Therefore, subsets depending on various criteria were selected for various regions to demonstrate the performance of all the simulation methods. These are detailed below.

For daily discharge, five groups of stations in the Neckar catchment for a time period of 55 years (1961-2015) each were selected. The groups included small and large head water sub-catchments whose surface areas varied between 100 and 2300 km^2.

For daily precipitation, four groups of five stations were selected from each corner of BW for the time period of 55 years (1961-2015) each such that their mean distance was about 30 to 40 km. For the case of far away stations, one group of five stations was selected such that their mean distance was about 161 km from each other for a time period 55 years (1961-2015). The description of these gauges was given in Chapter 4 Section 4.1 and Fig. 4.3.

For hourly precipitation, four groups of five stations from different corners of BW were selected for the time period of 6 years (2010-2015) such that their mean distance was about 30 to 40 km. One group of stations that were far away from each other (mean distance of about 160 km) was also selected. Only the case of the nearer stations is presented. The description of these gauges was given in Chapter 4 Section 4.1 and Fig. 4.4.

One hundred time series were simulated for each setting with 10,000 optimization iterations.

9.2 Objective functions used to simulate time series

In total, seven algorithms were described in Chapter 5 that may be used to simulate time series. A large number of combinations was possible given the number of methods and the objective functions. As it is not possible to show the results using all the methods, only a subset for each method are presented. Following are the details for each method.

The Amplitude Adjusted Fourier Transform (AAFT), Iterative Amplitude Adjusted Fourier Transform (IAAFT), Multivariate Iterative Amplitude Adjusted Fourier Transform (MIAAFT), Modified Multivariate Iterative Amplitude Adjusted Fourier Transform (MMIAAFT), Iterative Amplitude Adjusted Fourier Transform - Pearson's and Spearman's correlation 1 (IAAFT-PSC1) and Iterative Amplitude Adjusted Fourier Transform - Pearson's and Spearman's correlation 2 (IAAFT-PSC2) do not allow for simulation of time series that keep the required properties in a manner similar to Phase Annealing (PA), depending on the situation the resulting series can either be accepted or rejected. AAFT was the fastest method and needed only a single iteration to produce a time series that also had the cross-correlations to some extent. IAAFT required a slightly more but converged usually within 10 iterations to a solution where no improvement of the power spectrum could be achieved for the simulated series, here no cross-correlations were conserved. MIAAFT is very similar to IAAFT, the only difference was that the difference of the phase spectra were also conserved i.e., cross-correlations. It also converged within 25 iterations. MMIAAFT also took similar time.

The two variants of IAAFT-PSC, on the other hand required more computation as the αs for each series had to be optimized. The objective functions used for these were the Marginal Cumulative Periodogram ($CMPOW_p$) and Grade Cumulative Periodogram ($CMPOW_s$) and the pairwise cross-$CMPOW_p$ and $CMPOW_s$. All entropy related properties were not included, as they took too much time to compute on the given computing systems.

For PA, the most flexible algorithm among all of them, two settings were used. One where the lumped versions (introduced in Chapters 3 and 6) of the objective functions and the other where both the lumped and spectral versions were used. The two cases are referred to as *Lumped* and *Spectral* respectively in the figures. Here too, entropy related properties were not included while optimizing.

Finally, the measures described in Chapter 7 were only used as metrics to evaluate the quality independently. Here, only the results of d-dimensional cumulative marginal correlation periodogram ($CMCORR_p^d$) and d-dimensional cumulative grade correlation periodogram

($CMCORR_s^d$) are presented. All the other were demonstrated for observed and the IAAFT cases in the Chapters 7 and 8 already.

9.3 Quality assessment of simultaneous extremes

To assess whether the simulations have the the same characteristics of simultaneous extremes as those of the reference, two independent (i.e., not considered during optimization whatsoever) tests are performed. One being the comparison of the distributions of the sums (spatial-sums) while the other being conditional probabilities of multuplets and their subsets bringing extreme values of a given or higher return period. These are described below.

Extreme flows may be a result of high precipitation at a single or multiple locations and that also for a longer duration. All the three possibilities are investigated as follows. First, the distributions of the sums of precipitation or discharge using all series for each time step with their original resolution i.e., daily or hourly. The second case is where each time series is aggregated to higher temporal scales e.g., daily to weekly or hourly to six-hourly and then series of each point are evaluated independent of the others. The third being the spatial-sums of the temporally aggregated series of all series.

For evaluating the conditional probabilities of multuplets of series, the test is conducted in the following manner. For a given multivariate series of M stations, A sized combinations are drawn out of it. Then, the conditional probabilities of all the series having values of return periods or greater versus only a subset of them having the extreme value for the same return period are compared. For example consider five discharge time series having a length of 100 years each for the same time period. Select all the time steps where at least four of them have a flood of a return period of one year or higher and then among these time steps only those where all five of them have floods with return periods of one year or more. The probability of the second case will always be less than or equal to the first case (in general lesser). These probabilities are compared for reference and simulations. Ideally, the simulations should produce a cloud that contains the reference. If the reference is significantly away from the cloud of simulations, then it can be concluded that the simulations are not very representative of the reference. Otherwise, they are accepted. The problem with such a test is that as the size of the combinations to test grows, the number of points that can be compared reduces very fast. For this reason, the results are shown for cases where combinations of four stations were drawn out of the five and then the probabilities are compared for the case when the four bring a value of a return period of two years or higher versus the case when only three of them do the same for the daily series while for the hourly the return period of six months is used due to the shorter data length.

9.4 Daily discharge

Fig. 9.1 shows the auto-ρ_p functions for time series of a randomly selected gauge. AAFT shows a larger loss as compared to the others. IAAFT-PSC2 and PA (Spectral) also show a small bias. The rest perform acceptably. Fig. 9.2 shows the auto-Spearman's correlation

functions (auto-ρ_s) functions. Again, AAFT shows the largest loss comparatively. IAAFT MIAAFT and MMIAAFT also show a much larger bias in that they could not simulate series that have higher correlations consistently. PA shows the best results but it could seen as over-fit. In that regard, IAAFT-PSC2 seems to show a much more realistic variance. Coming to copula order asymmetry (AO) and copula directional asymmetry (AD) (Fig. 9.3 and 9.4 respectively), none perform in an acceptable manner but only get closer to the reference. IAAFT-PSC2 does get very close, but still shows a bias for AO.

Pairwise cross-ρ_p, ρ_s, AO and AD are shown in Fig. 9.5, 9.6, 9.7 and 9.8 respectively for the various methods. It should be noted that IAAFT was run with cross-correlation conservation for the annual and seasonal cycles only, which are quiet large for the case of discharges here, that consequently resulted in significant cross-correlations. The rest of the methods perform well. It is interesting to note that the cross-AO and AD by AAFT are very close to zero compared to the others, it is what is expected from a generator that is supposed to produce a symmetric dependence. Another important aspect to notice is that MIAAFT and MMIAAFT overestimate the cross-ρ_p and AO consistently with a few exceptions. This has bearing on the extremes synchronizing too much in space, which will certainly lead to larger sums at any given time step compared to the reference.

All performed acceptable for $CMPOW_p$ (Fig. 9.9). As AAFT did not conserve the auto-ρ_s, its $CMPOW_s$ (Fig. 9.10) did not perform well either. The simpler methods show large biases for the case of $CMPOW_{AO\_AUTO}$ (Fig. 9.11) for the lead of 1 step but all perform well for the $CMPOW_{AD\_AUTO}$ (Fig. 9.12). All these have implications for the overall dynamics, but their effect on the behavior of the extremes have to be independently evaluated (shown next).

For the various methods Fig. 9.13 and 9.14 show the $CMCORR_p^d$ and $CMCORR_s^d$ spectra respectively. Note that only the contributions from the frequencies whose periods are less than one year time are considered here. Important differences can be observed. For $CMCORR_p^d$, AAFT lacks the long term frequency strength but has a greater strength in the higher frequencies. IAAFT has little strength in the high frequency range. IAAFT-PSC2 has high frequency strength that is even greater than the reference while PA (Lumped) behaves in a manner similar to that of IAAFT but shows more loss. The rest show good performance. Here, it can be foretold already the the behavior of simultaneous extremes i.e., all the methods that perform well in the high frequency range will also produce results that are closer to those of the observed.

Coming to the most important metric which are distributions of the sums of extremes values, Fig. 9.15 shows them for the various methods using the sums of the daily discharges for all the five considered stations. The figures are made in a way to highlight the upper 10% of the distributions. Excluding IAAFT (mean reduction of about 40%), all of them show no strong biases compared to the observed case. Fig. 9.16 shows the sums of the weekly-mean distributions. Again, no bias even for IAAFT but a problem can be seen with IAAFT-PSC2 which increases the values by about 30%. Similarly, Fig. 9.17 shows the distribution of the weekly-mean of one randomly selected station among the five, it also shows no clear bias for any of them except IAAFT-PSC2. Here, an important conclusion ca be drawn i.e., the overall cross asymmetries are not related to the simultaneous extremes. For example,

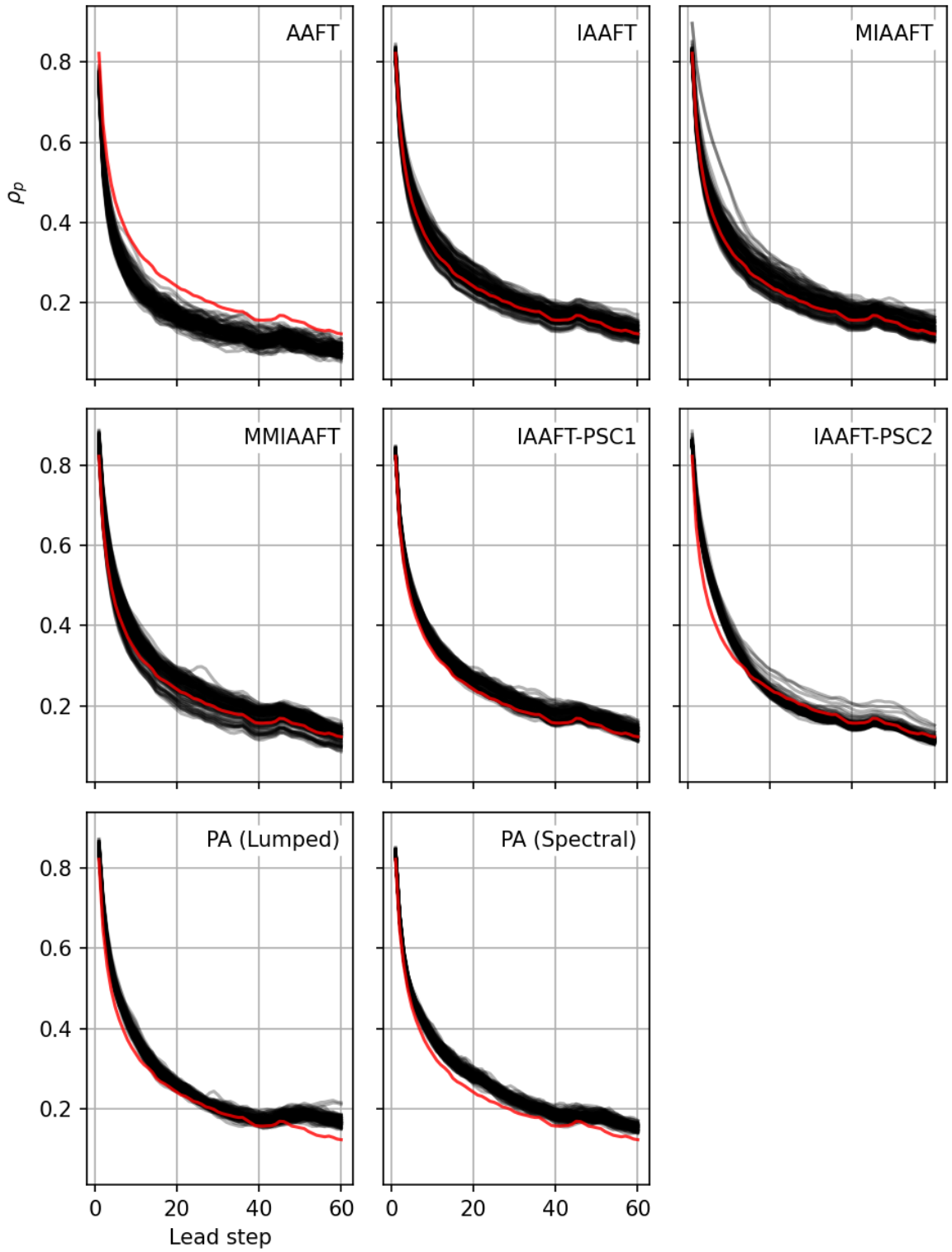


Figure 9.1: Pearson auto-correlation functions using various methods for daily discharge. In red are those of the observed while the black are using simulation.

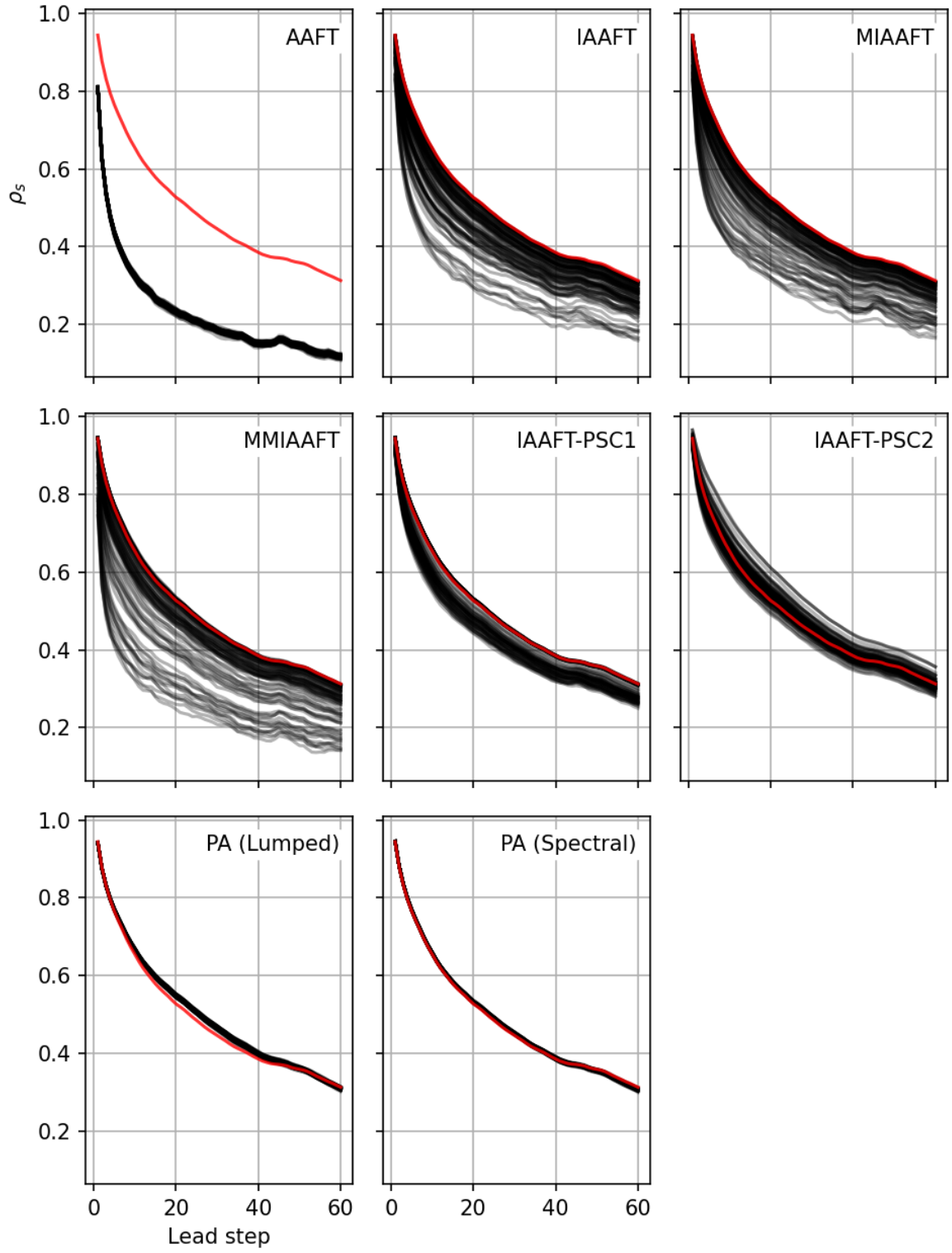


Figure 9.2: Spearman auto-correlation functions using various methods for daily discharge. In red are those of the observed while the black are using simulation.

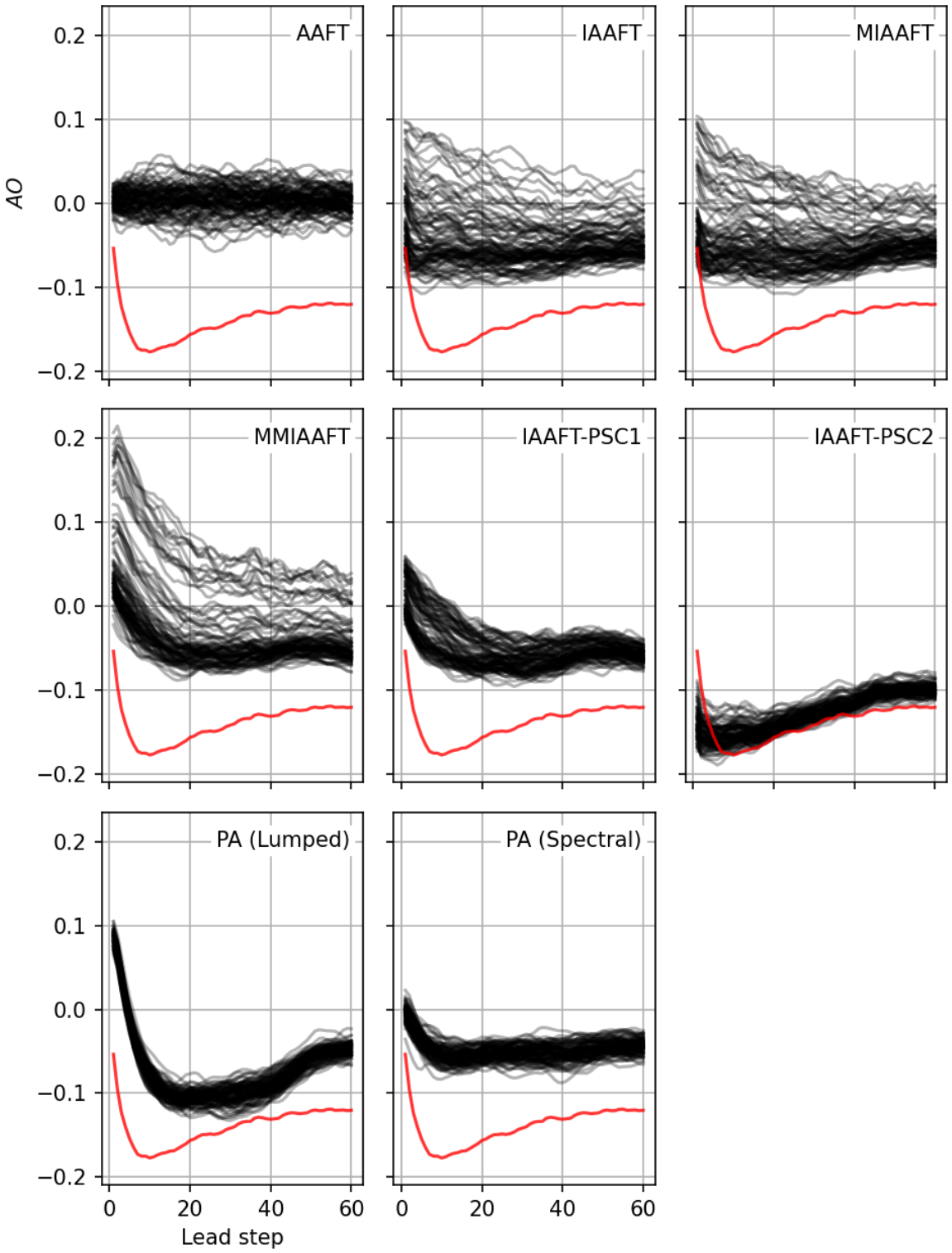


Figure 9.3: Auto-order asymmetry functions using various methods for daily discharge. In red are those of the observed while the black are using simulation.

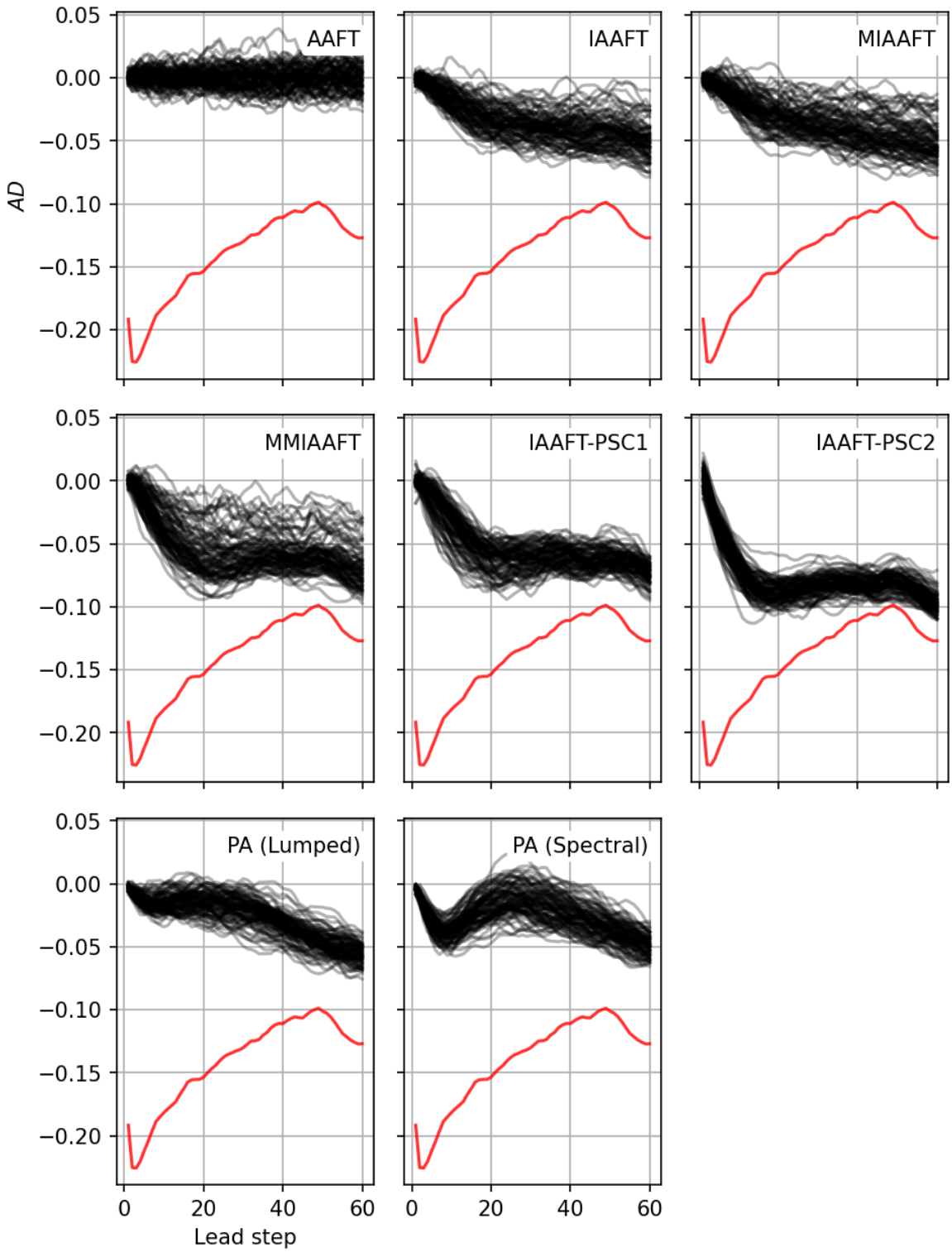


Figure 9.4: Auto-directional asymmetry functions using various methods for daily discharge. In red are those of the observed while the black are using simulation.

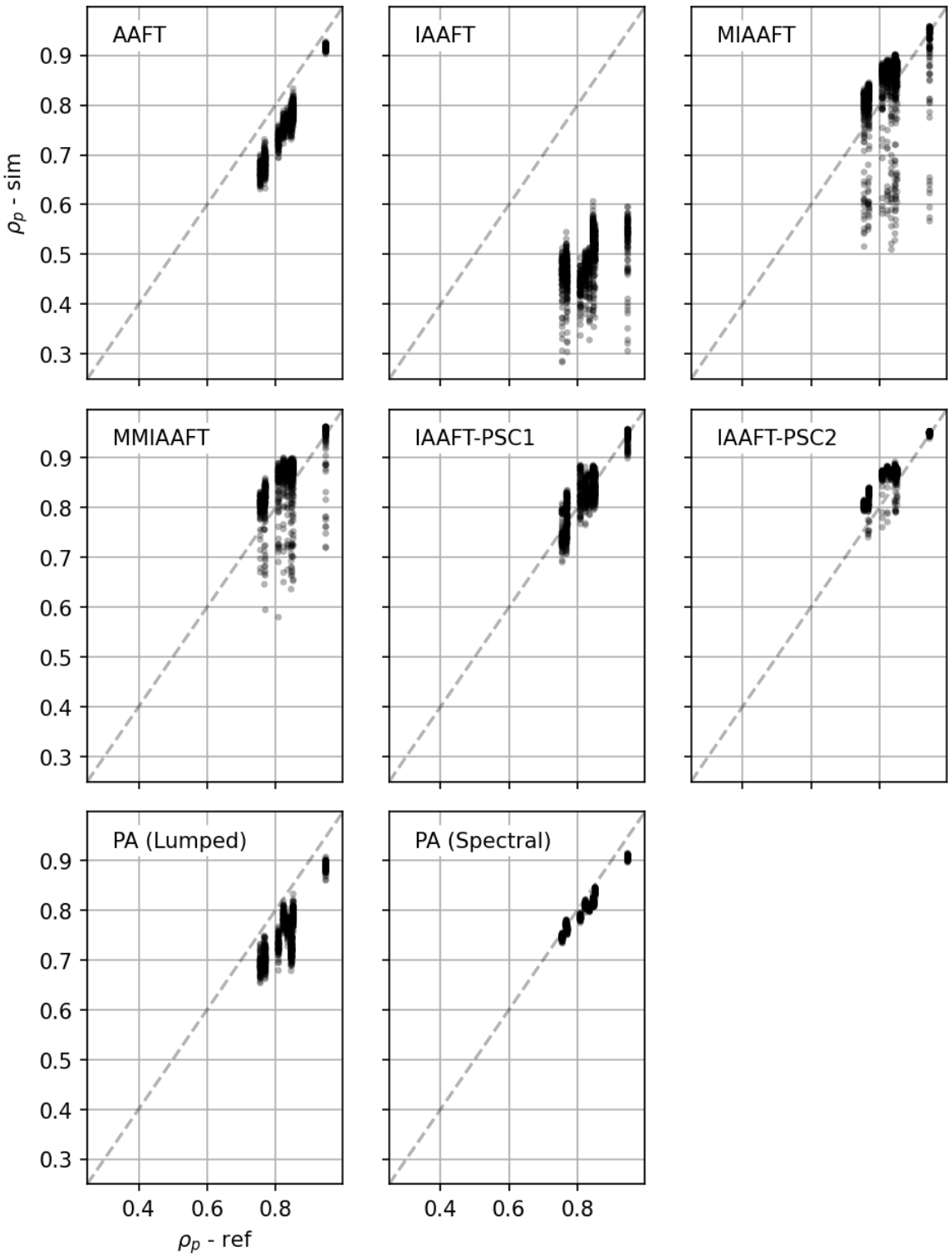


Figure 9.5: Pairwise Pearson cross-correlations using various methods for daily discharge.

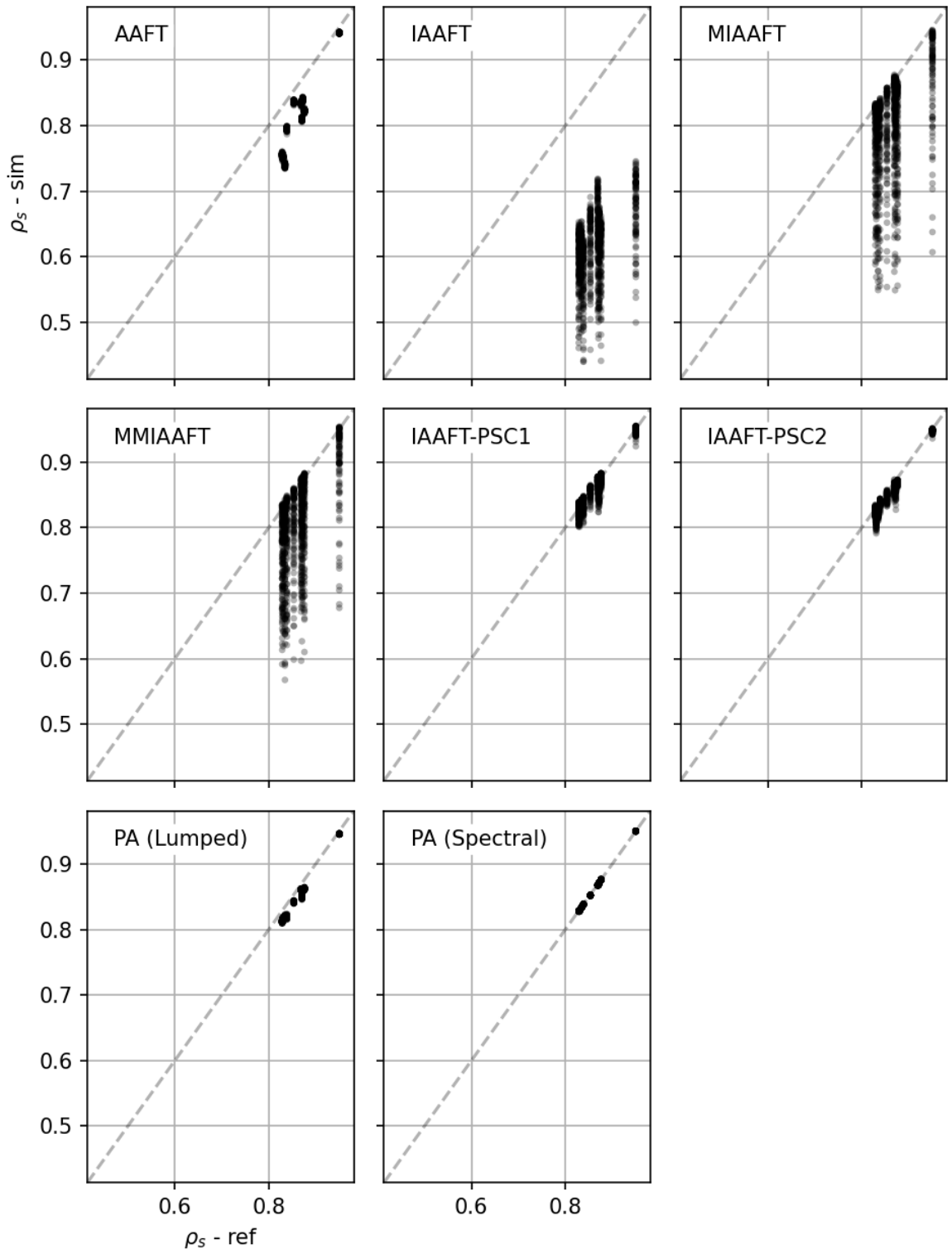


Figure 9.6: Pairwise Spearman cross-correlations using various methods for daily discharge.

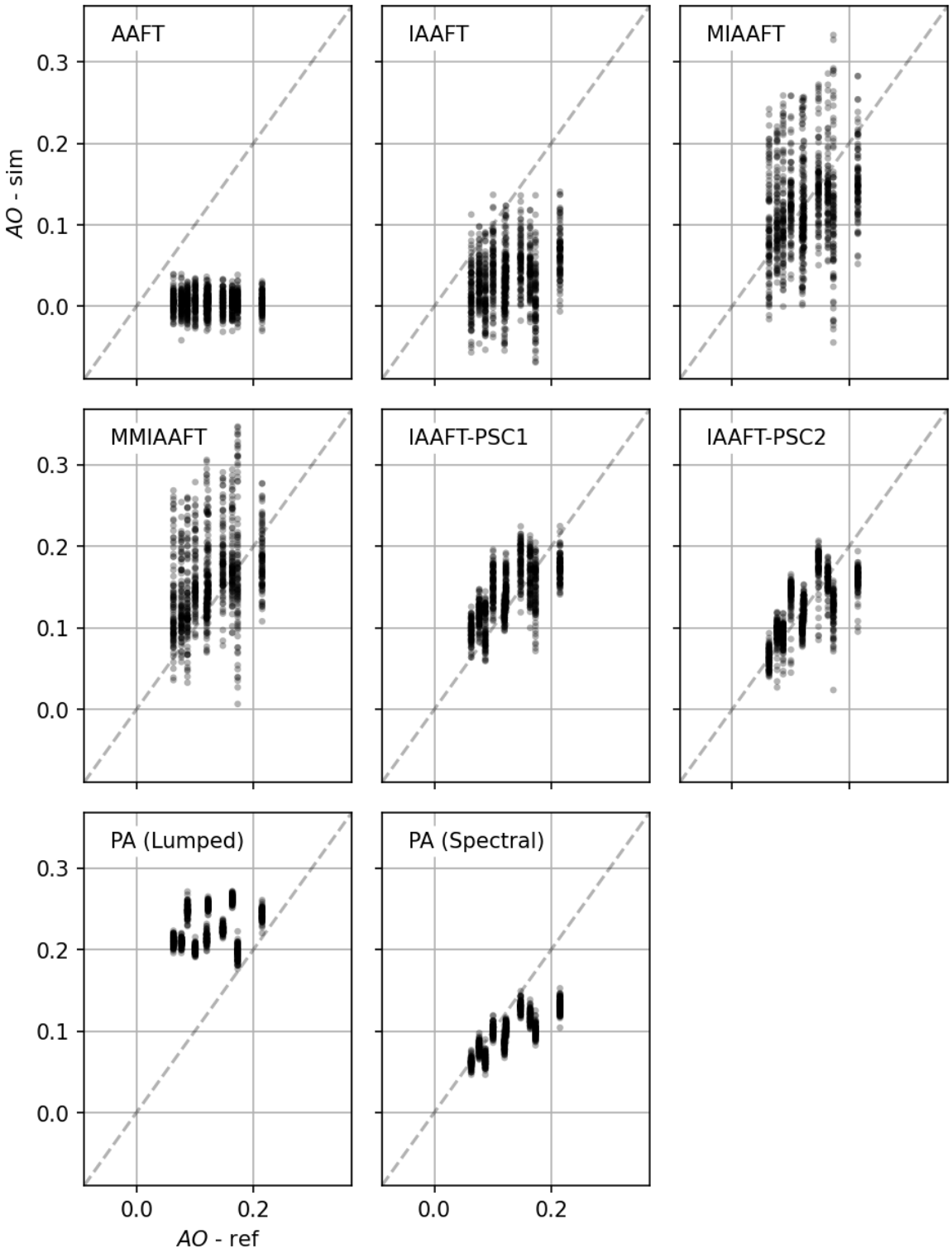


Figure 9.7: Pairwise cross-order asymmetries using various methods for daily discharge.

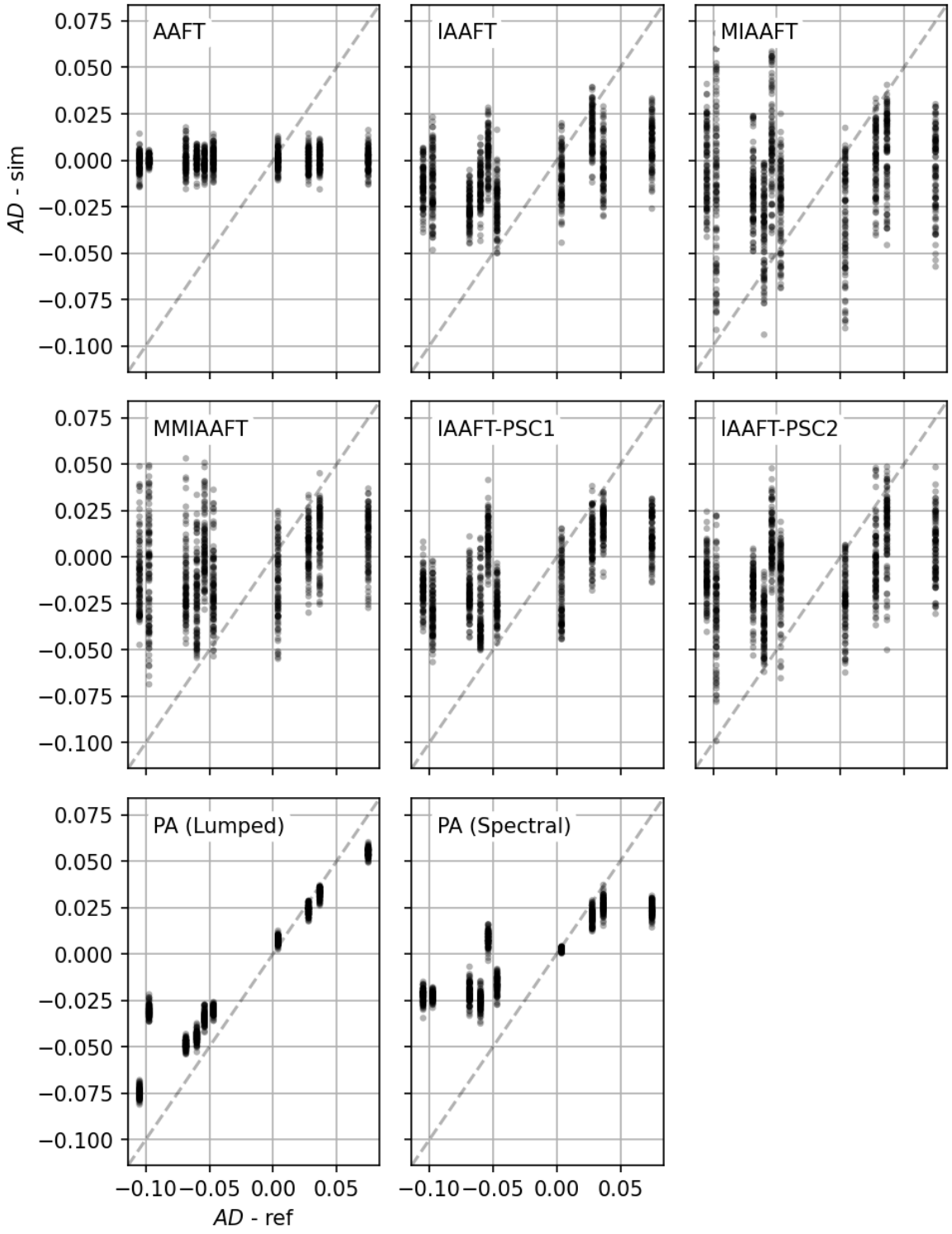


Figure 9.8: Pairwise cross-directional asymmetries using various methods for daily discharge.

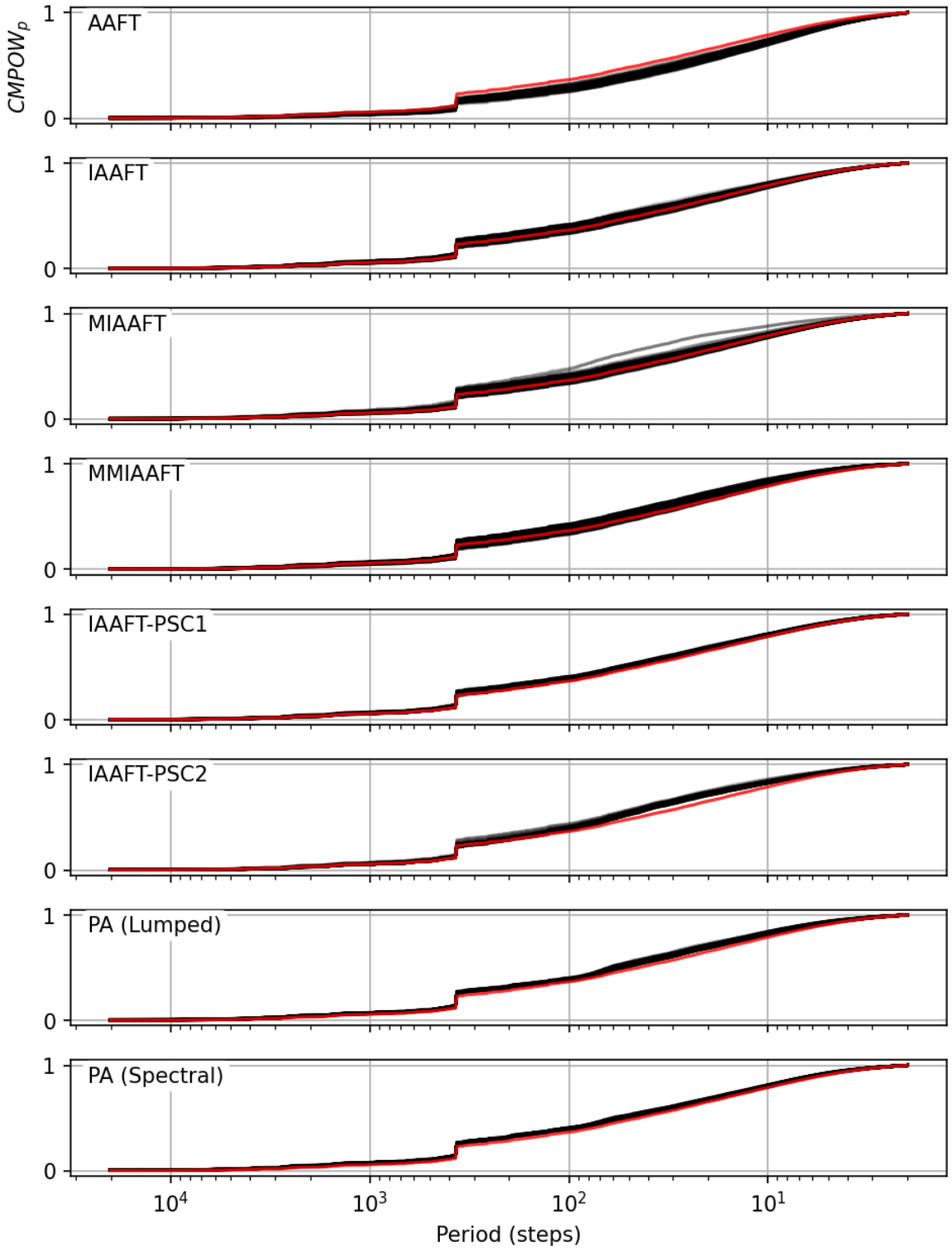


Figure 9.9: $CMPOW_p$ for various methods for daily discharge. Red are observed while black are the simulations.

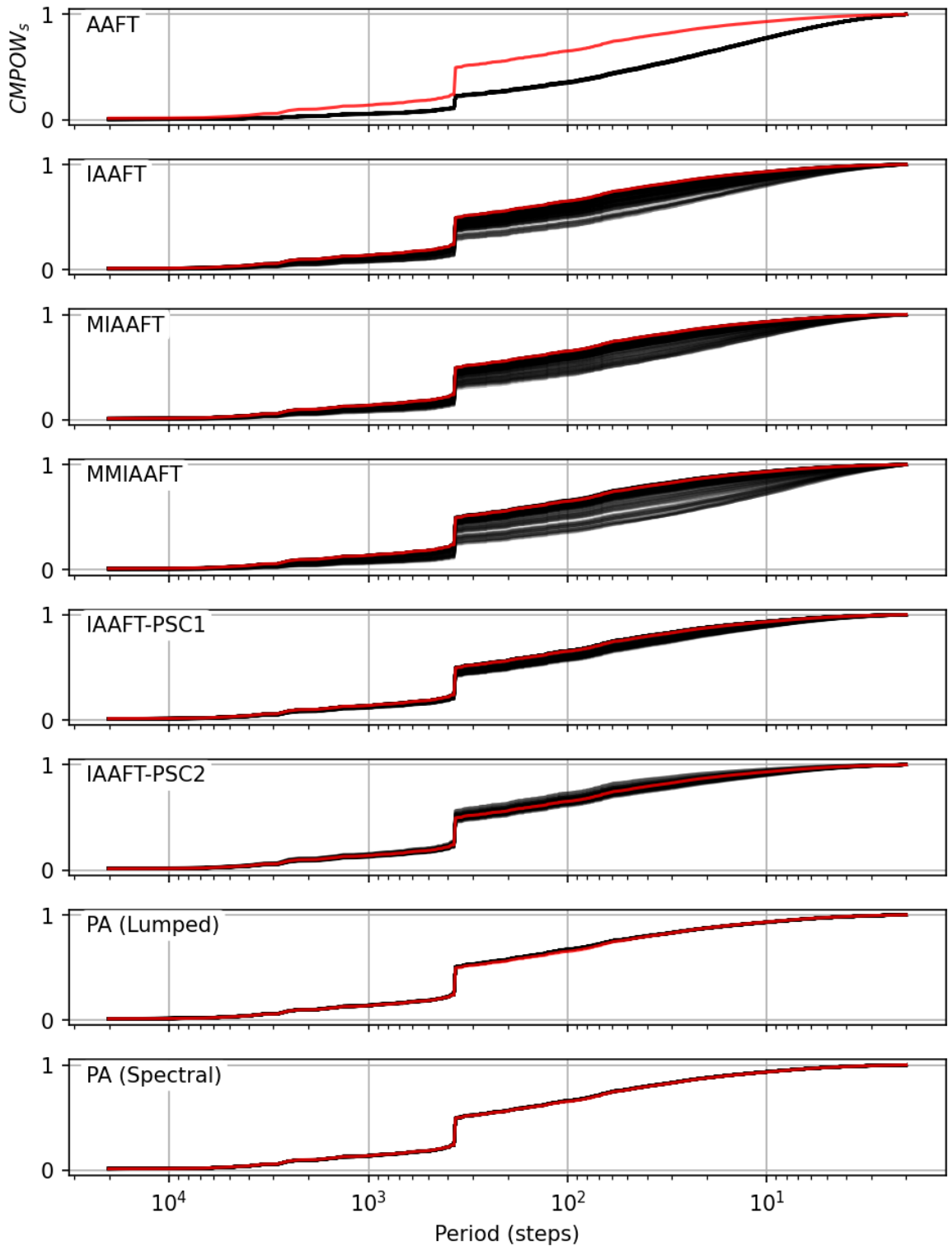


Figure 9.10: $CMPOW_s$ for various methods for daily discharge. Red are observed while black are the simulations.

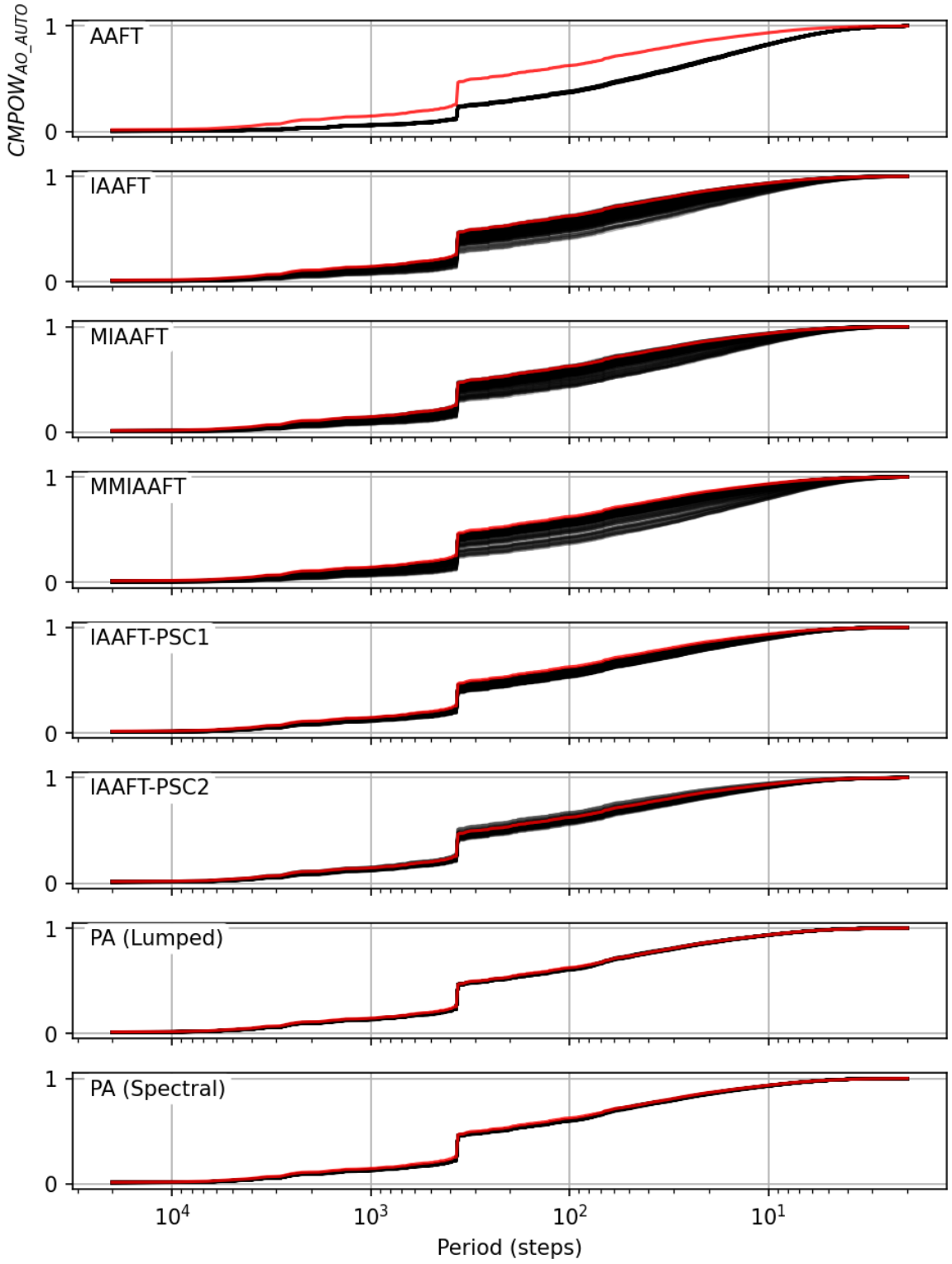


Figure 9.11: $CMPOW_{AO\_AUTO}$ with a lead step of 1 for various methods for daily discharge. Red are observed while black are the simulations.

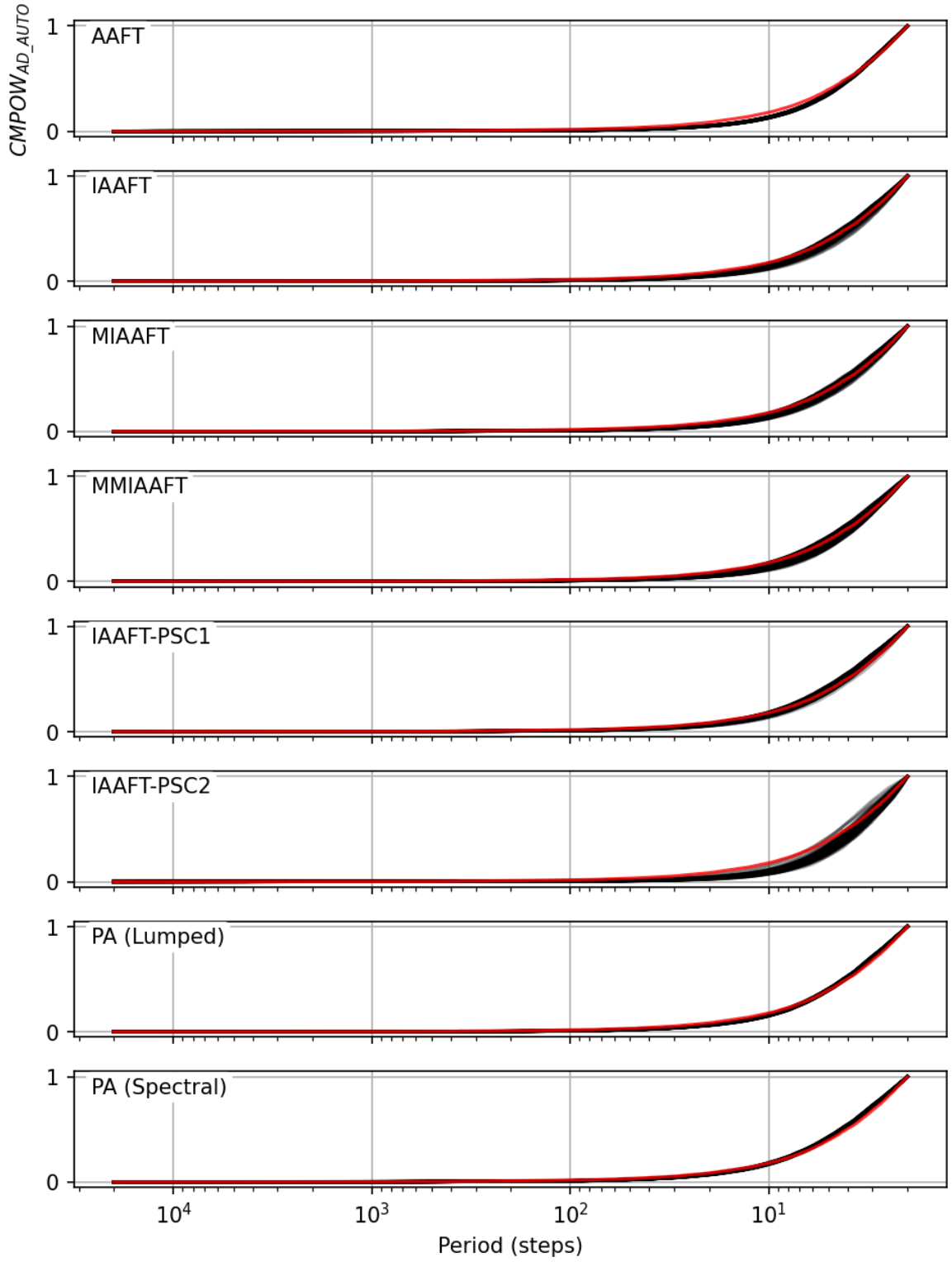


Figure 9.12: $CMPOW_{AD\_AUTO}$ with a lead step of 1 for various methods for daily discharge. Red are observed while black are the simulations.

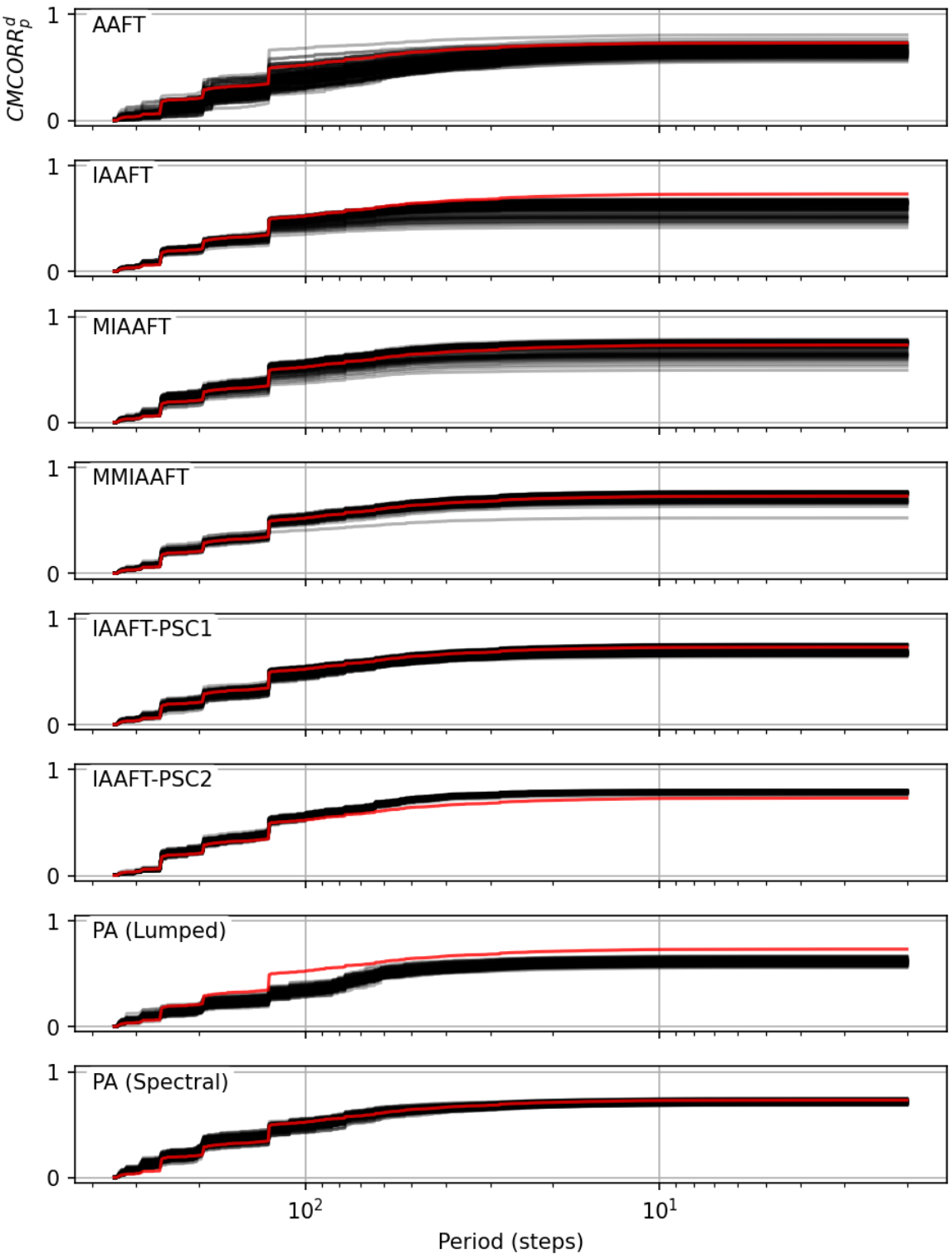


Figure 9.13: $CMCORR_p^d$ using various methods for daily discharge. Red are observed while black are the simulations.

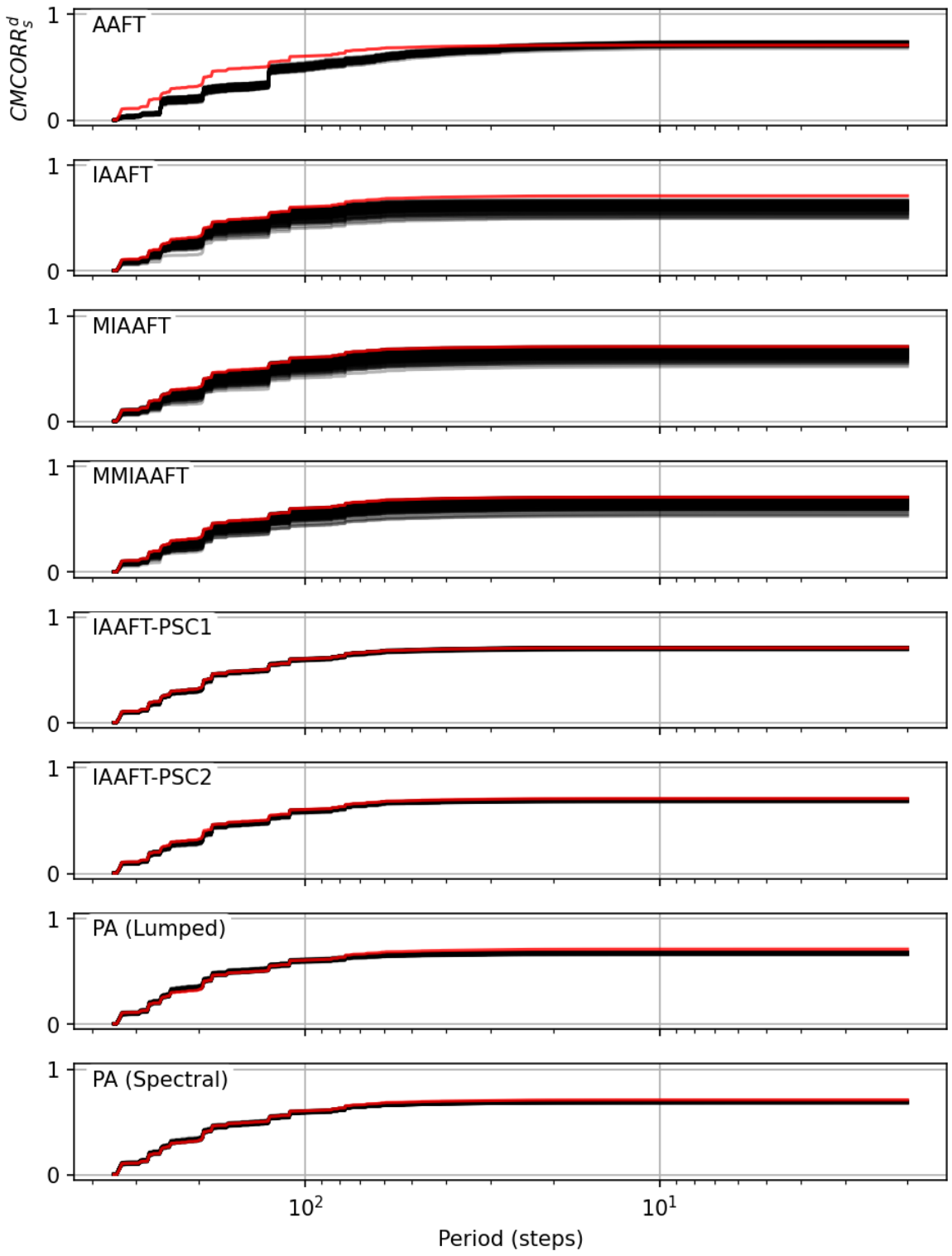


Figure 9.14: $CMCORR_s^d$ using various methods for daily discharge. Red are observed while black are the simulations.

AAFT showed practically, zero cross asymmetries while IAAFT did have asymmetries that had more variability comparatively. MIAAFT and MMIAAFT had pairwise cross-ρ_p values that were overestimated, and their daily spatial-sum distributions show relatively larger values as well. This means, that the extremes are more related to the measures defined on the marginal's domain rather than the rank domain. Another difference to notice is that all methods except for IAAFT slightly overestimate the spatial-sums, but not all overestimate the cross-ρ_p and all except AAFT do overestimate the auto-ρ_p. This all points to the fact that the spatial-sums are strongly related to both auto- and cross-ρ_p.

Fig. 9.18 shows the conditional probabilities of cases where at least three out of four stations produced a value of a return period of two year or more versus the case where all of them produced the same. Here, it can be clearly seen that the methods that consistently produced higher spatial-sums also consistently overestimate the conditional probabilities compared to observations. AAFT and IAAFT-PSC1 perform very well with PA (Lumped) being the worst.

9.5 Daily precipitation

Fig. 9.19 shows the auto-ρ_p functions for time series of a randomly selected gauge. All perform acceptably but a slight underestimation for AAFT and some overestimation using MIAAFT and MMIAAFTis visible. Fig. 9.20 shows the auto-ρ_s functions. Similar to discharge, AAFT shows a larger loss as compared to the others. IAAFT and MIAAFT also suffers from a small bias and so does PA. The rest perform acceptably. Coming to AO and AD (Fig. 9.21 and 9.22 respectively), all perform in an acceptable manner with the exception of AAFT. IAAFT-PSC1 overestimates AO slightly. All methods were unable to reach the very small AD of the first few lead steps.

Looking at the pairwise cross-properties (Fig. 9.23, 9.24, 9.25 and 9.26), AAFT shows some underestimation for ρ_p. MIAAFT and MMIAAFT similar to discharge, also show strong bias towards higher values while the rest perform as well. IAAFT was run in a setting to keep only the cross-correlations of the annual and seasonal cycles, as they are very small for precipitation, their effects are not observable. Only IAAFT-PSC1 shows good performance for the cross-ρ_s. For cross-AO, all methods show bias but IAAFT-PSC1 performs better than the rest. For the case of cross-AD, MIAAFT performs the best. MMIAAFT and IAAFT-PSC1 do stay close to the reference more than the others.

Similar to discharge, all performed well for $CMPOW_p$ (Fig. 9.27). As AAFT did not conserve the auto-ρ_s so much, its $CMPOW_s$ (Fig. 9.28) did not perform well either. No large biases for the case of the cumulative power spectrum of auto-order asymmetry ($CMPOW_{AO\_AUTO}$; Fig. 9.29) for the lead of 1 step can be observed and all perform well for the Cumulative power spectrum of auto-directional asymmetry ($CMPOW_{AD\_AUTO}$; Fig. 9.30). These results show, in contrast to daily discharge, that precipitation is more suited for simulation with these algorithms due to its more Gaussian-like behavior in space-time.

For the various methods, Fig. 9.31 and 9.32 show the $CMCORR_p^d$ and $CMCORR_s^d$ spectra respectively. Note that only the contributions from the frequencies whose periods are less

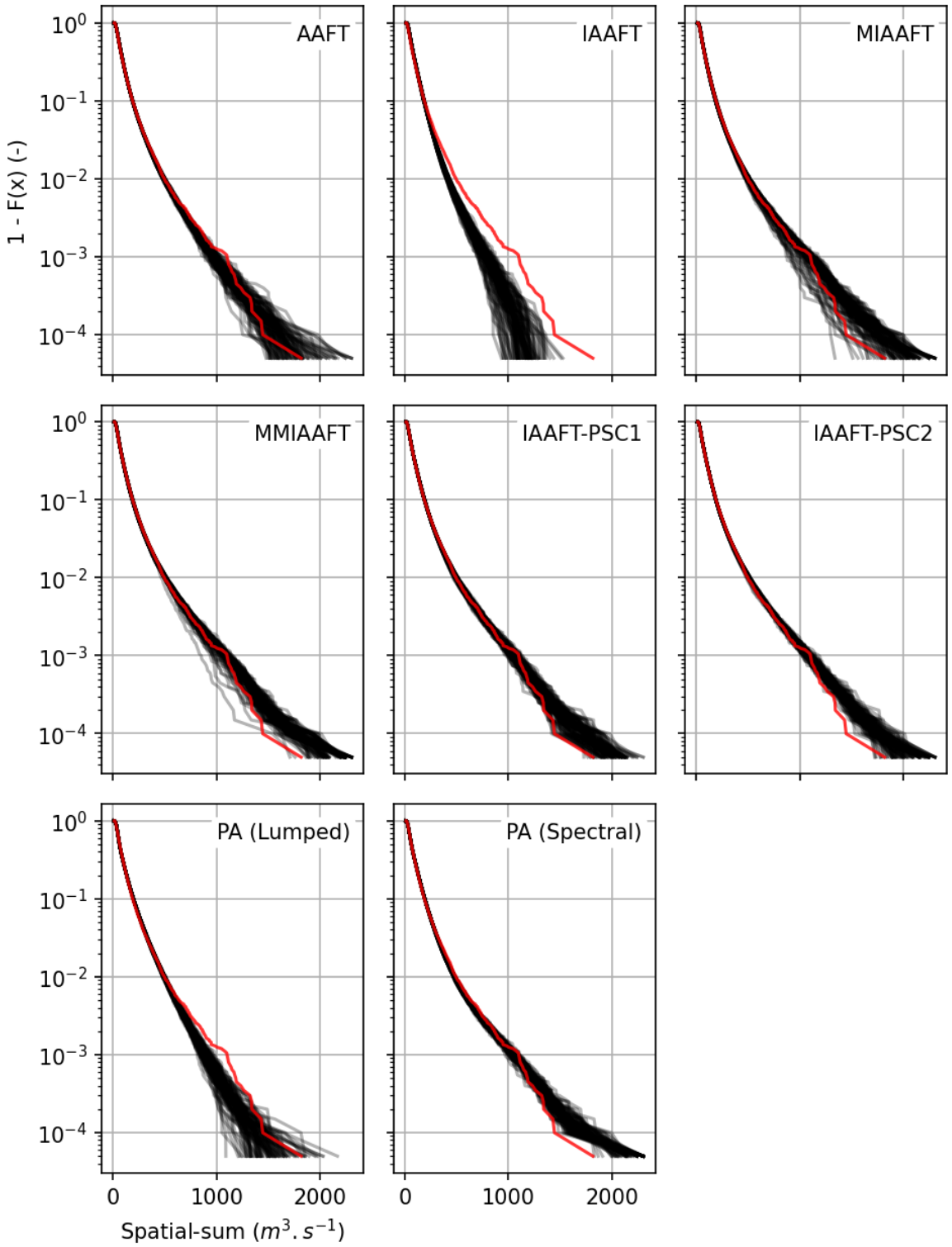


Figure 9.15: Distribution of daily spatial-sums of discharge for five large head water catchments in the study area. Red are observed while black are the simulations.

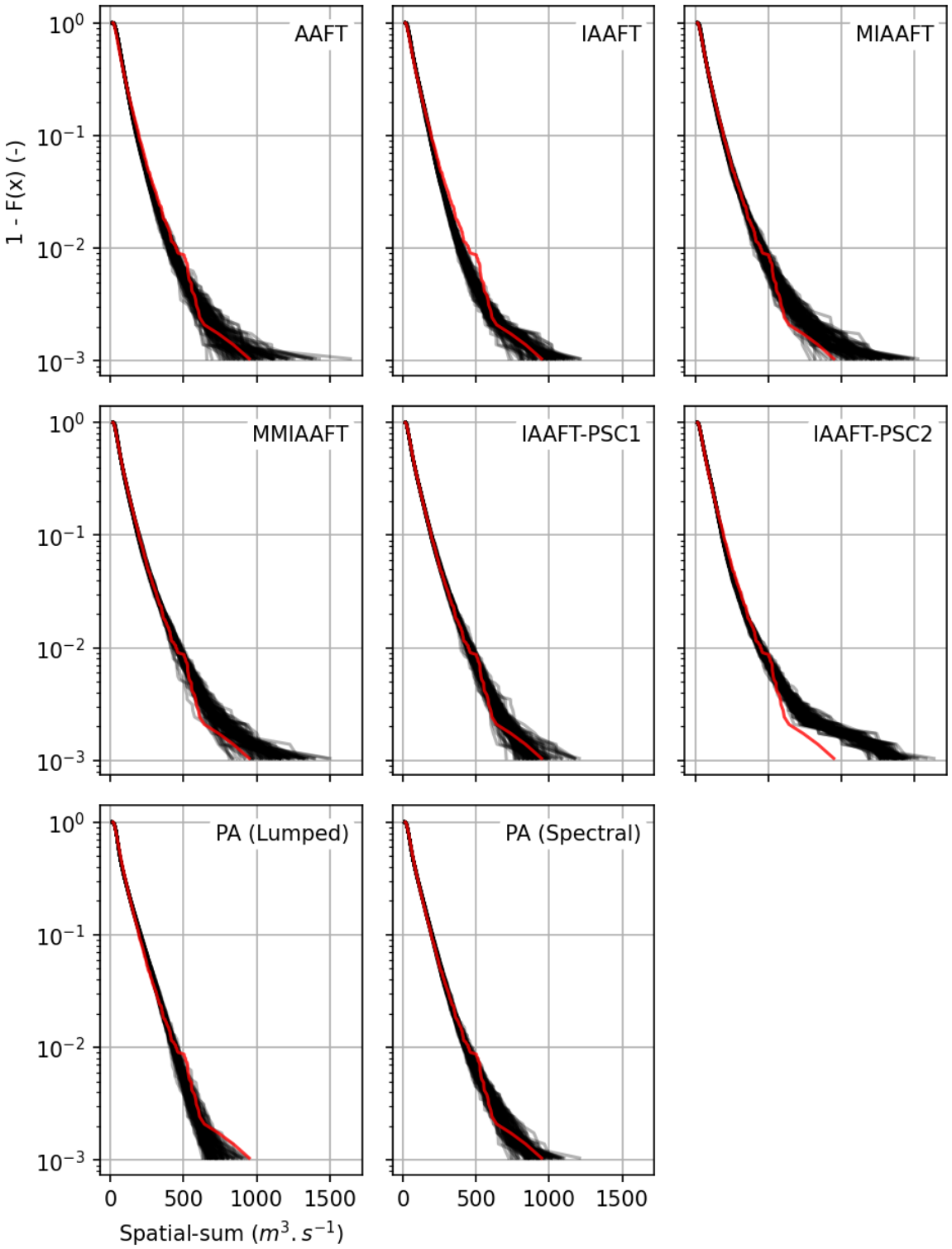


Figure 9.16: Distribution of weekly spatial-sums of discharge for five large head water catchments in the study area. Red are observed while black are the simulations.

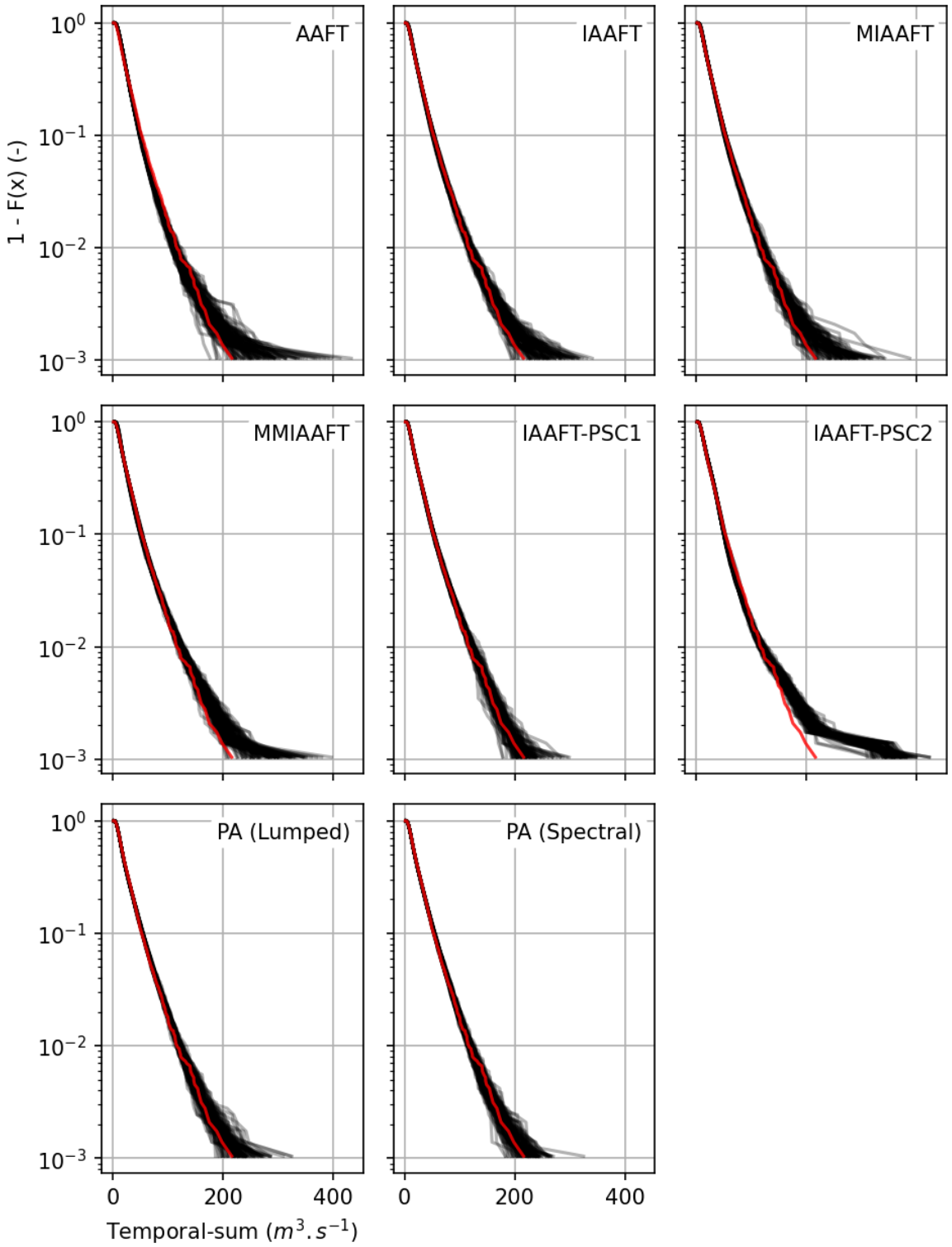


Figure 9.17: Distribution of weekly mean discharge for randomly selected head water catchment in the study area. Red are observed while black are the simulations.

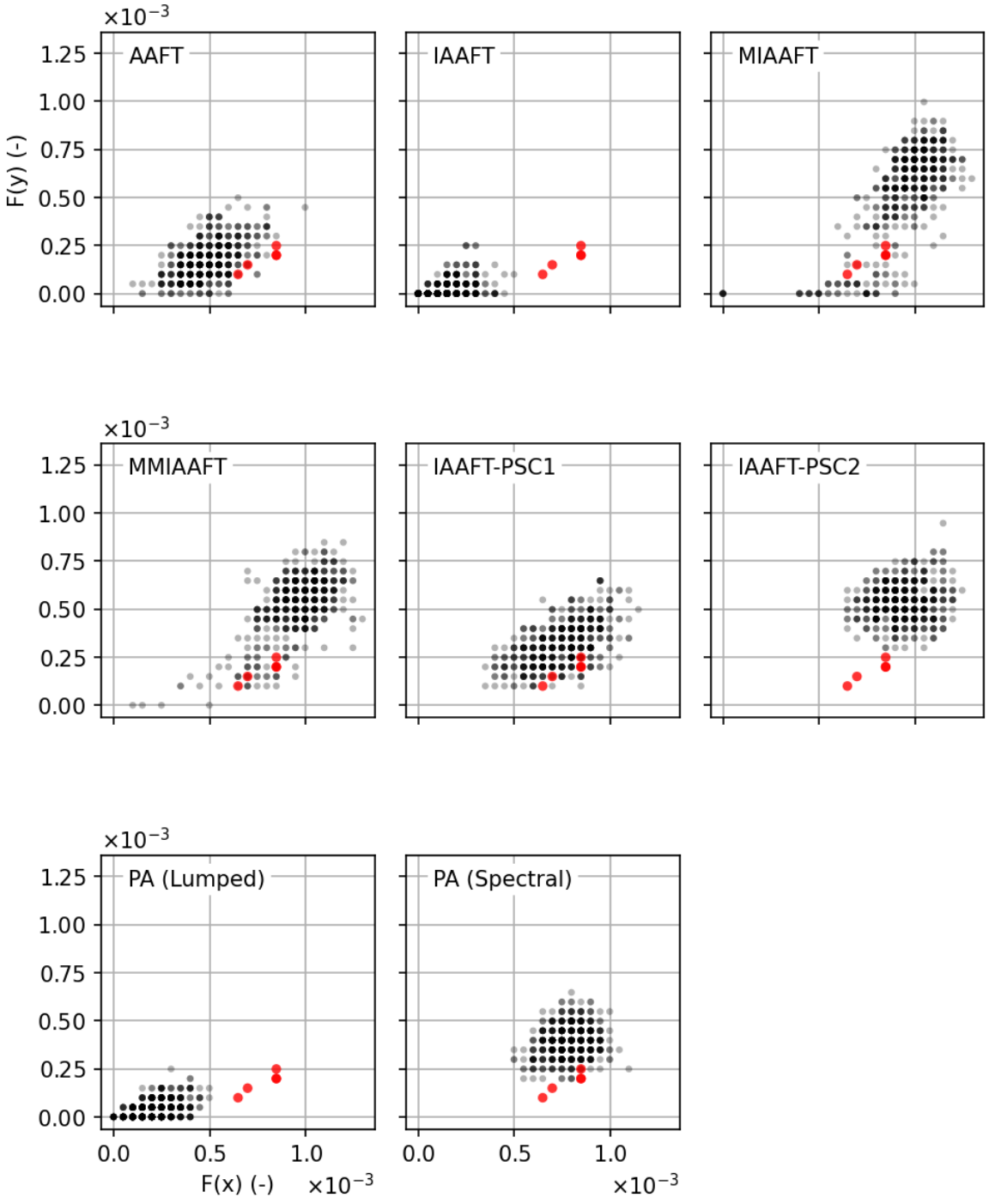


Figure 9.18: Conditional probability comparisons for daily discharge of three out of four have values with return periods of two or greater (horizontal axis) versus all of them having values equal or greater than two years (vertical axis) at the same time step.

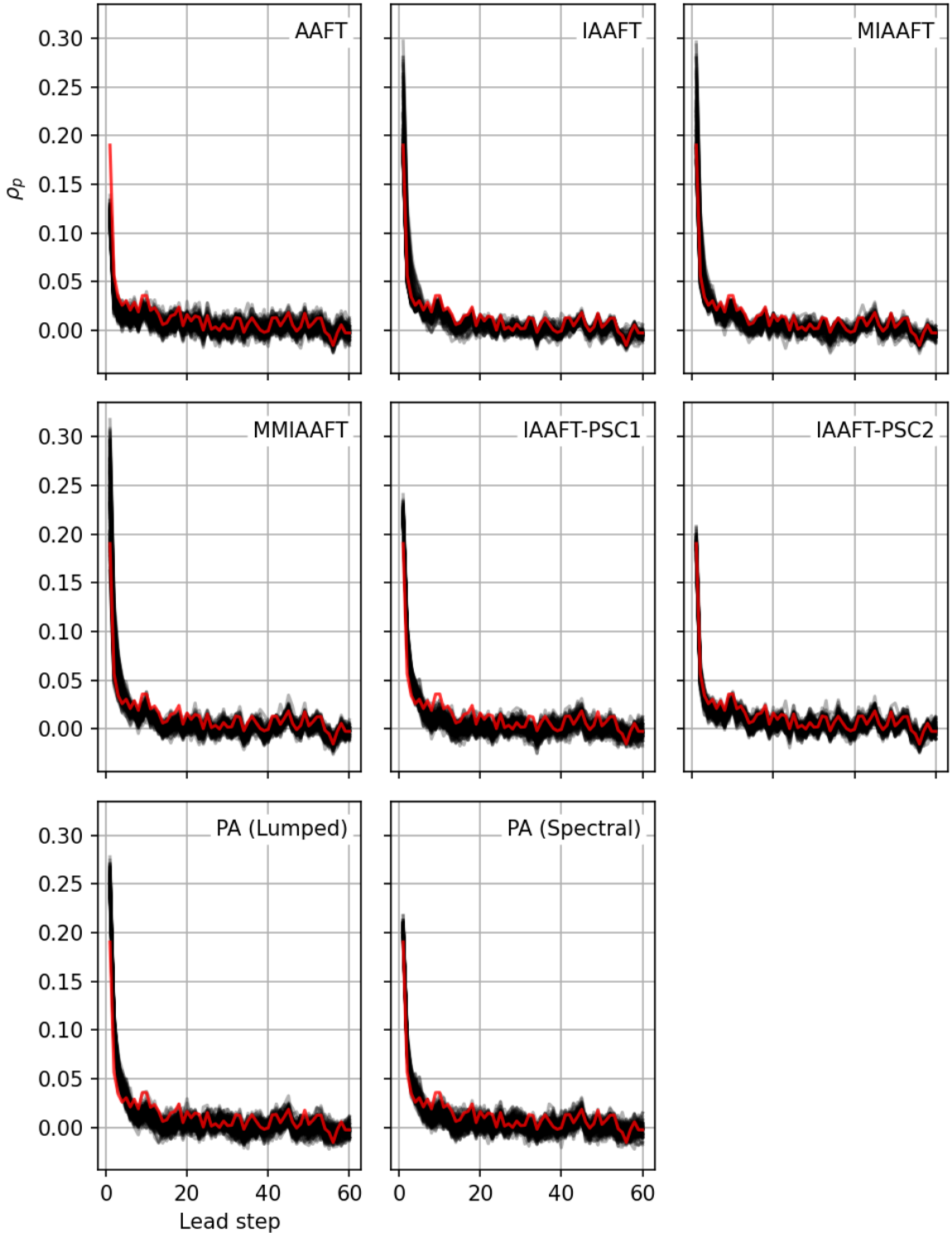


Figure 9.19: Pearson auto-correlation functions using various methods for daily precipitation. In red are those of the observed while the black are using simulation.

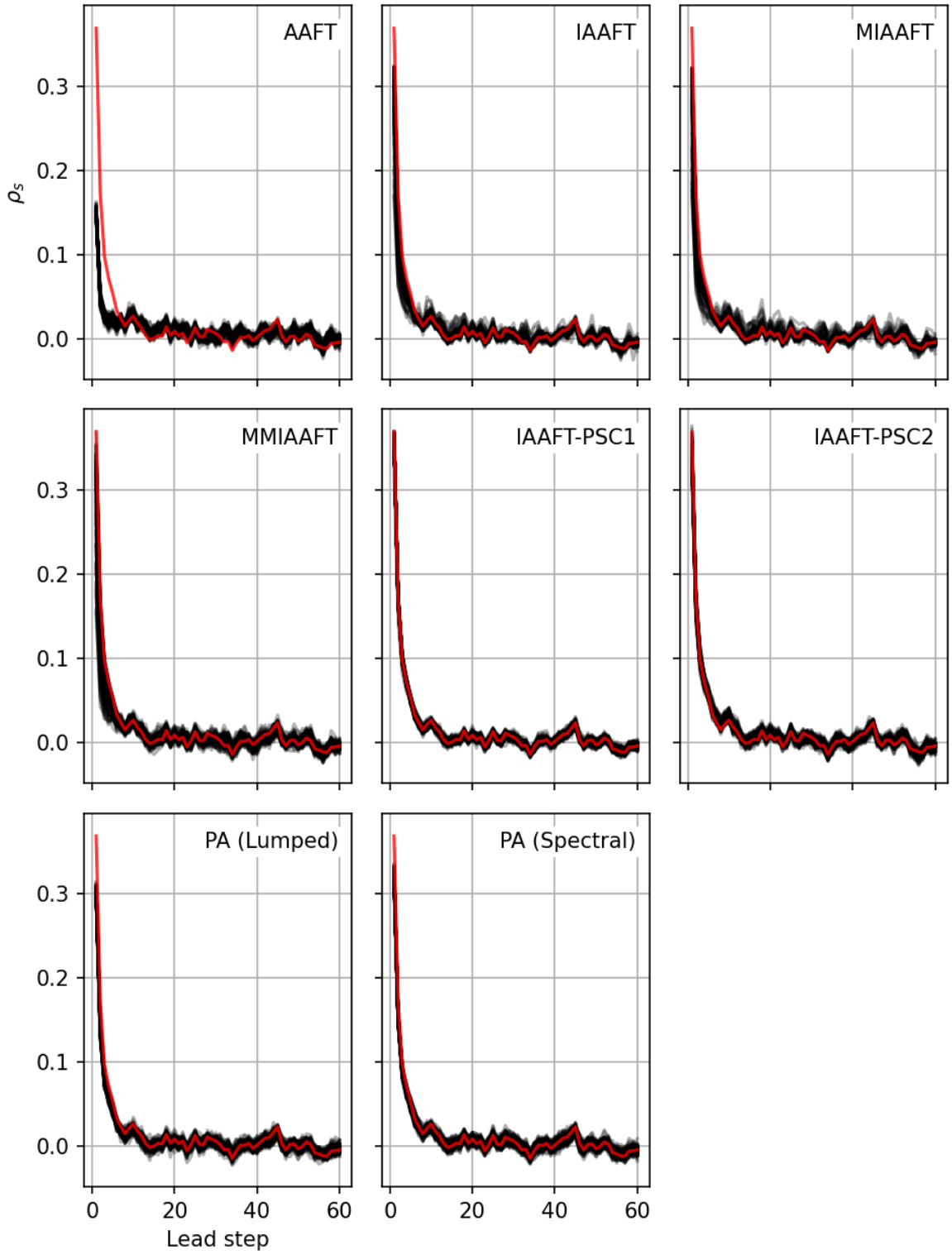


Figure 9.20: Spearman auto-correlation functions using various methods for daily precipitation. In red are those of the observed while the black are using simulation.

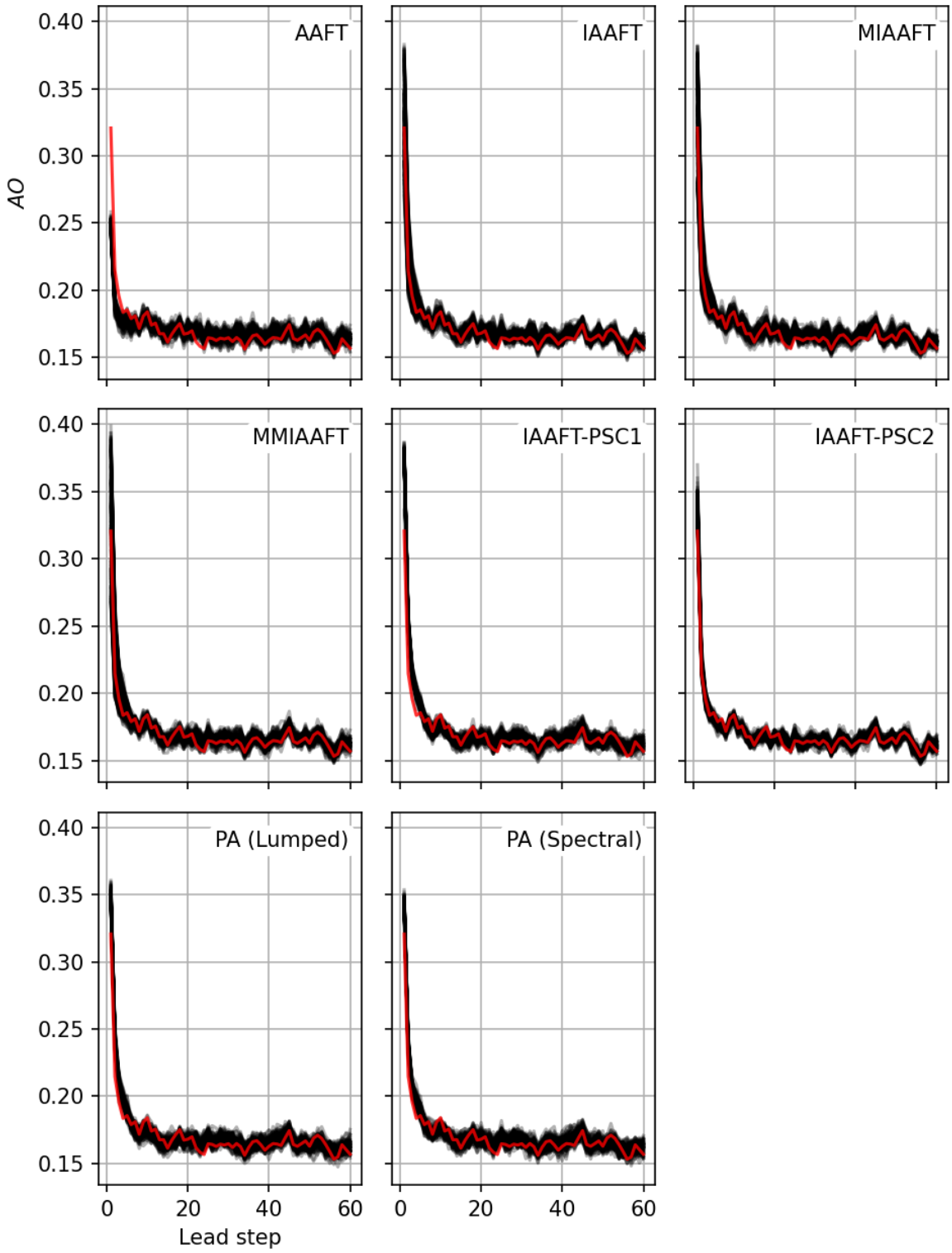


Figure 9.21: Auto-order asymmetry functions using various methods for daily precipitation. In red are those of the observed while the black are using simulation.

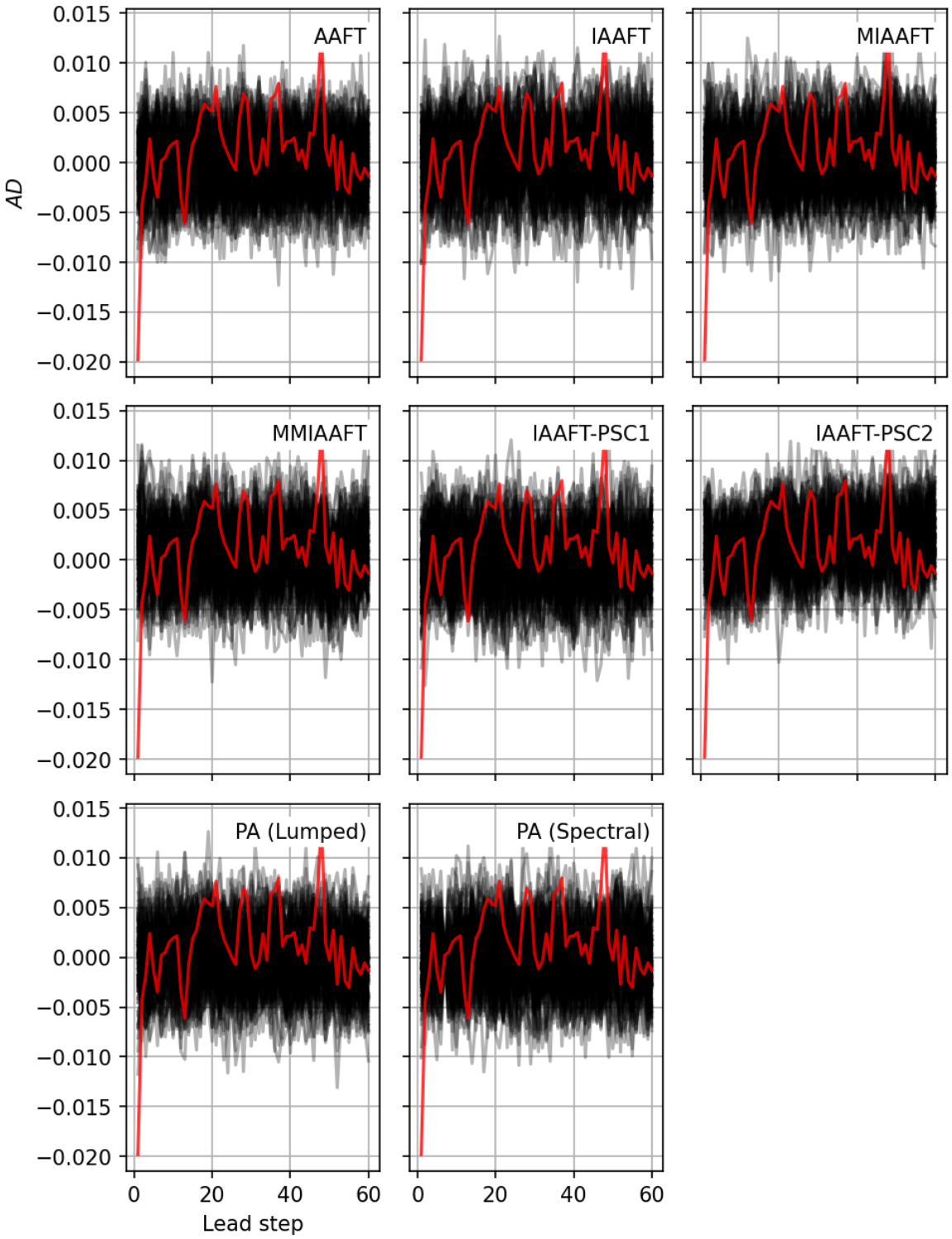


Figure 9.22: Auto-directional asymmetry functions using various methods for daily precipitation. In red are those of the observed while the black are using simulation.

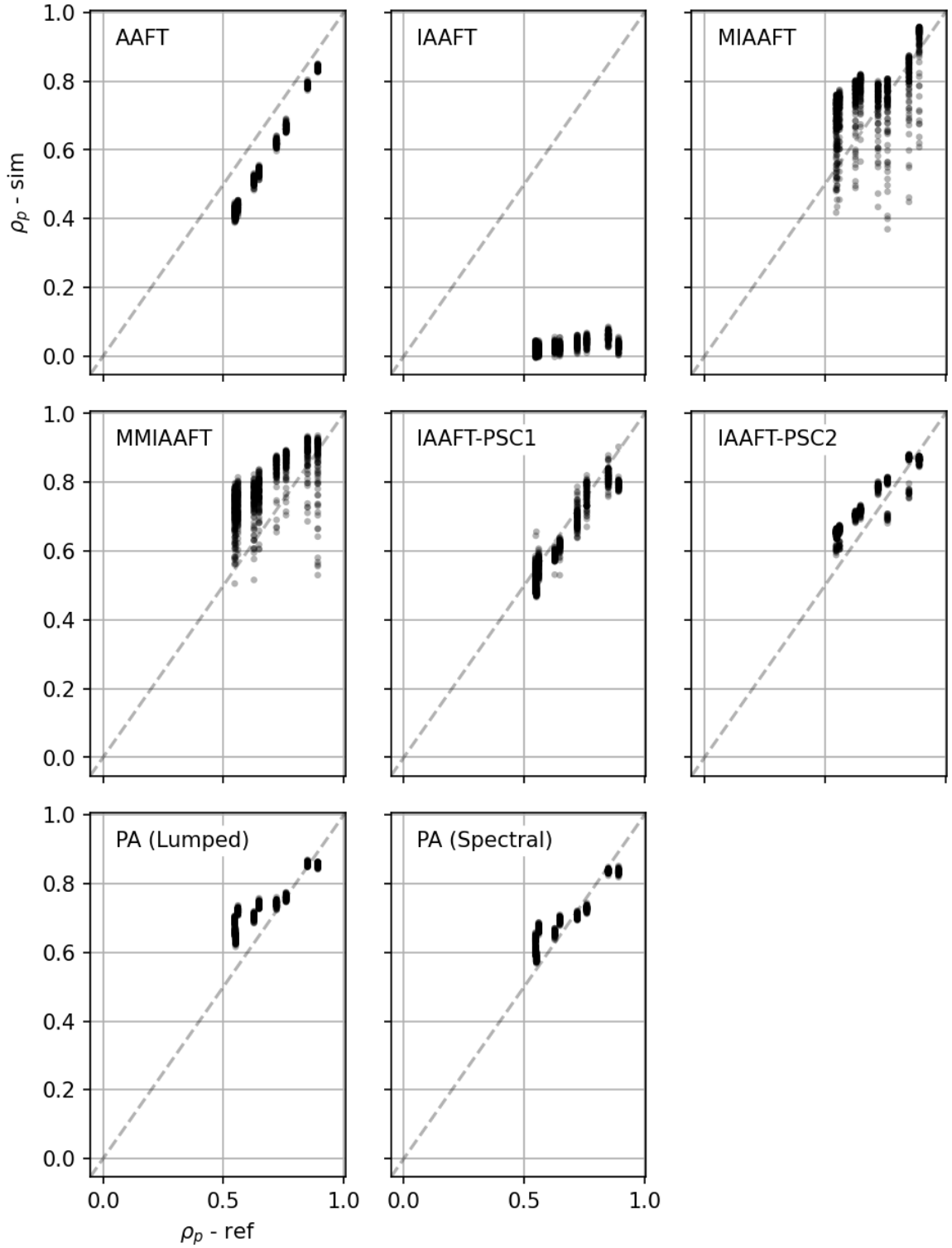


Figure 9.23: Pairwise Pearson cross-correlations using various methods for daily precipitation.

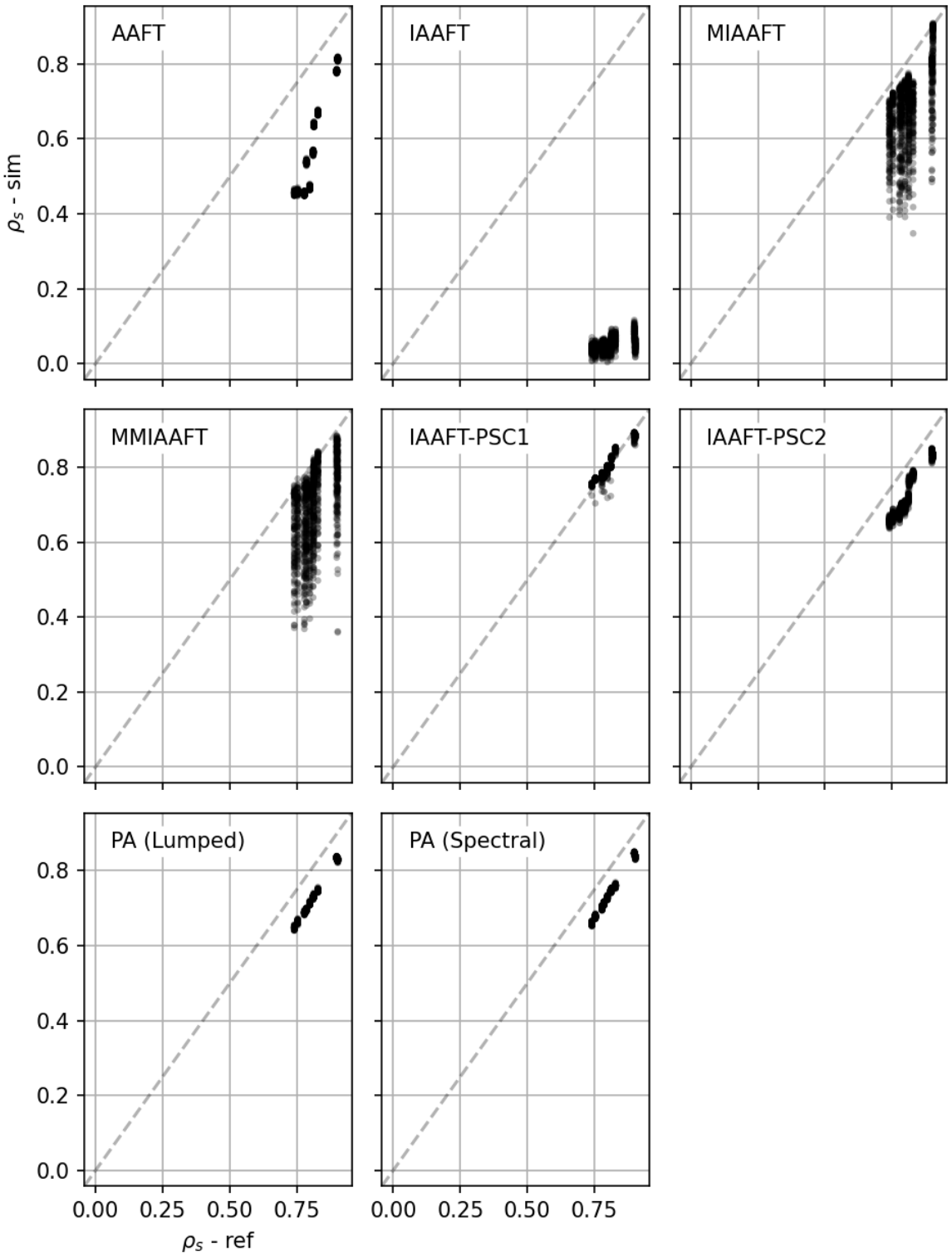


Figure 9.24: Pairwise Spearman cross-correlations using various methods for daily precipitation.

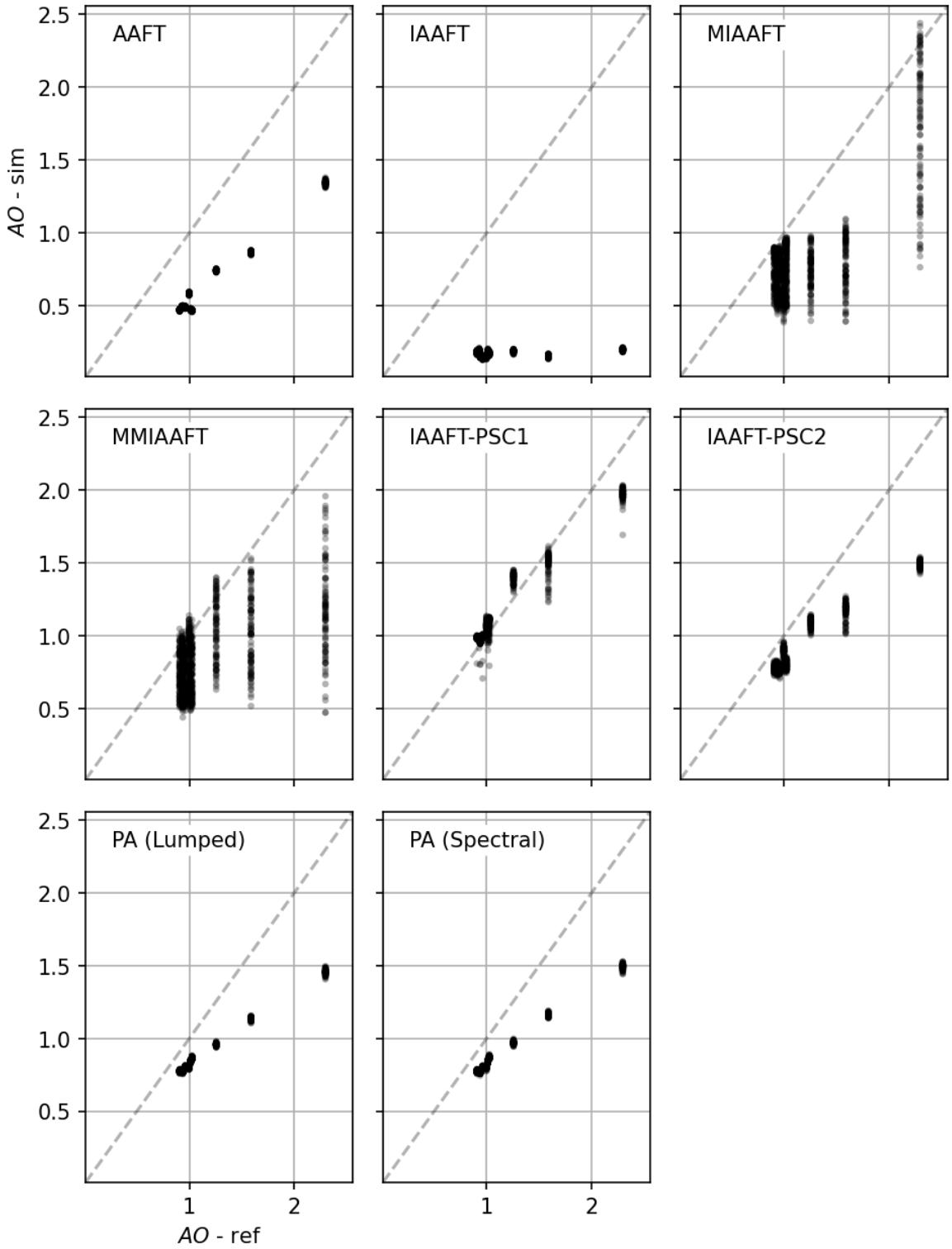


Figure 9.25: Pairwise cross-order asymmetries using various methods for daily precipitation.

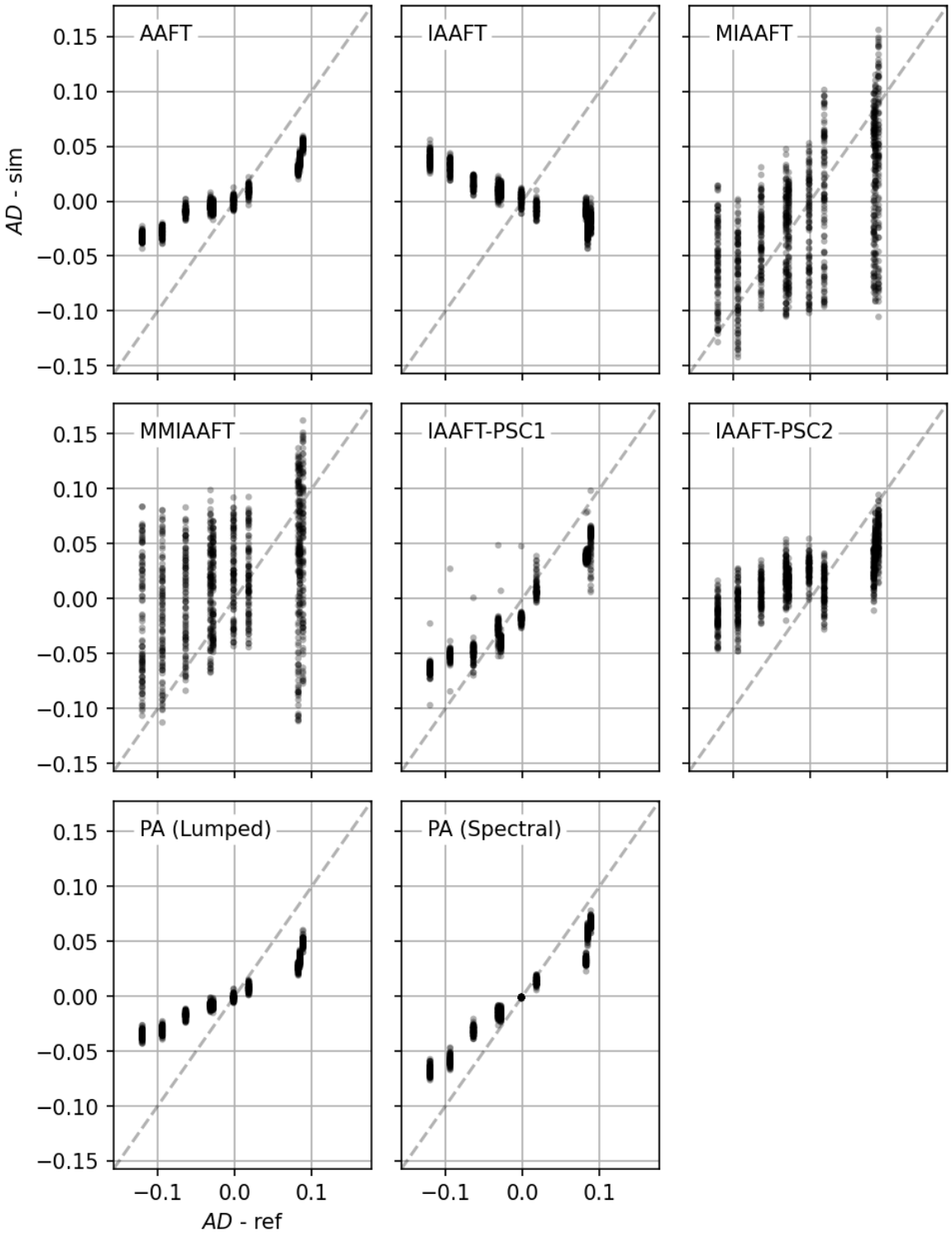


Figure 9.26: Pairwise cross-directional asymmetries using various methods for daily precipitation.

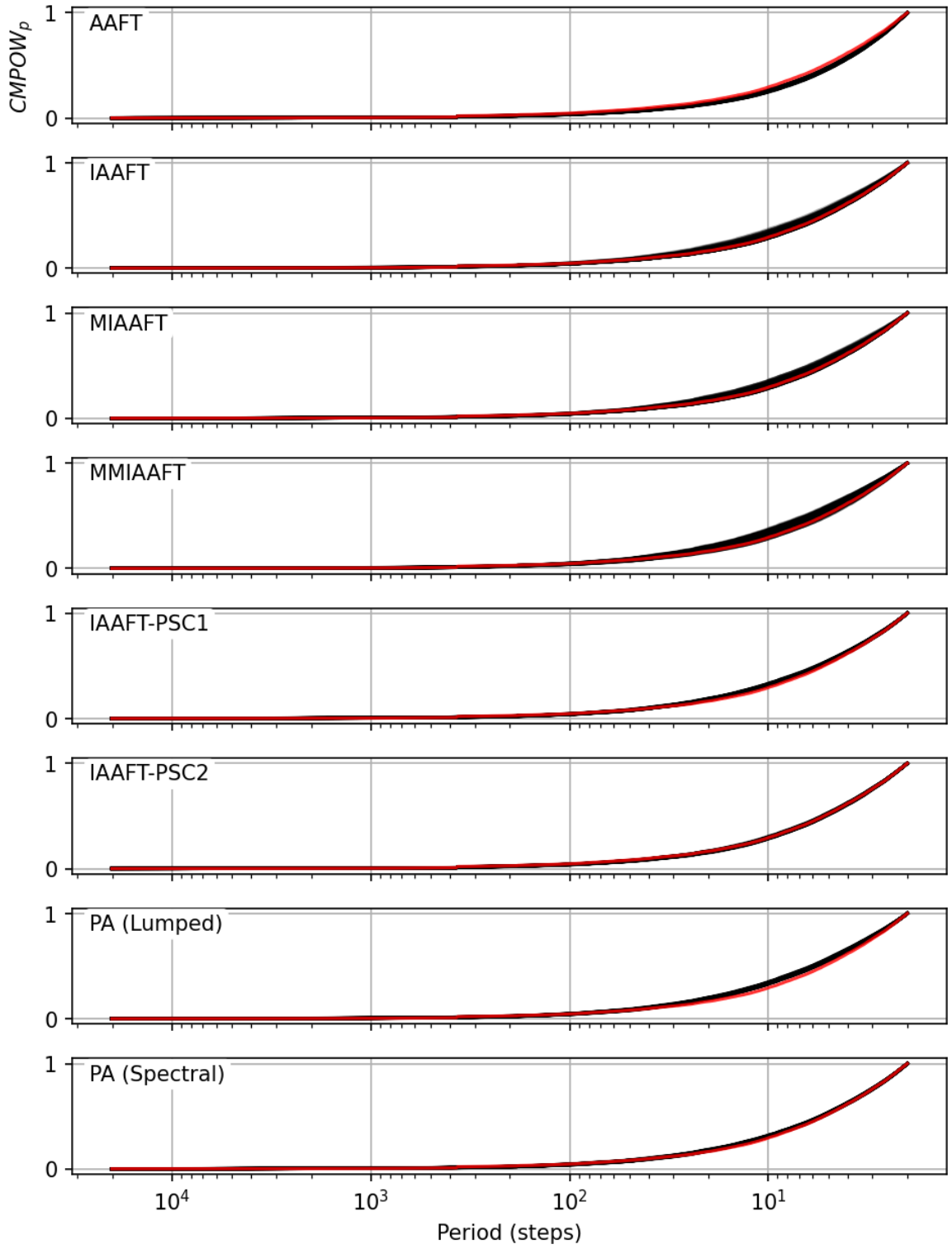


Figure 9.27: $CMPOW_p$ for various methods for daily precipitation. Red are observed while black are the simulations.

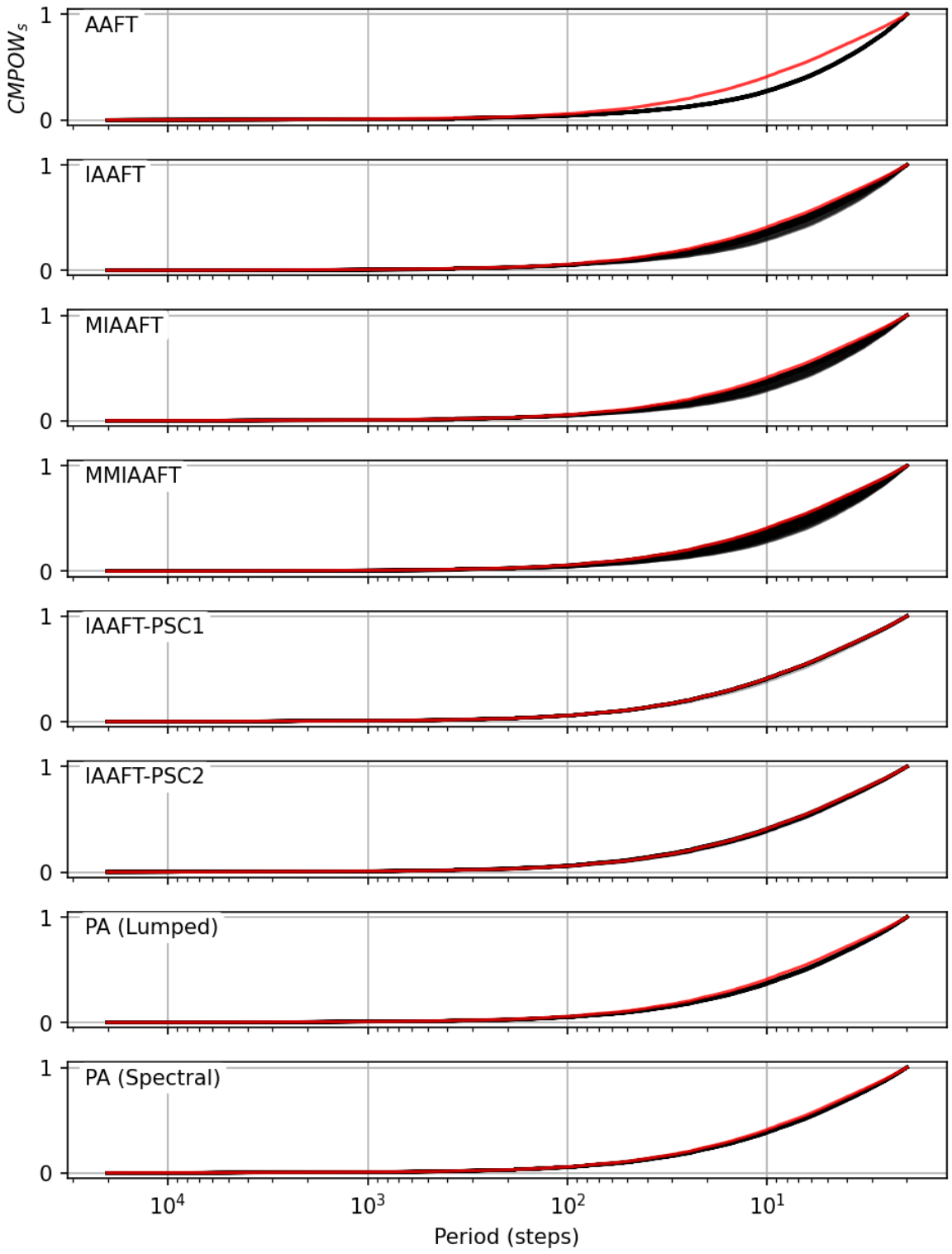


Figure 9.28: $CMPOW_s$ for various methods for daily precipitation. Red are observed while black are the simulations.

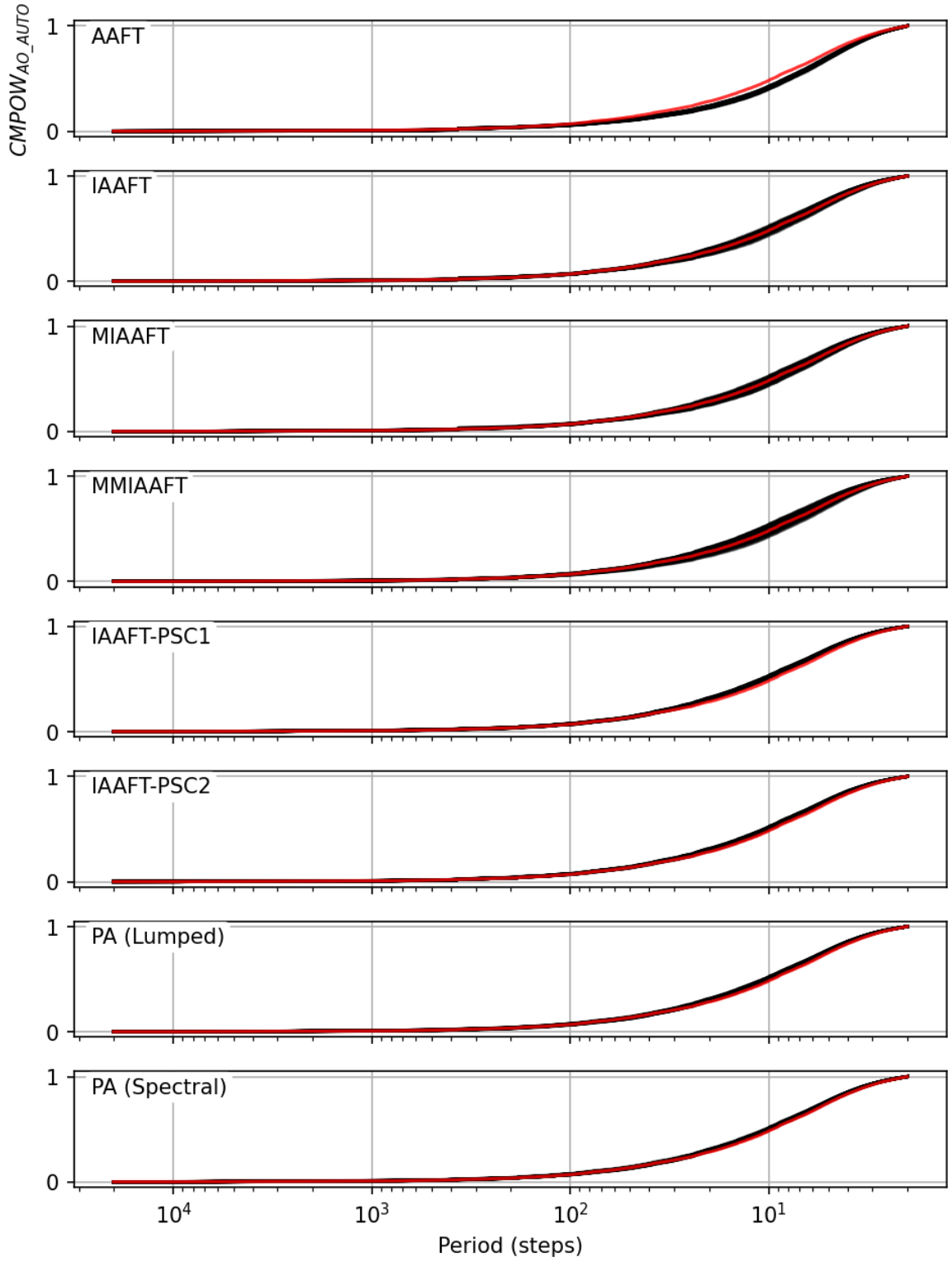


Figure 9.29: $CMPOW_{AO\_AUTO}$ with a lead step of 1 for various methods for daily precipitation. Red are observed while black are the simulations.

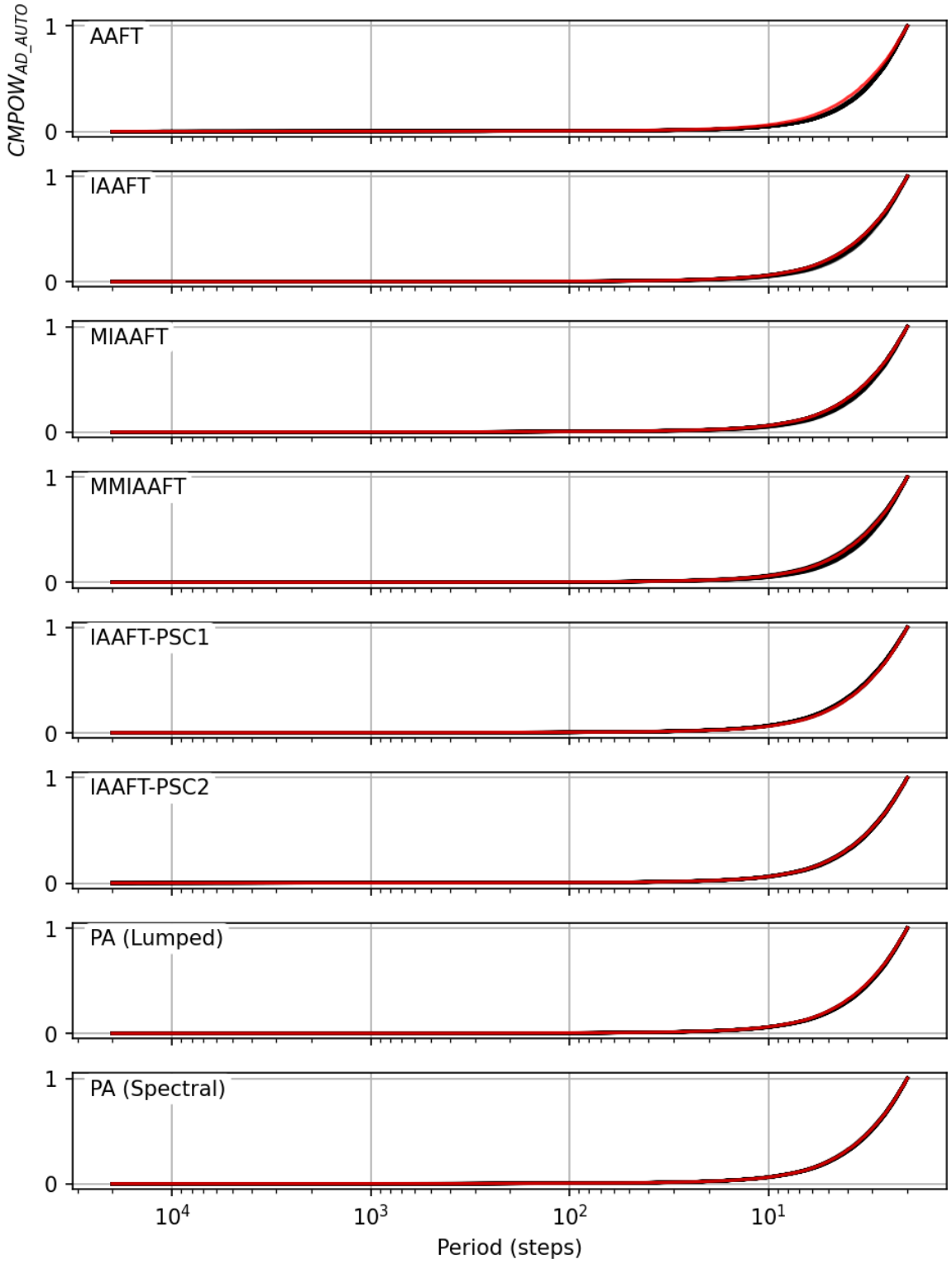


Figure 9.30: $CMPOW_{AD\_AUTO}$ with a lead step of 1 for various methods for daily precipitation. Red are observed while black are the simulations.

than one year time are considered here. Similar to discharge, tor the very high frequency range (greater than 10 days), all methods show good synchronization for $CMCORR_p^d$. As expected, IAAFT diverges away from the reference in the upper range by staying flat. On the other hand, comparing to discharge, there are cases where both $CMCORR_p^d$ and $CMCORR_s^d$ show the opposite behavior. For example, except for PA (Spectral), IAAFT-PSC1 and IAAFT-PSC2, a considerable offset between the two curves for each method is present. This points to the problem of the older simulation methods simulating only one type of correlation correctly but not the other i.e., adding an offset between the two. The newer methods seem to circumvent this to a large degree. Finally, the contributions in the high frequency range are reproduced well by all methods except for IAAFT.

Regarding the distributions of the upper tails, similar to discharge, it can be observed (Fig. 9.33) that the daily spatial-sum distributions using all the methods are acceptable. There are biases though. MIAAFT, MMIAAFT and to some extent IAAFT-PSC2 show bias towards larger values. AAFT, IAAFT-PSC1 and PA (Spectral) show very good agreement with the reference. As IAAFT was run without conserving the cross-correlations, it shows that the sums are reduced by almost 60% on average. This also serves as a check. The distribution of the spatial-sum (Fig. 9.34) of the weekly series show similar results i.e., the biases towards higher values become much clearer. PA (Lumped) shows an additional problem of larger values. Fig. 9.35 shows the distribution of the weekly-sum series of a randomly selected station. All algorithms match the reference very well.

Fig. 9.36 shows the conditional probabilities of cases where at least three out of four stations produced a value of a return period of two years or more versus the case where all of them produced the same. Similar to daily discharge, AAFT and IAAFT-PSC1 perform very well with IAAFT-PSC2, PA (Spectral) and (Lumped) not being far off. As expected, IAAFT has no synchronization of the high values.

9.6 Daily model inputs

The case of using simulated model inputs is the most curious of them all. Firstly, lumped HBV using spatially averaged values using observations for each time step were used to calibrate it using the Differential Evolution algorithm. The NSE values varied from 0.82 to 0.9 for the 10 year's case and from 0.8 to 0.88 for the 55 year's case for the four considered sub-catchments. Using the calibrated model parameters, the simulated inputs were now used to produce new simulated discharge series (called model discharge from here onward). The properties of these are compared to the properties of the series that were obtained using the calibrations (using observed inputs). Furthermore, the observed discharge curves are also shown along side for comparison (in *green*) in the figures. Unfortunately, PA could not be used as it resulted in simulations times that were too long. Following are the results in detail.

Fig. 9.37, 9.38, 9.39 and 9.40 show the auto-ρ_p, ρ_s, AO and AD functions for a randomly selected gauge. All the measures show very good agreement with the reference and the observed except for AD. It can be seen that the properties of the simulation and the calibrated

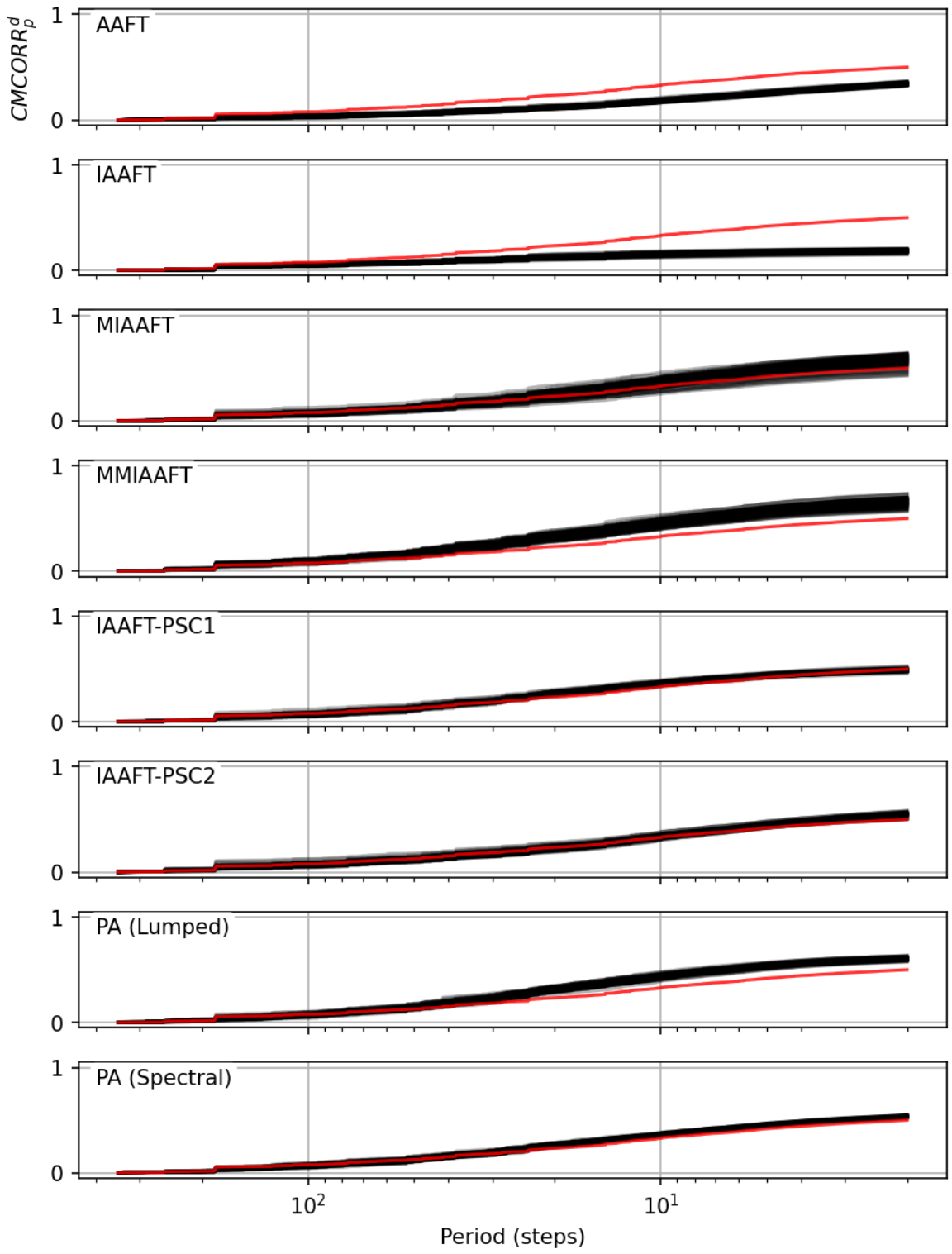


Figure 9.31: $CMCORR_p^d$ using various methods for daily precipitation. Red are observed while black are the simulations.

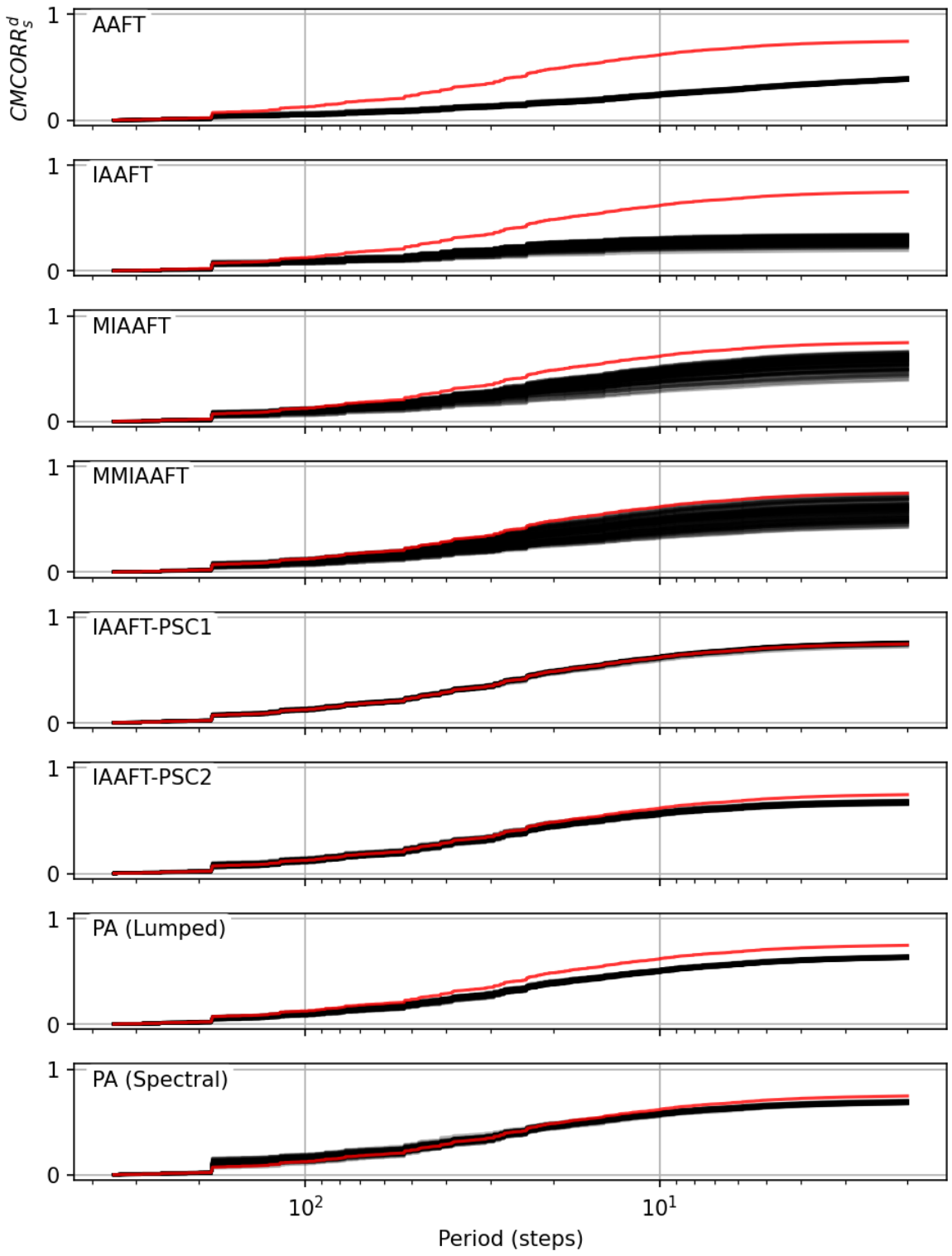


Figure 9.32: $CMCORR_s^d$ using various methods for daily precipitation. Red are observed while black are the simulations.

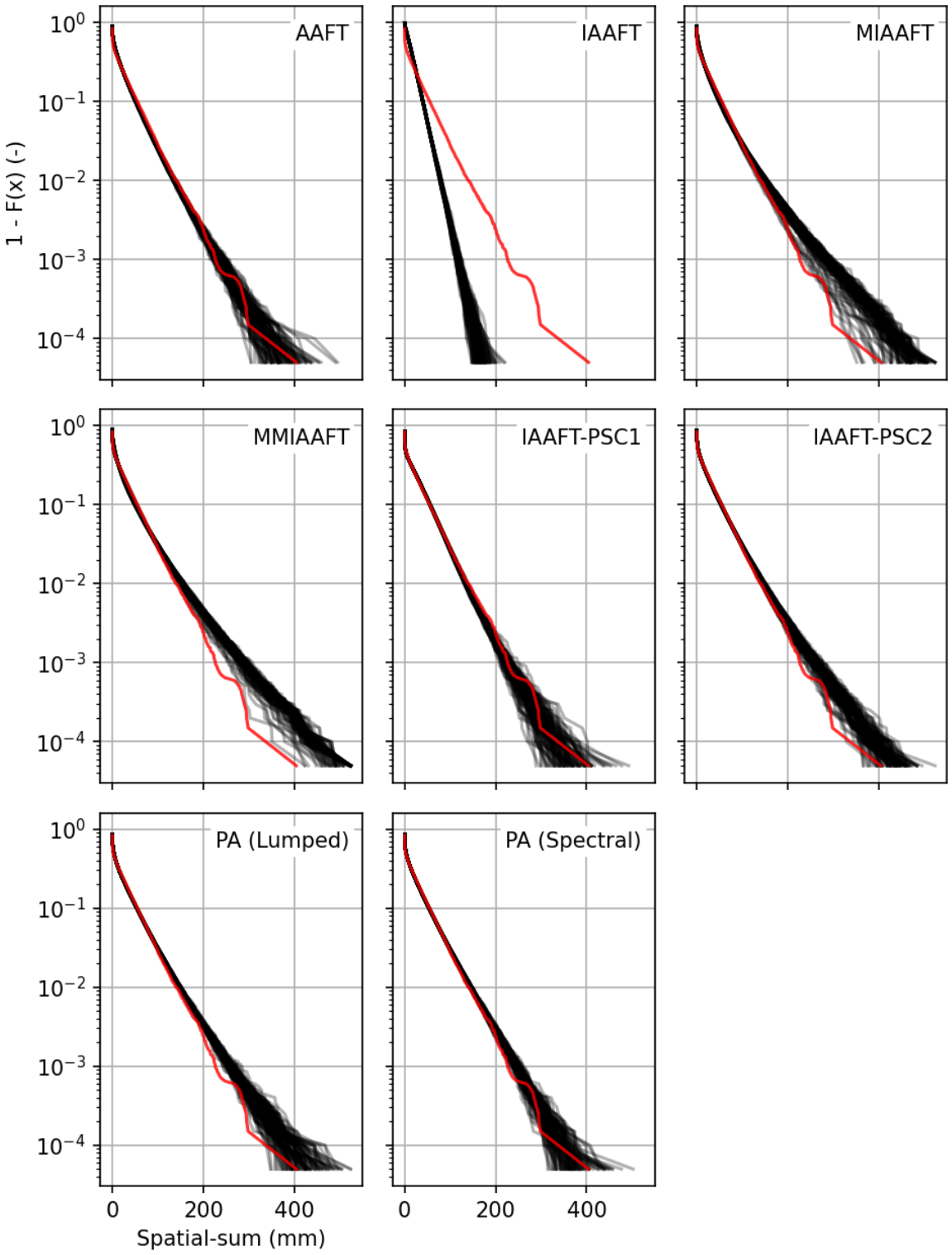


Figure 9.33: Distribution of daily spatial-sums of precipitation for five stations in the study area. Red are observed while black are the simulations.

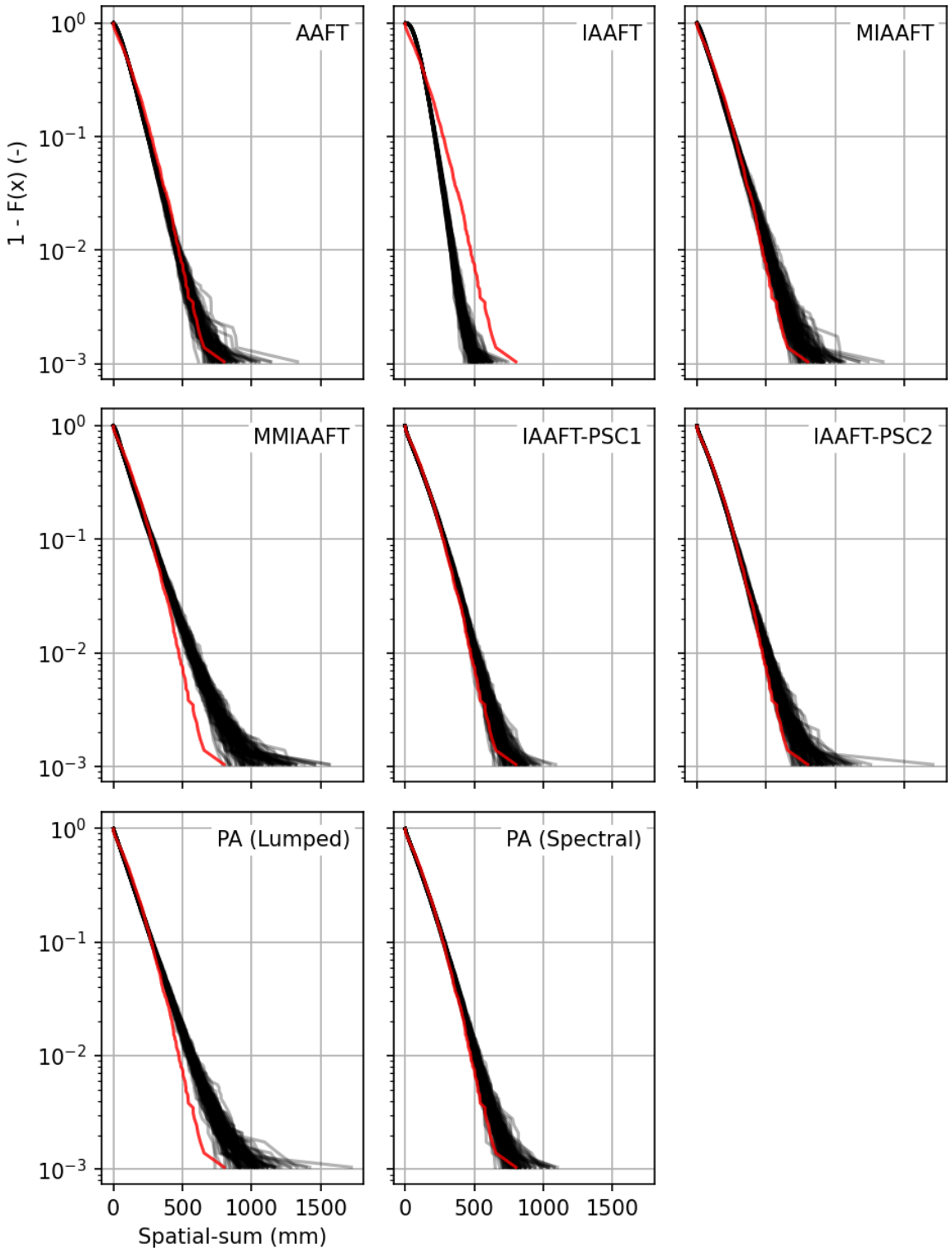


Figure 9.34: Distribution of weekly spatial-sums of precipitation for five stations in the study area. Red are observed while black are the simulations.

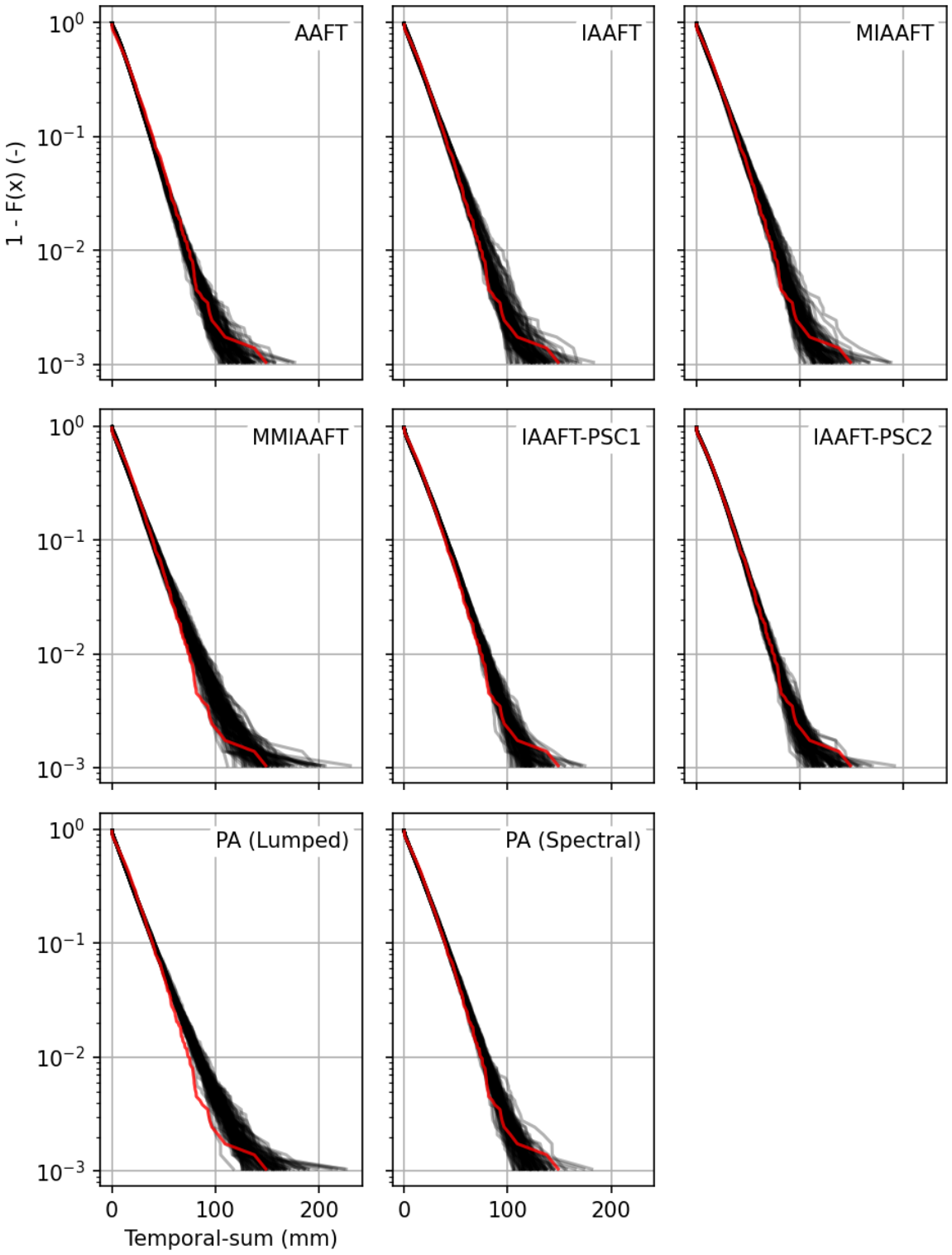


Figure 9.35: Distribution of weekly mean precipitation for randomly selected station in the study area. Red are observed while black are the simulations.

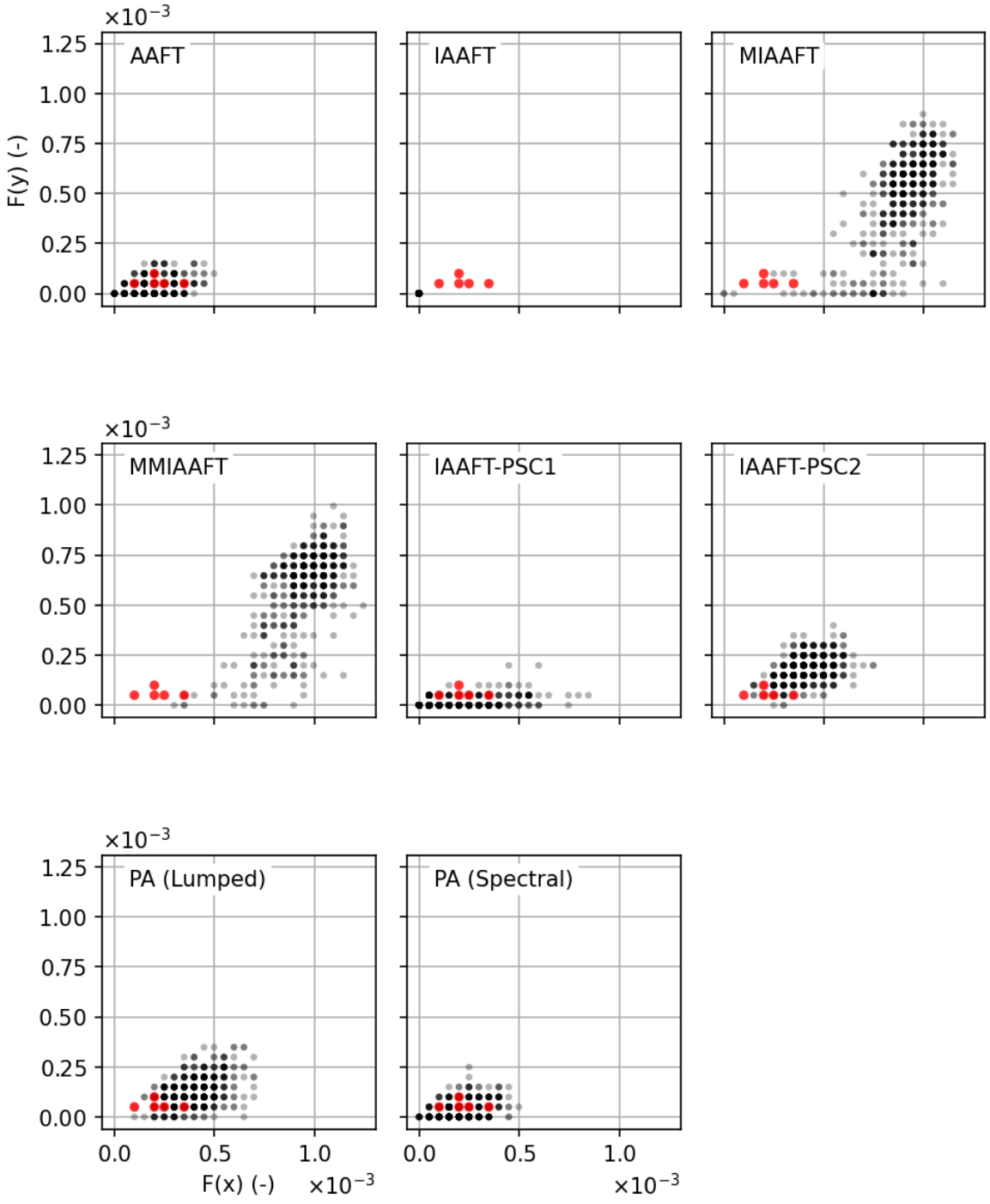


Figure 9.36: Conditional probability comparisons for daily precipitation of three out of four have values with return periods of two or greater (horizontal axis) versus all of them having values equal or greater than two years (vertical axis) at the same time step.

discharge are very similar but the observed values are very different when looking at lead steps within one and three week windows. The model discharges lose their asymmetric behavior rather quickly afterwards but then plunge towards the observed values again. This has nothing to do with the simulated inputs but rather the model itself. The model seems to be lacking in the reproduction of the rising and recession limbs similar to those of the observed consistently. This indicates that the model needs improvement in some of it aspects. This behavior could also be catchment specific.

All the pairwise cross-properties all perform acceptably (Fig. 9.41, 9.42, 9.43 and 9.44). It is strange to see that the ρ_p values tend to be underestimated but not overestimated as often. This is most probably related to the simulated inputs. The methods that overestimated the cross-ρ_p, also show some overestimation here but not as much as the precipitation. Such a problem can also be linked to the fact that cross-asymmetries (pairwise) were also not conserved so well for precipitation. Furthermore, observed data have some error in them but the model discharge are more deterministic. This is an area where more investigation can be carried out in the future.

Coming to the distributions of the upper tails, similar to discharge and precipitation, the daily spatial-sum distributions (Fig. 9.45) using all the methods show acceptable results. As IAAFT was run without conserving the cross-correlations, it shows that the sums are reduced by almost 40% on average. The distribution of the spatial-sum of the weekly-mean series show similar results (Fig. 9.46). Fig. 9.47 shows the distribution of the weekly-mean series of a randomly selected station. Again, no major biases but a tendency to underestimate at the higher aggregations can be observed. This is inline with reduced cross-properties of the simulations, that manifest their effects when considering many values rather than individual ones. It is noteworthy that the simulated discharges using simulated inputs are larger than the observed values, even though the discharge from the original calibration was consistently lower in the upper tails as compared to observed values and the distributions of inputs are exactly the same everywhere.

Fig. 9.48 shows the conditional probabilities of cases where at least three out of four stations produced a value of a return period of two years or more versus the case where all of them produced the same. Here, all algorithms perform very similar and acceptable but it can be observed that the observed probabilities are away from the simulations and the reference.

9.7 Hourly precipitation

So far, all the presented algorithms performed very similar for the distributions of extremes of the daily cases. This lead to the question: *Would these perform similarly for the hourly case as well?*. Unfortunately, PA could not be used as it took too long to converge. Furthermore, only the more important properties are presented here.

For the various methods Fig. 9.49 and 9.50 show the $CMCORR_p^d$ and $CMCORR_s^d$ spectra respectively. Note that only the contributions from the frequencies whose periods are less than one year time are considered here. The differences in spatial coherence can be seen here much more clearly. In the case of $CMCORR_p^d$, IAAFT shows no correlation in frequencies of

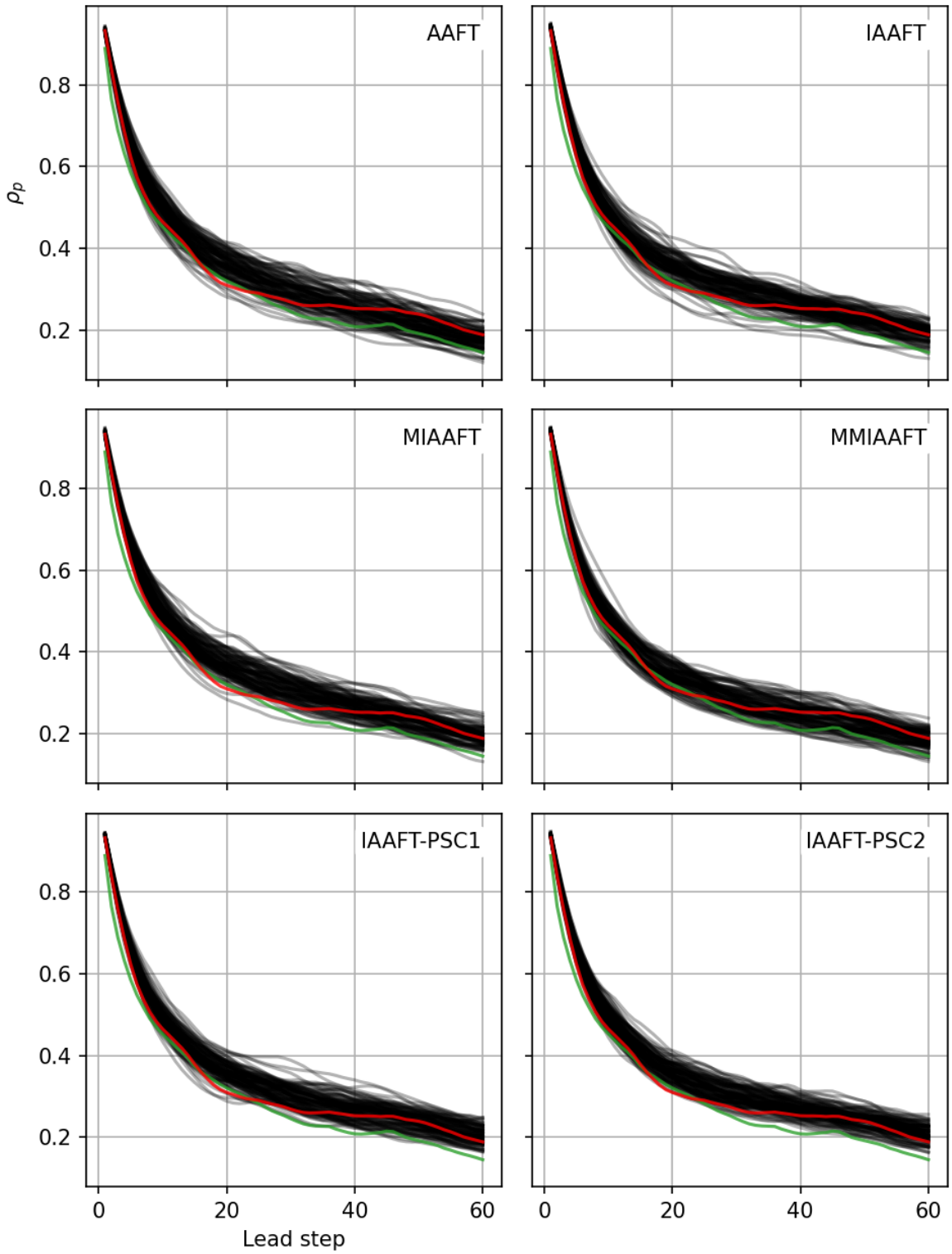


Figure 9.37: Pearson auto-correlation functions using various methods for daily model discharge. In red are those of the reference calibration, green are the observed while the black are using simulation.

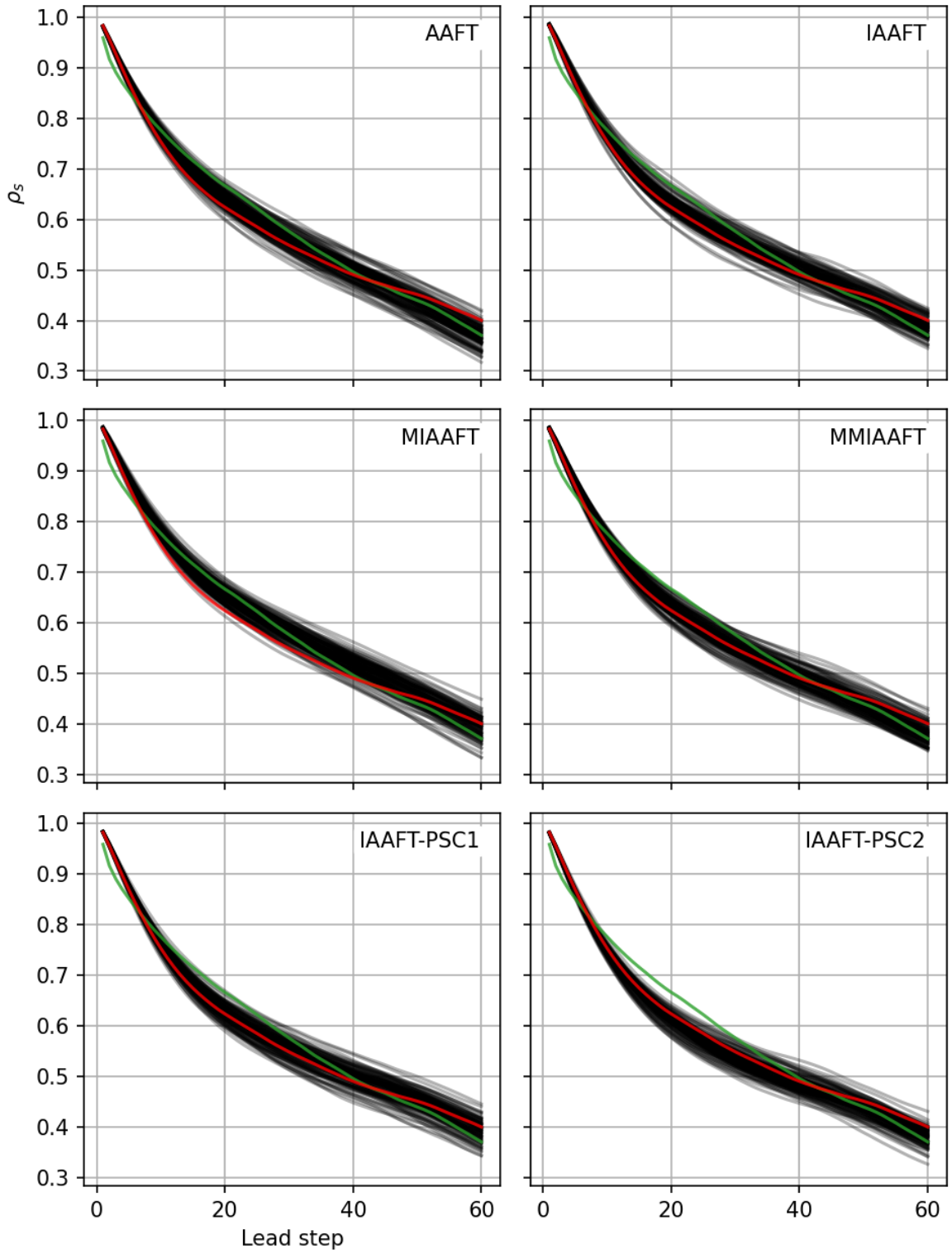


Figure 9.38: Spearman auto-correlation functions using various methods for daily model discharge. In red are those of the reference calibration, green are the observed while the black are using simulation.

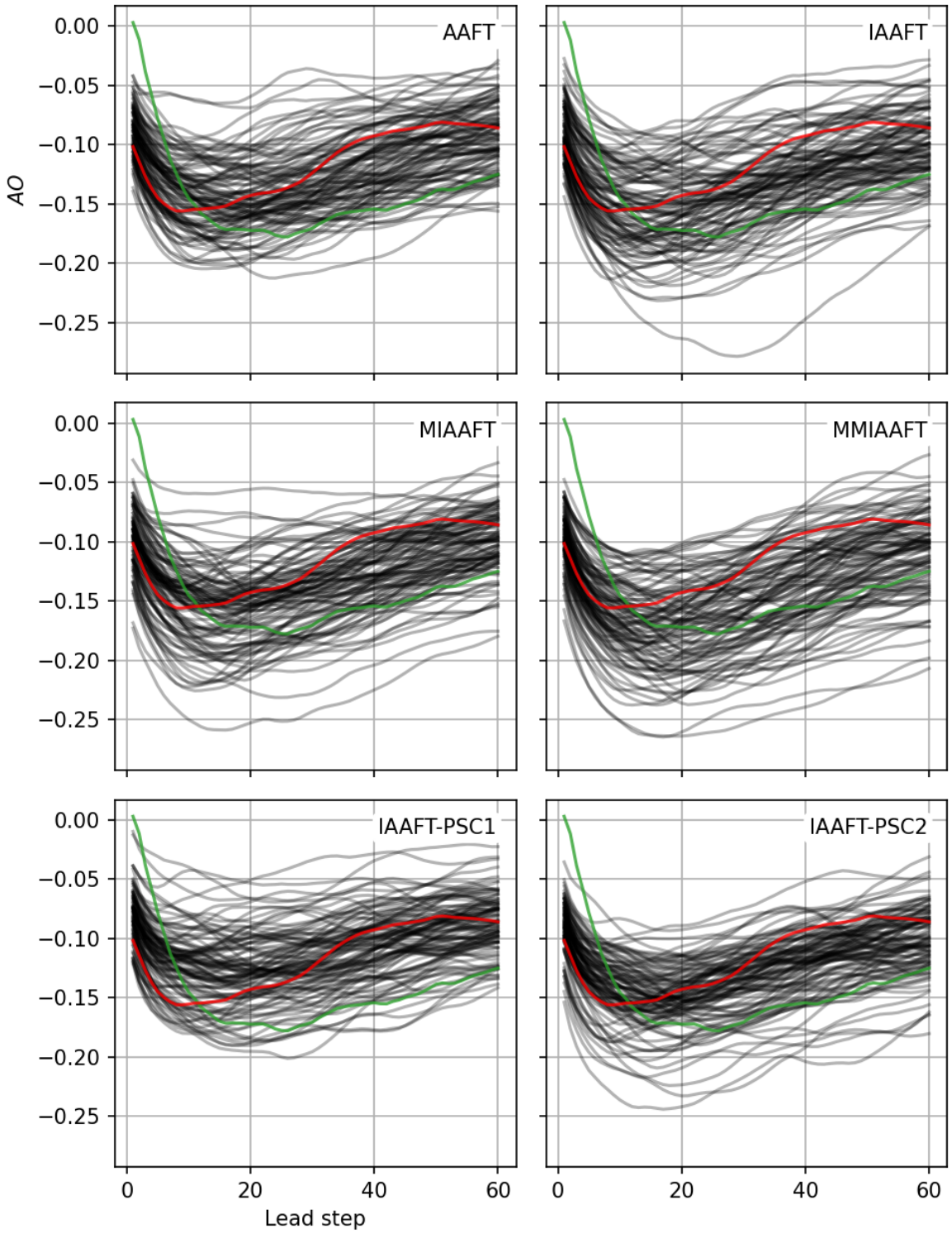


Figure 9.39: Auto-order asymmetry functions using various methods for daily model discharge. In red are those of the reference calibration, green are the observed while the black are using simulation.

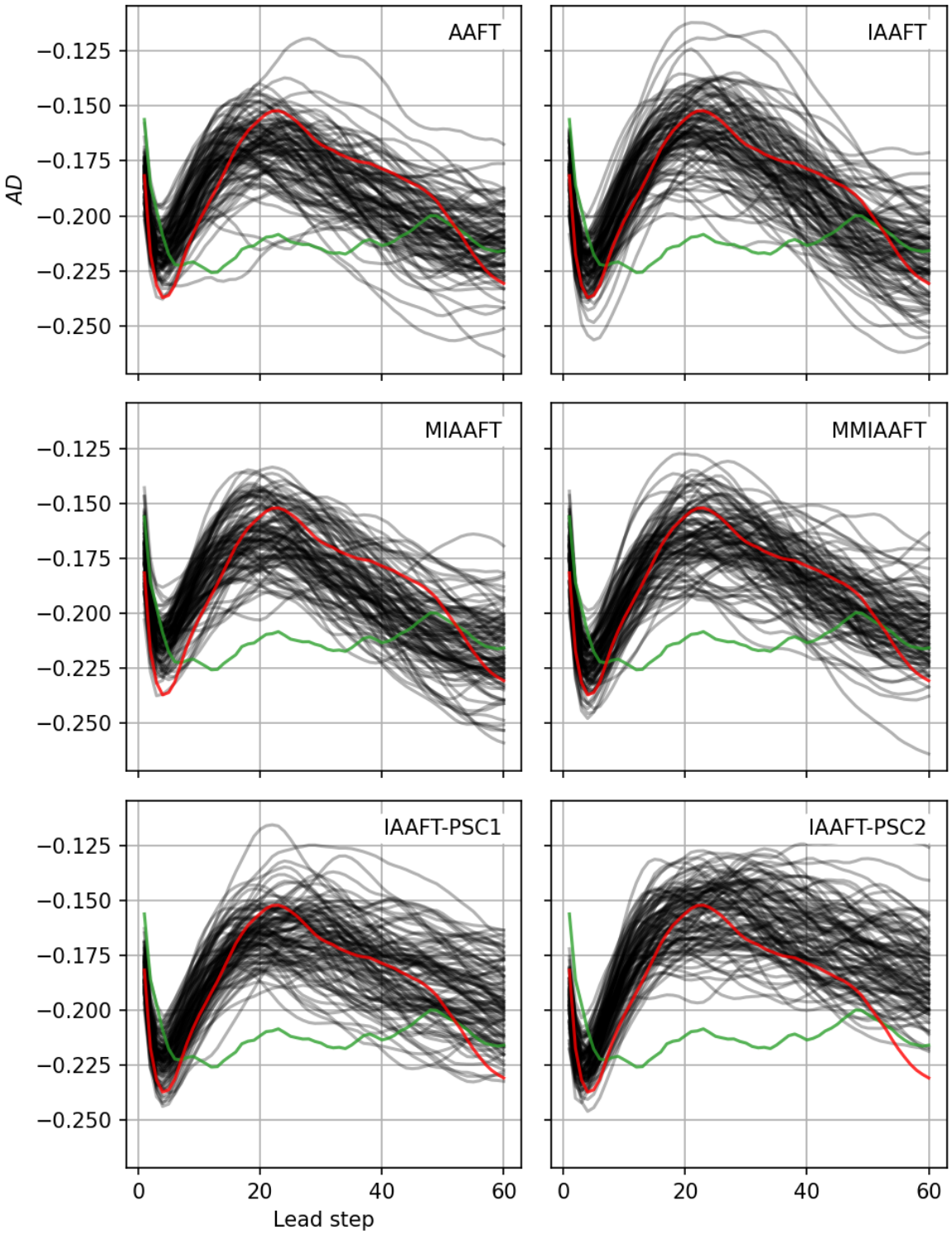


Figure 9.40: Auto-directional asymmetry functions using various methods for daily model discharge. In red are those of the reference calibration, green are the observed while the black are using simulation.

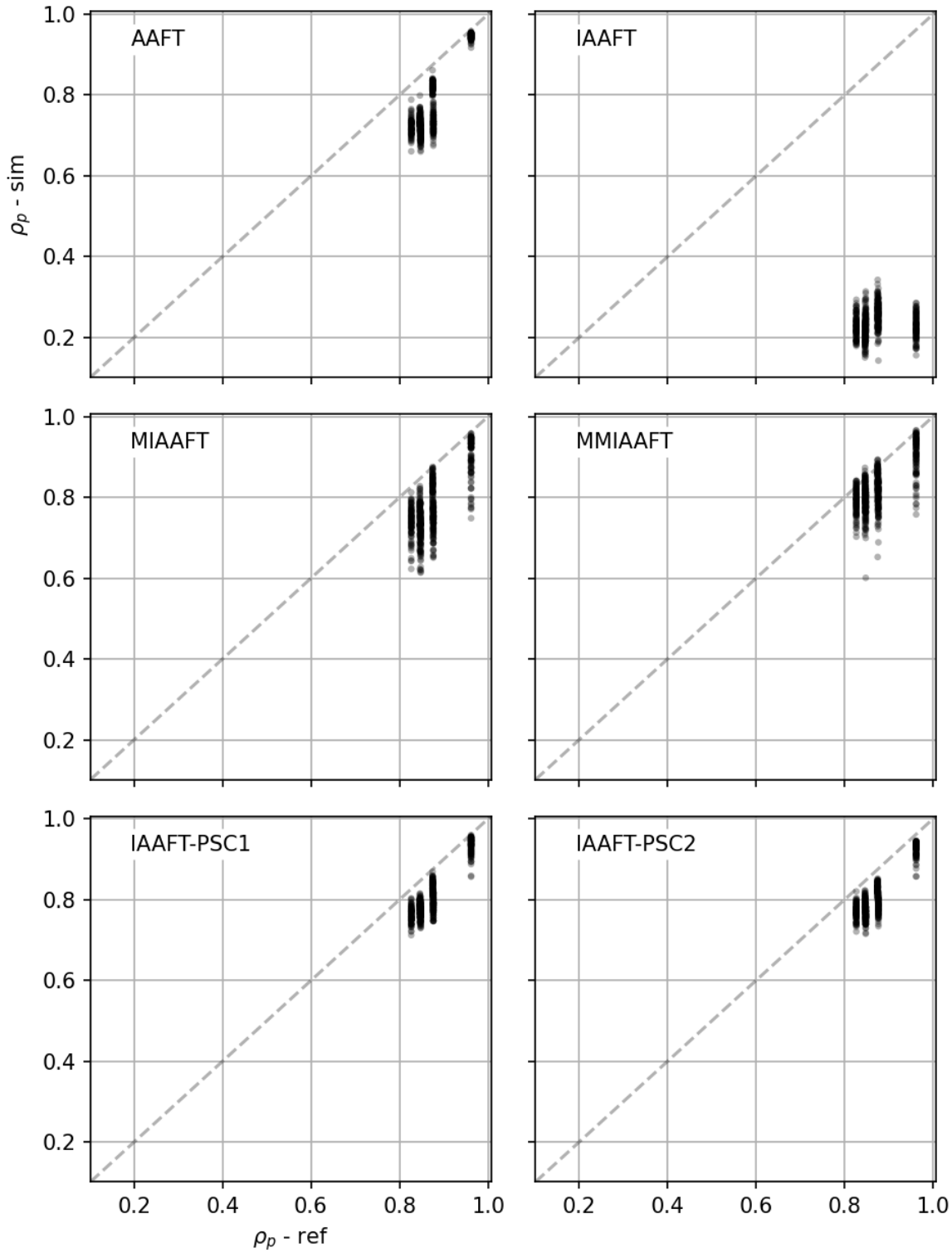


Figure 9.41: Pairwise Pearson cross-correlations using various methods for daily model discharge.

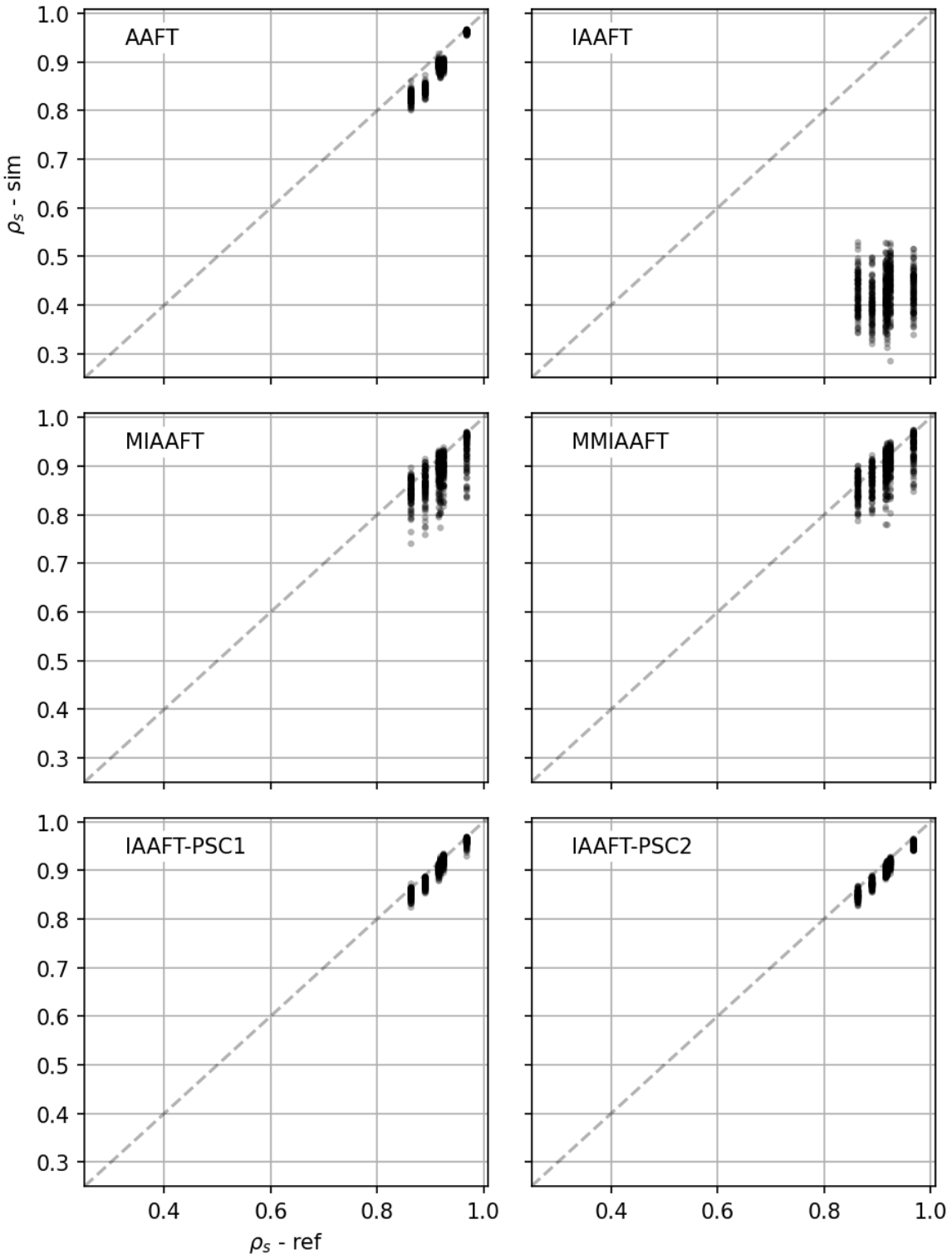


Figure 9.42: Pairwise Spearman cross-correlations using various methods for daily model discharge.

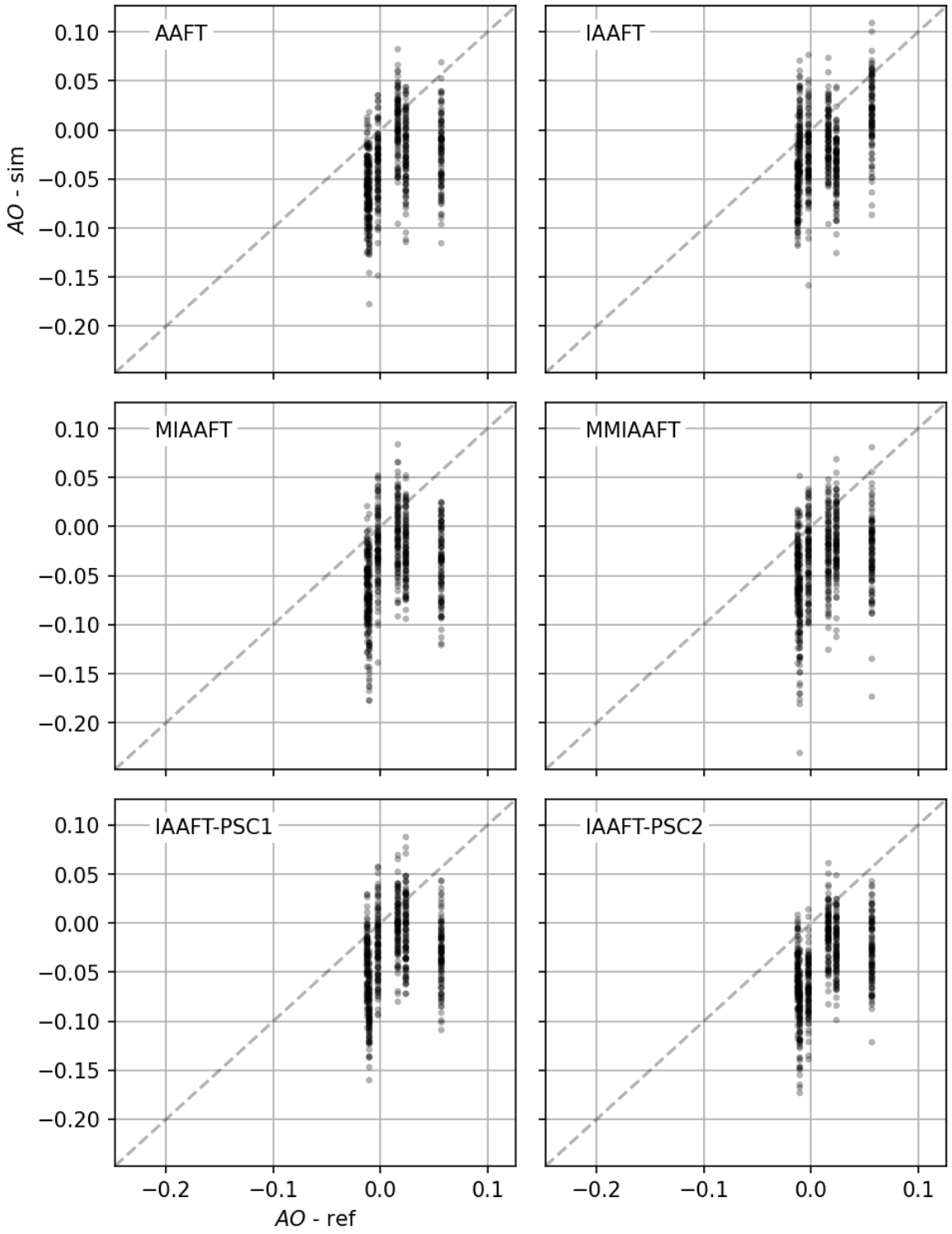


Figure 9.43: Pairwise cross-order asymmetries using various methods for daily model discharge.

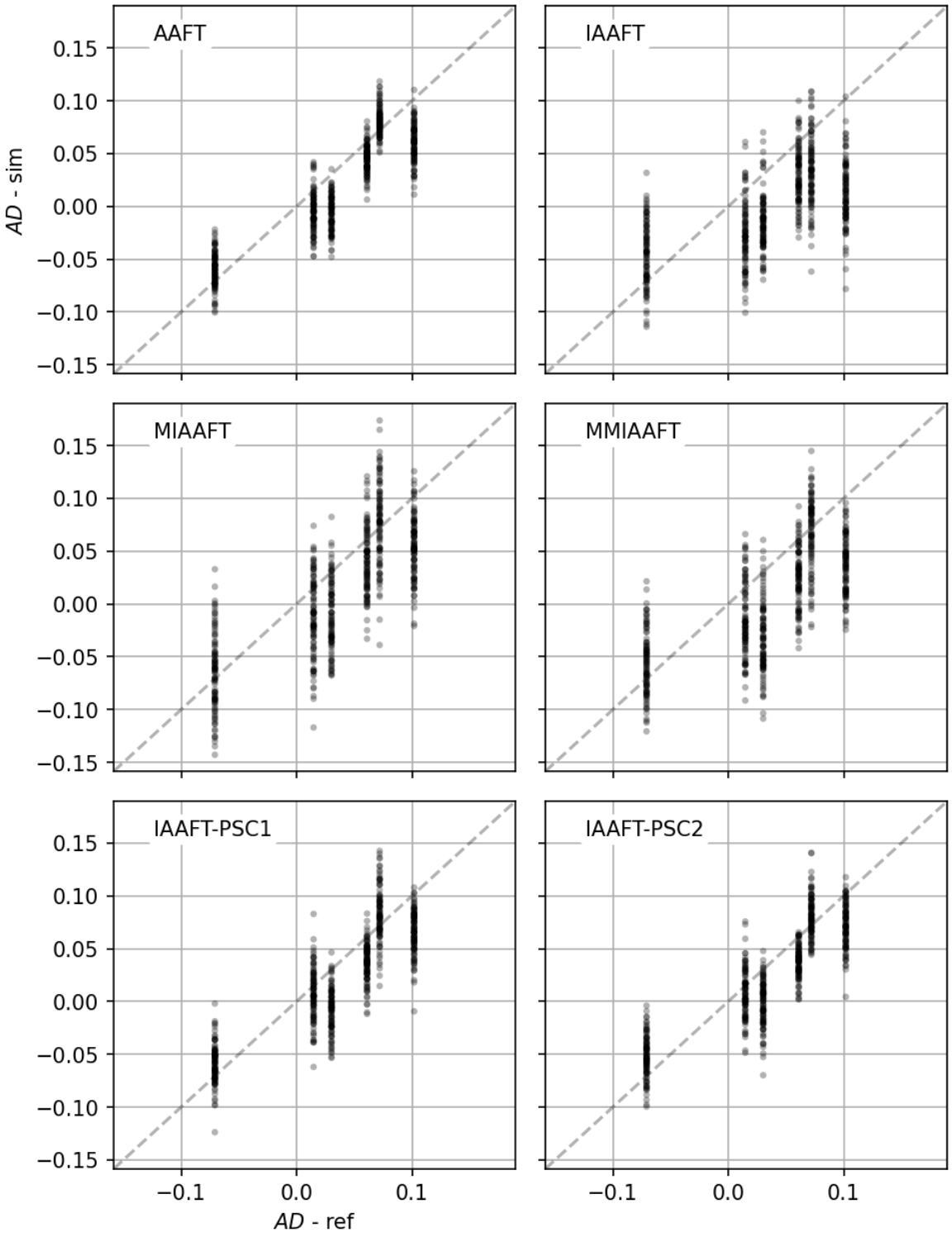


Figure 9.44: Pairwise cross-directional asymmetries using various methods for daily model discharge.

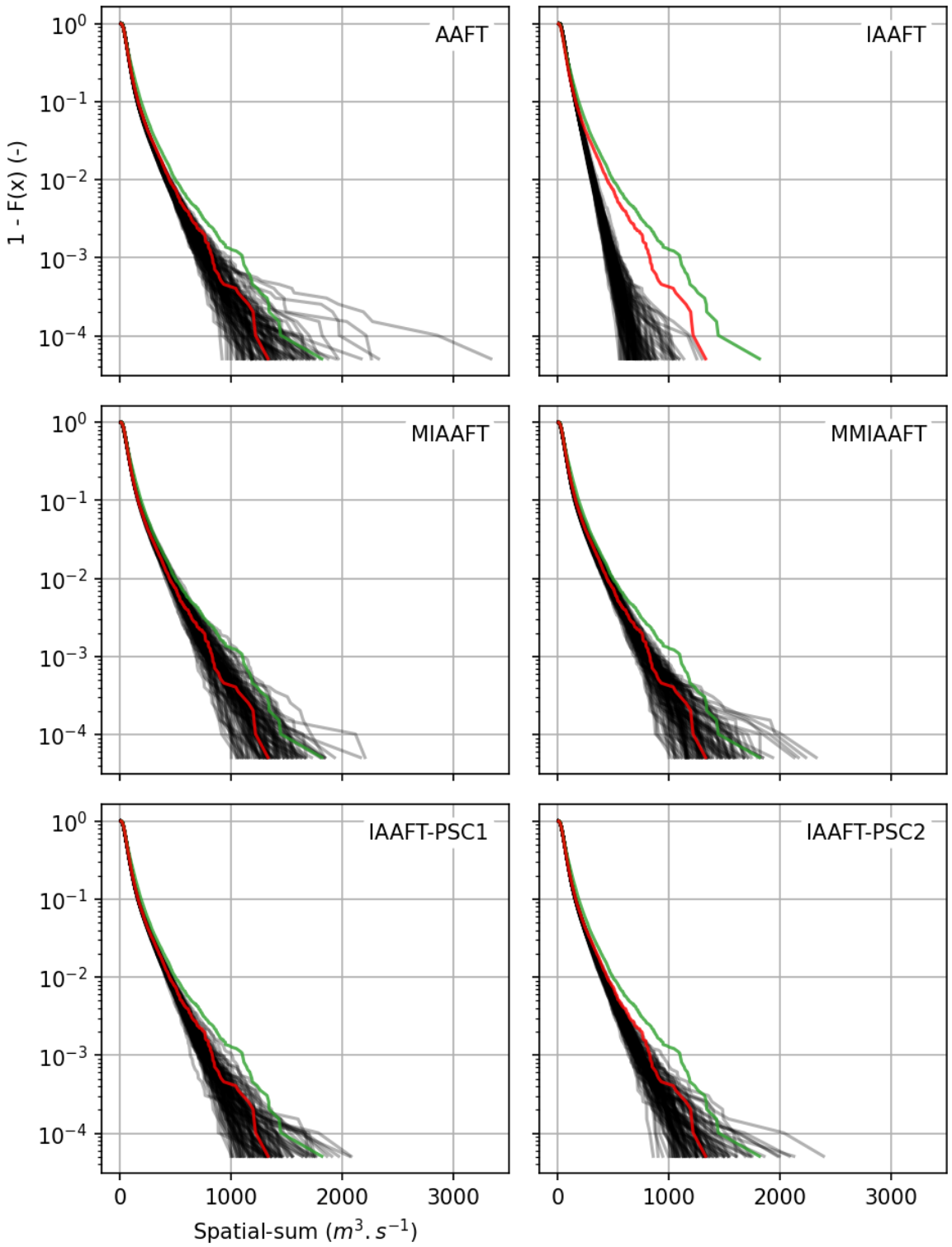


Figure 9.45: Distribution of daily spatial-sums of model discharge for five large head water catchments in the study area. Red are the reference calibration, green are the observed while black are the simulations.

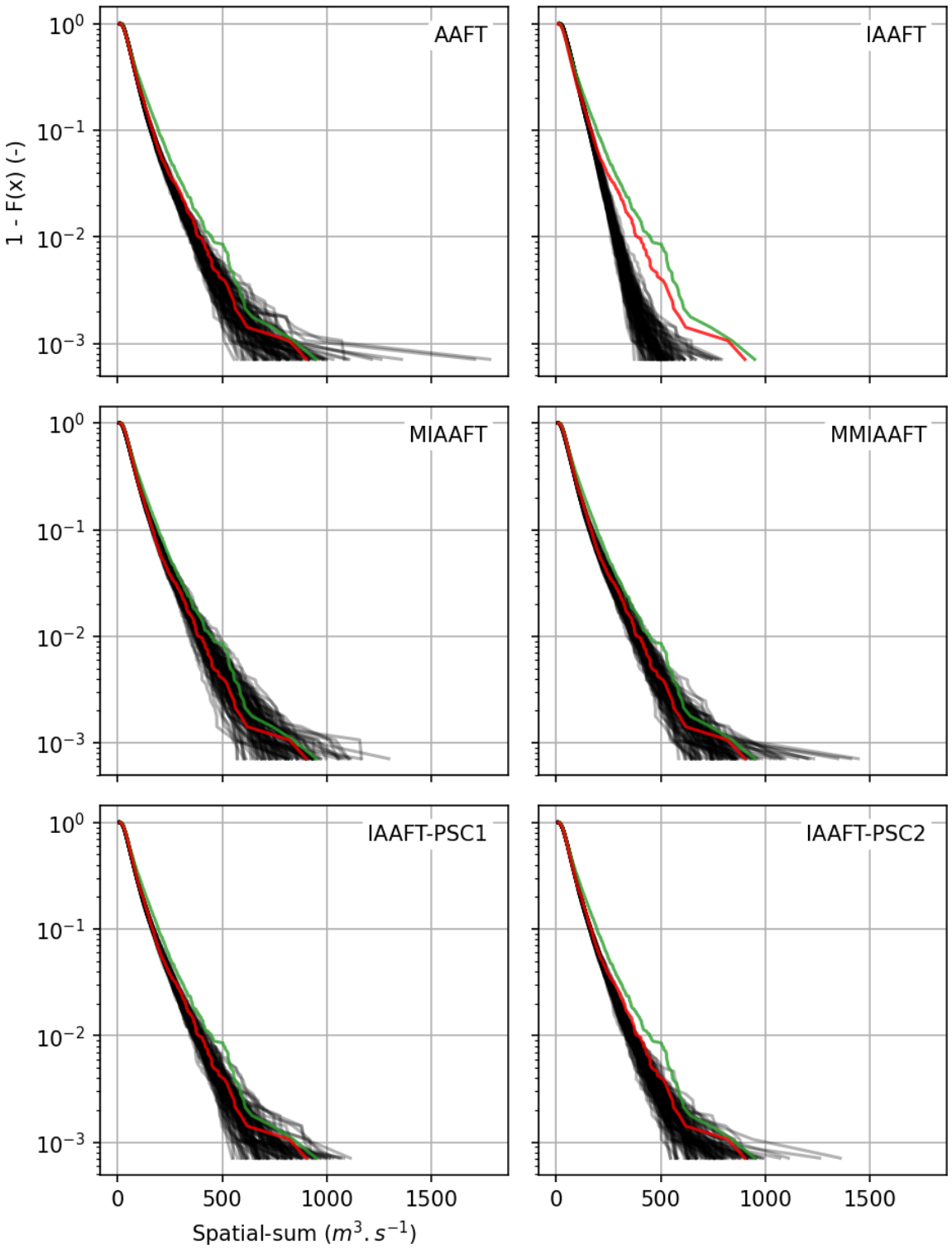


Figure 9.46: Distribution of weekly spatial-sums of model discharge for five large head water catchments in the study area. Red are the reference calibration, green are the observed while black are the simulations.

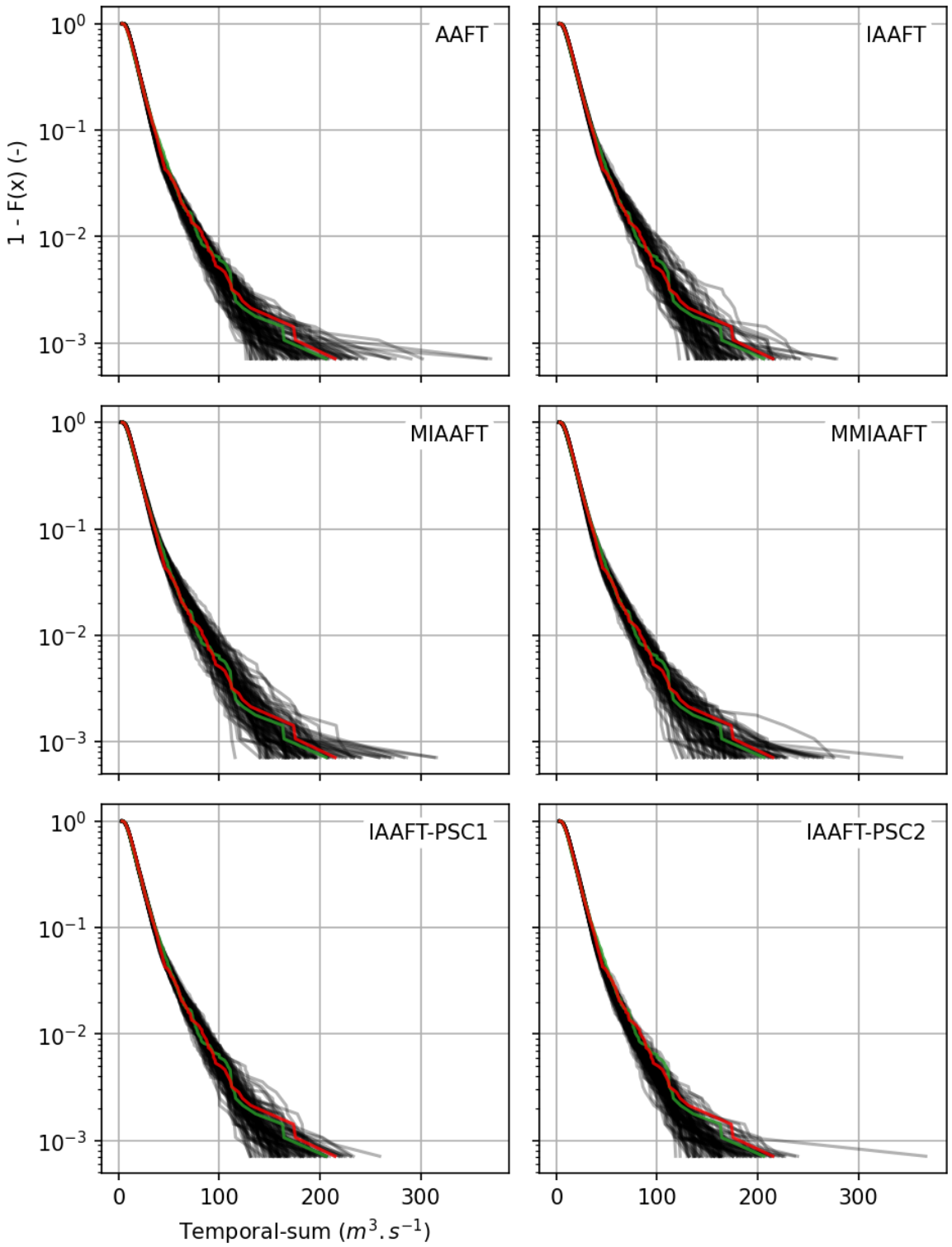


Figure 9.47: Distribution of weekly mean model discharge for randomly selected head water catchment in the study area. Red are the reference calibration, green are the observed while black are the simulations.

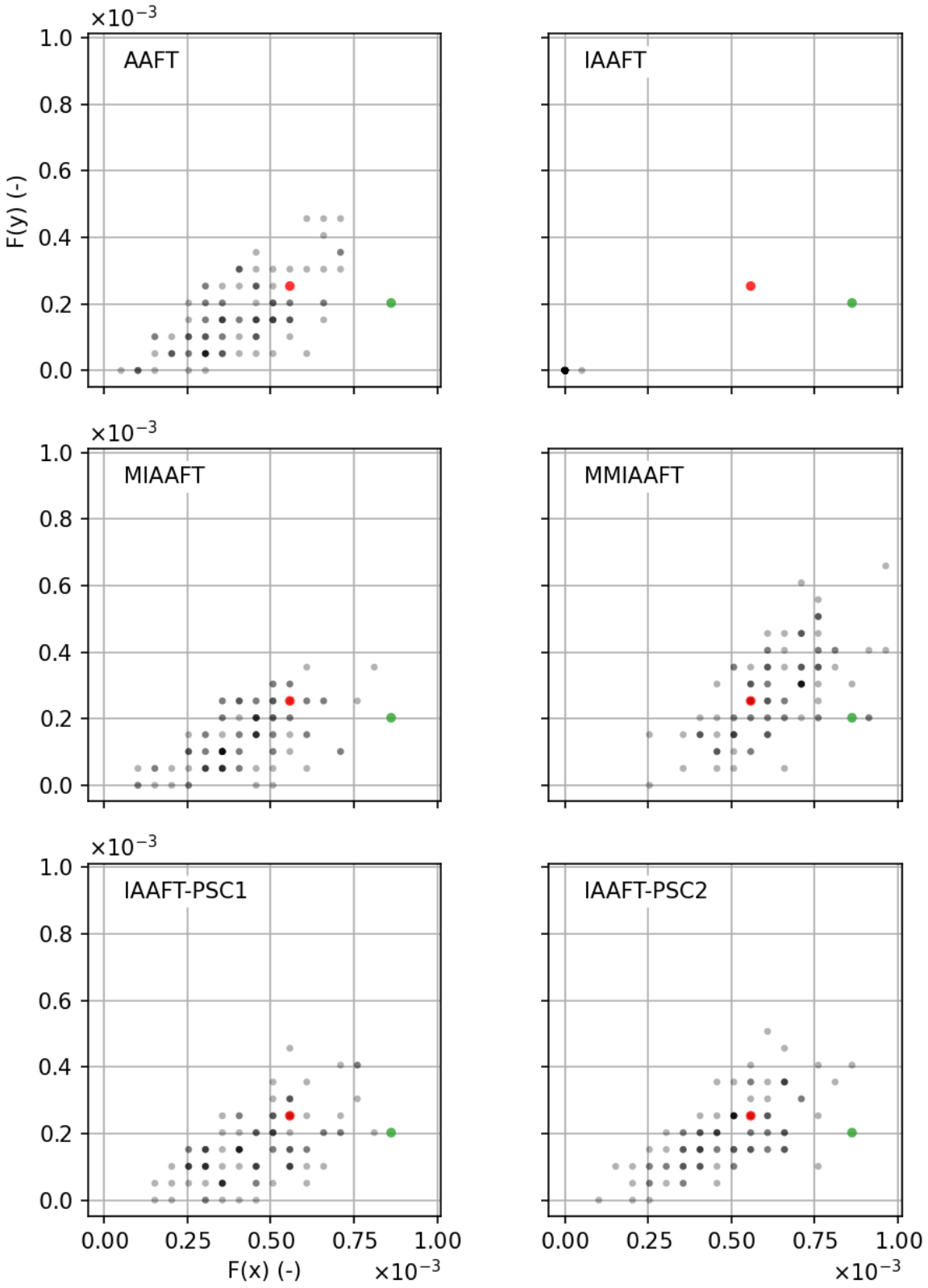


Figure 9.48: Conditional probability comparisons for daily model discharge of three out of four have values with return periods of two or greater (horizontal axis) versus all of them having values equal or greater than two years (vertical axis) at the same time step. Red are the reference calibration, green are the observed while black are the simulations.

about one year to one day. Then, it steadily rises up and slows down till the final frequency of about one hour. This results in it only achieving the very few largest spatial extremes correctly while underestimating the rest. AAFT shows a much reduced form of the same behavior. MIAAFT and MMIAAFT overestimate it, IAAFT-PSC1 both over-fits and over-estimates it while only IAAFT-PSC2 seems to follow the reference without any problems. The same is true for the case of $CMCORR_s^d$, where the offsets of the older methods become even larger. This points to the problem of under- and over-estimation of spatial extremes that many of these methods will exhibit next.

Fig. 9.51 shows the spatial-sums of the hourly series. Here, the results are very mixed results. AAFT shows only good results for the very high sums but has bias for the lower ones. MIAAFT overestimates by a very large margin (over 100%) while MMIAAFT seems to over-fit in some cases and matches the observed exactly while overestimating in the other cases. This is due to the high values of $CMCORR_p^d$. IAAFT-PSC1 over-fits as well and shows very little variability. IAAFT-PSC2 overestimates in general but not as much as the others. The over-fitting by some of the methods is most probably due to the very high frequency of zeros in the hourly data (more than 90%). Fig. 9.52 shows the spatial-sum of the resampled daily series. Here, biases can be observed as well but much less than those for the hourly. Fig. 9.53 shows the distribution of hourly precipitation resampled to daily for a randomly selected station. Strangely, IAAFT shows the best performance here.

Fig. 9.54 shows the conditional probabilities of cases where at least two out of four stations produced a value of a return period of six months or more versus the case where all of them produced the same. The chances of all four having high values is zero but in exceptional circumstances, two stations may have some synchronization for the observed case. Similar to the very high spatial-sums, MIAAFT and MMIAAFT seem to overestimate the probabilities much more than the rest. AAFT and IAAFT-PSC2 perform acceptably. But it should be kept in mind that these probabilities are very small and highly sensitive in general.

9.8 Summary and discussion of results

To summarize, the spatial-sum distributions using various settings of daily discharge series show that, even though many properties such as the copula asymmetries for the auto- and cross-cases are not kept exactly by the simulations, the behavior of extremes is well represented by enveloping the reference spatial-sum distributions. Daily precipitation shows similar results to that of the daily discharge, but the conservation of various properties is much better. Daily discharge using simulated model inputs show very good results for the distribution of extremes and other properties as well. However, it should be mentioned that as the spatial-sums of the discharges of the calibrated model were smaller than those of the observed discharges, the results of models can only be used when their discharge distributions match those of the observed in an acceptable manner. One the other hand, the results of the hourly precipitation show that only IAAFT-PSC2, performs better than the existing ones in a significant manner but still the results are not as good as compared to the daily simulations.

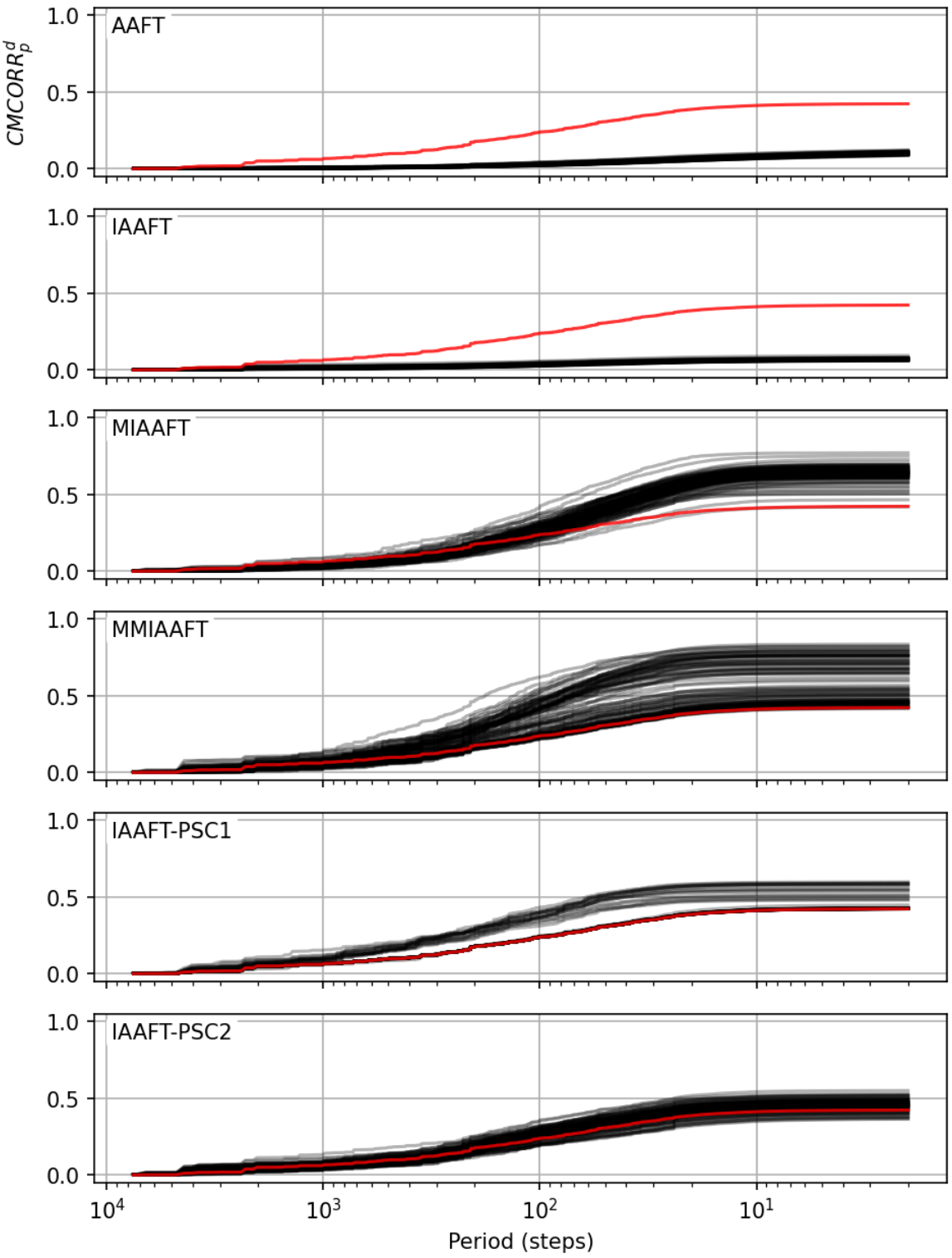


Figure 9.49: $CMCORR_p^d$ using various methods for hourly precipitation. Red are observed while black are the simulations.

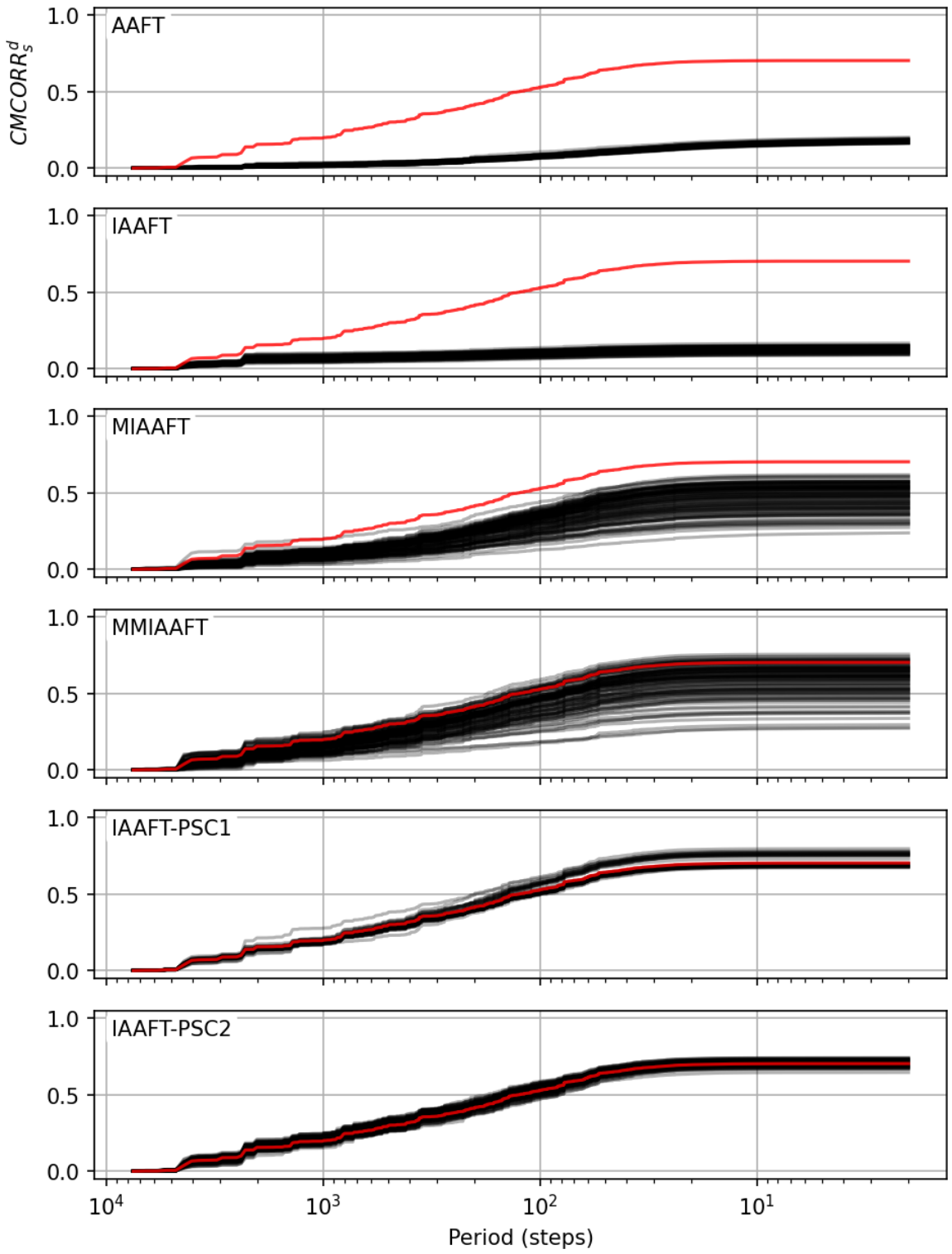


Figure 9.50: $CMCORR_s^d$ using various methods for hourly precipitation. Red are observed while black are the simulations.

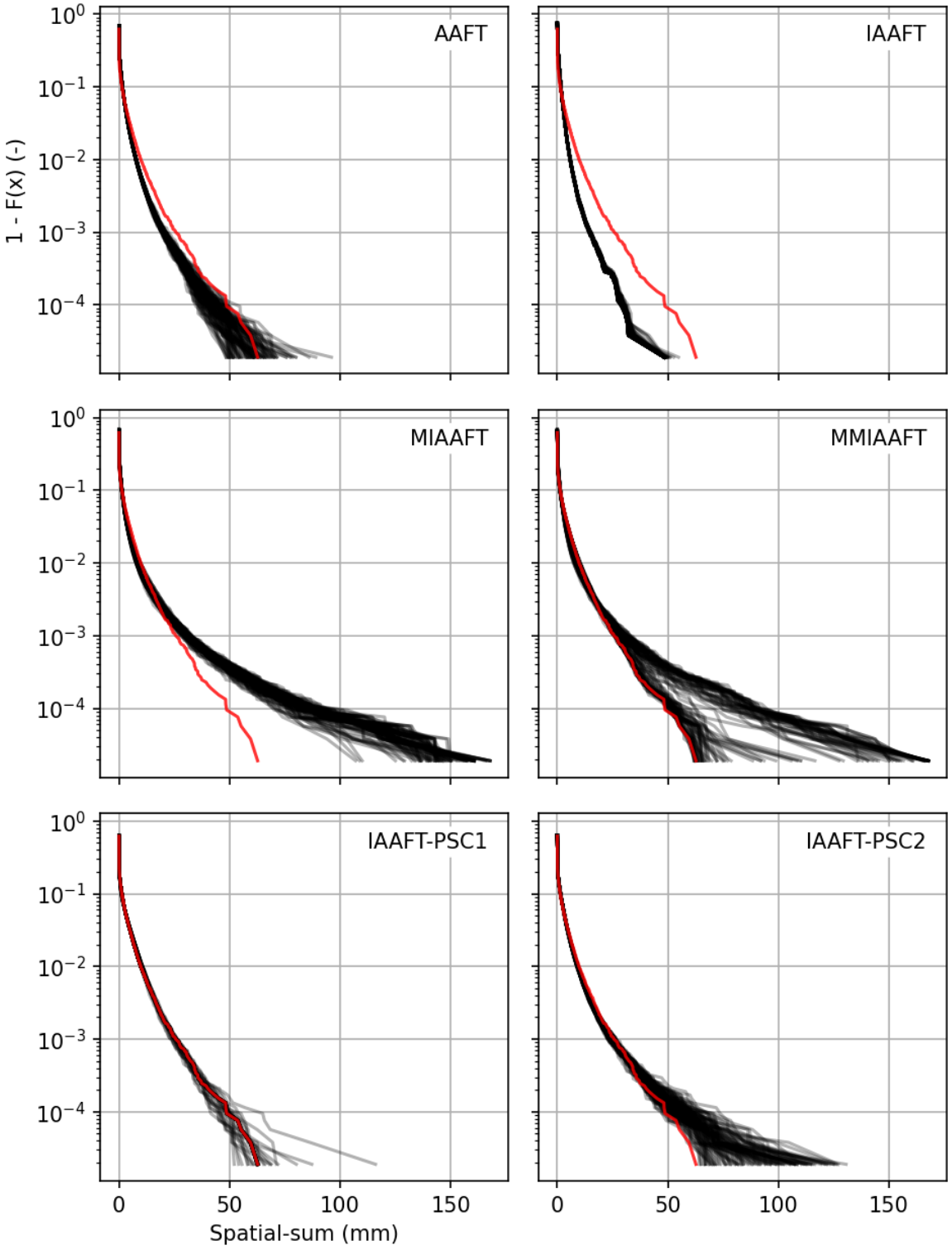


Figure 9.51: Distribution of spatial-sums of hourly precipitation for five stations in the study area. Red are observed while black are the simulations.

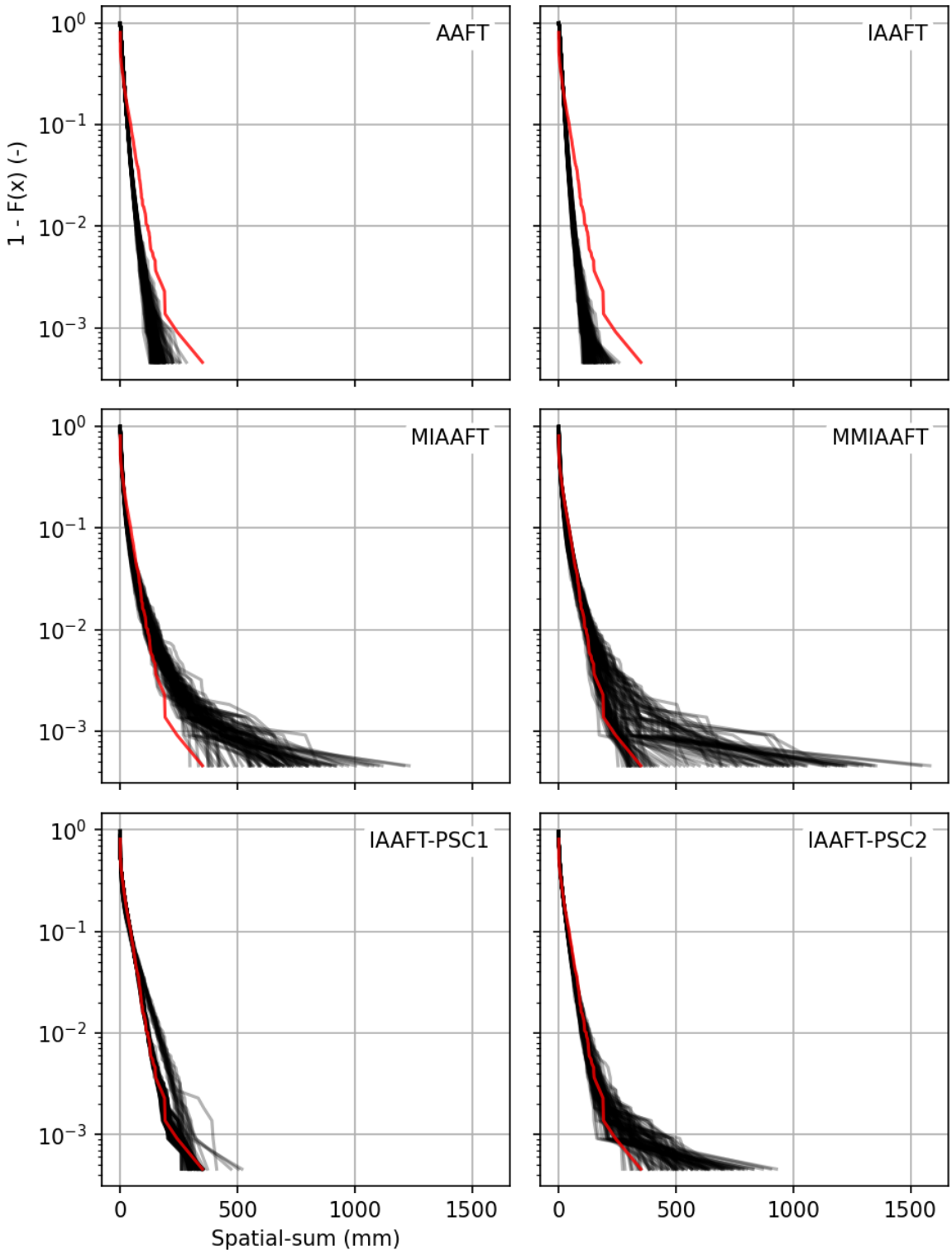


Figure 9.52: Distribution of resampled daily spatial-sums of hourly precipitation for five stations in the study area. Red are observed while black are the simulations.

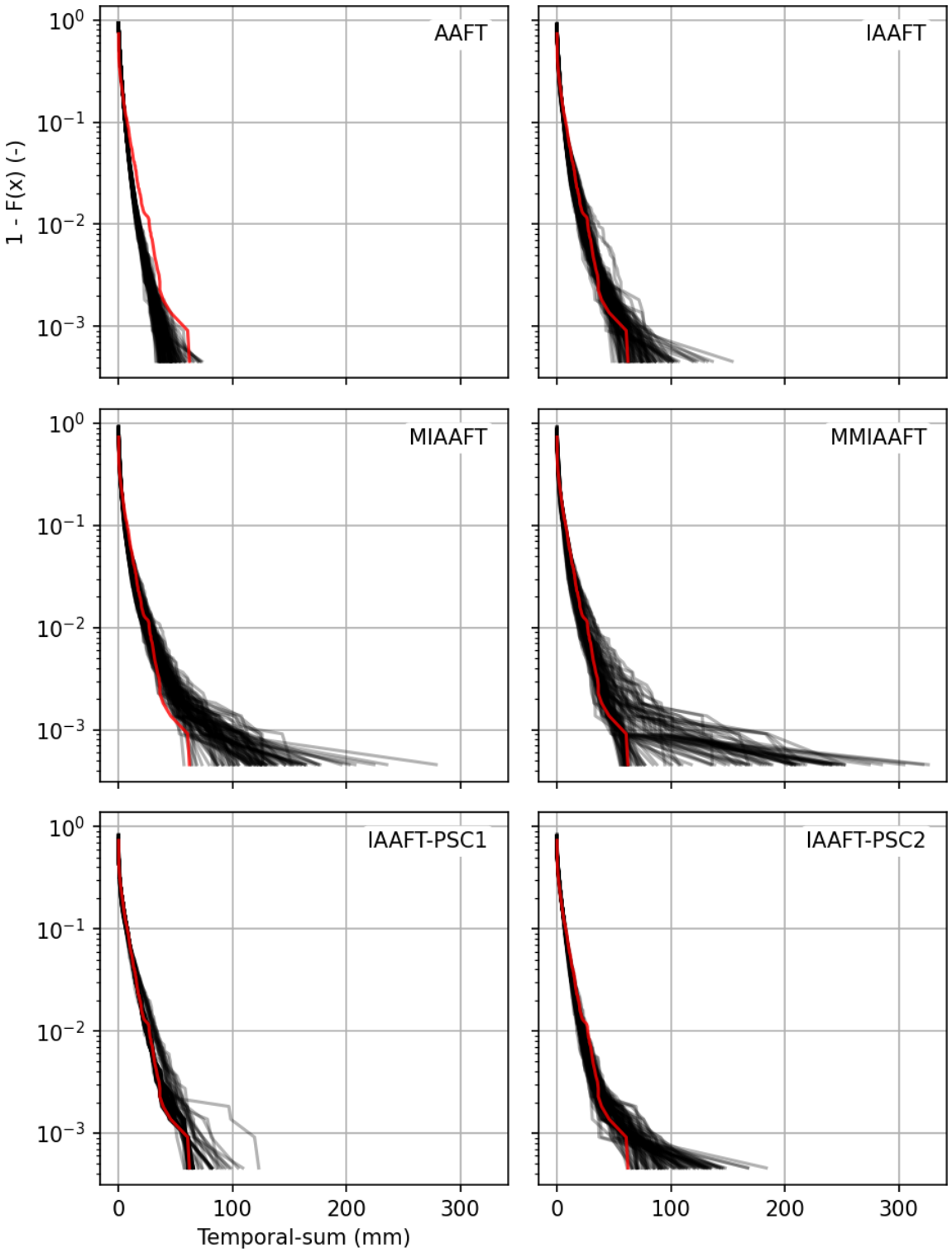


Figure 9.53: Distribution of hourly resampled to daily precipitation for a randomly selected station in the study area. Red are observed while black are the simulations.

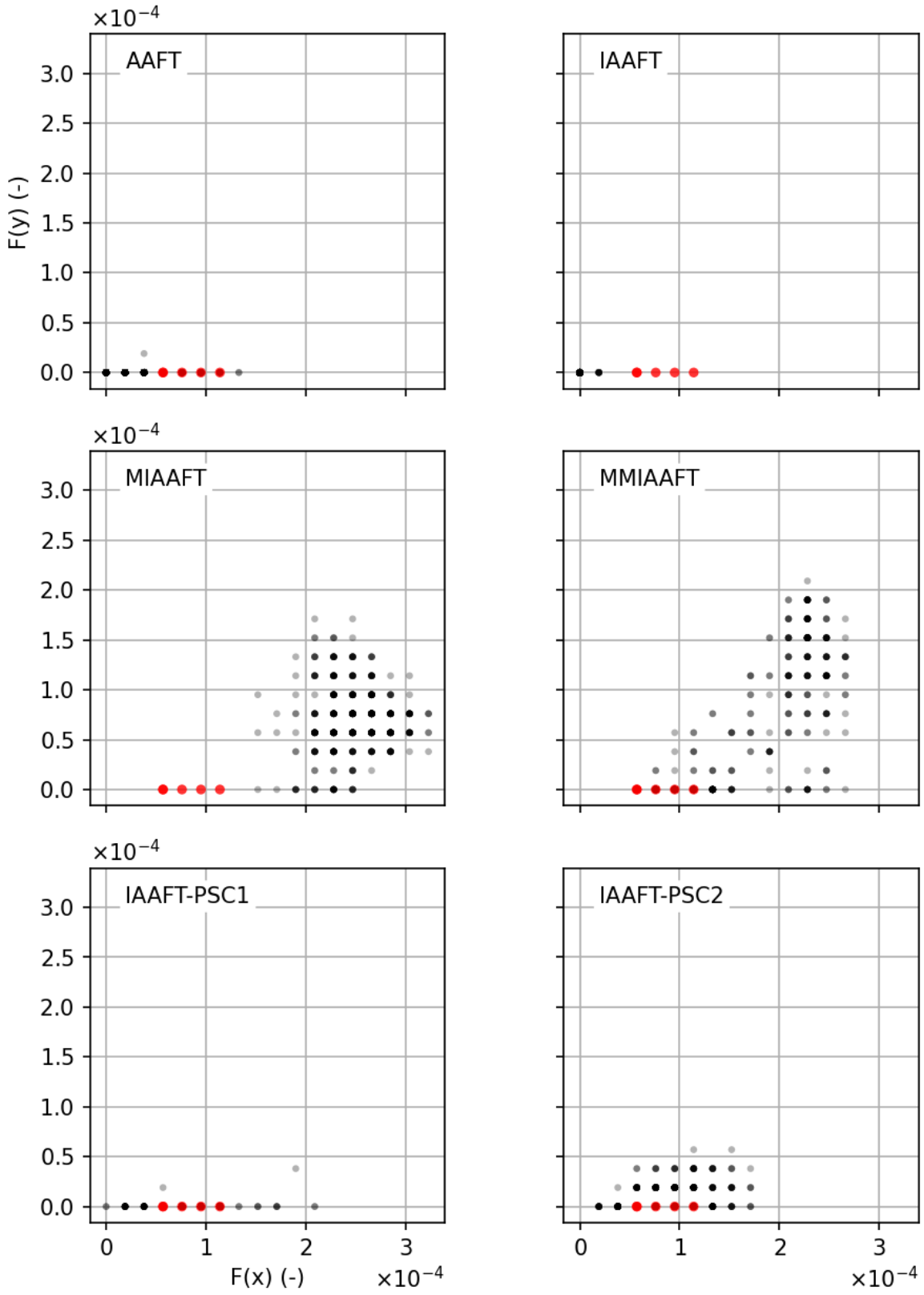


Figure 9.54: Conditional probability comparisons for daily precipitation of two out of four have values with return periods of six months or greater (horizontal axis) versus all of them having values equal or greater than six months years (vertical axis) at the same time step.

Finally, the following question can be answered: *which measures are to be conserved to have better representation of the spatial extremes?* One of the simulation methods, namely IAAFT, had no cross-correlation preservation but had the annual and seasonal cycles locked like the rest of the methods. It is the only method here that produced spatial-sums much lower than the rest for all the considered variables. And what were the properties that it lacked? From results it can be concluded that $CMCORR_p^d$ is the measure that consistently explains the desired behavior. Furthermore, it is evident that biases in the auto- and cross-ρ_p functions are also directly responsible for the tail behavior. The effects become clearer when looking at higher temporal aggregations. Considering the case of AAFT, which does well for the extremes despite having consistently lower cross-correlations and $CMCORR_p^d$. It was shown that it does keep (in this case) the correlation in the high frequency range. It should be kept in mind that measures cannot be interpreted one-to-one for discharge (a continuous distribution) and precipitation (a mixed distribution). An example could be of the daily discharge, where PA (Lumped) shows some underestimation for the $CMCORR_p^d$ and produces spatial extremes that are slightly underestimated. Comparing this to the case of daily precipitation, where AAFT shows an underestimation of the $CMCORR_p^d$'s final value, but does produce spatial extremes that envelop the reference very nicely. Contradicting information? The answer to this is, most probably, that discharge is a result of a low-pass filtering. High discharge values do not happen in isolation similar to precipitation. This means that, comparatively speaking, discharge spatial extremes are more explained by some lower frequencies as well.

Another important conclusion that can be drawn is that of simulating rainfall-runoff model inputs rather than simulating discharge directly. Judging from the properties of the HBV discharge compared to the reference and observed using simulated inputs, it is clear that they are much better in the reproduction of properties compared to direct discharge simulation. This boils down to the fact that precipitation is much more random/chaotic in nature as compared to discharge which is highly deterministic. Hence, using simulated precipitation combined with a rainfall-runoff model provides a better chance at simulating discharge series that have realistic properties in less computational time by needing less objective functions overall, even though temperature and potential evapotranspiration have to be simulated simultaneously as well. It is not important to do so for simulating series that better preserve simultaneous extremes only, but for other purposes that may require more realistic discharge series. It goes without saying, the final results depend on how the model and calibrations were, to begin with.

Finally, the simulation of the hourly precipitation is where the new variants shine performance-wise. There, it can be clearly seen that IAAFT-PSC1 and IAAFT-PSC2 are the winners in overall performance. This also bodes well for cases where finer temporal scales hold more relevance such as those of small river along cities or small catchments in general where time-of-concentrations are less than 24 hours.

9.9 Simulation methods' leader board

So, which method wins? What is recommended? The answer is: *it's depends!* All methods show biases in one form or another. AAFT performed very well overall for the distributions of extremes and the conditional multuplet distributions of exceedence probabilities for the daily case. It is surprising but true. Given the very little time that it takes to simulate, it is recommended for the daily case while only conserving the individual time series power spectra, cross phase spectra and (if possible) $CMCORR_p^d$. Then, the next obvious choices are IAAFT-PSC1 and PA (Spectral). MIAAFT or MMIAAFT are not recommended for the cross-cases as the results show that they can significantly over-estimate in some cases. For the daily auto-case, IAAFT-PSC2 is the method of choice as it was best able to conserve the properties at a point. For the hourly precipitation case, IAAFT-PSC2 performs very well and is recommended for multi-site simulation as it preserves the auto- and cross-properties of the precipitation. Finally, PA is recommended when enough computational power and time are available as it is able to produce series with arbitrarily specified properties.

10 Conclusions and Outlook

Existing and new methods to simulate time series using various methods were presented in this thesis. The main aim was to have time series that have properties similar to a reference with more focus on those that define the occurrence of spatial extremes in an objective manner. The simulated series with acceptable quality allows for the ability to test any system under scrutiny for many scenarios. It was demonstrated that the aspect of spatio-temporal coherence is very important and has to be reproduced by the simulations. It was also learned that precipitation and discharge properties are stable in space-time for long time periods with some drift as the environment evolves. The Earth's climate is under constant change be it due humans or otherwise. The methods proposed and tested here are not capable of taking such changes into account while simulating time series. Research in this area exists and is a focus of many these days. In future, the findings of others can be incorporated into these methods. For example, the consensus is that increasing temperatures result in more intense precipitation then, simply put, the outputs of the techniques here can be post-processed by, yet, another algorithm to adjust them accordingly. Coming back to learning, it was also learned that properties that define the overall behavior of precipitation and other variables in the copula domain do not have to be conserved exactly as they are only partly responsible for the cases of spatial extremes. The cross-correlations were concluded to be much more important though and even those need not be kept exactly the same. The coherence of high frequency components was found out to be the main indicator of simultaneous extremes in the frequency domain. This makes sense, as big floods are a result of the integration of the precipitation process over time where little deviations in timing can be forgiven. The important point is that the high values at multiple locations maintain their coherency similar to that of the reference. One more issue that was not addressed here was that of the *forecasting*. This requires much more work; but in theory, the methods presented here are capable of doing so. Techniques such as Random Mixing (Bárdossy and Hörning, 2016) exist already and can be used to have constrained realizations. A more realistic result is to have three-dimensional fields where precipitation is simulated on a plane at each time step while keeping coherence in space-time. The presented methods may be adopted to produce such results but then the main challenge is evaluating the results at the unknown locations and the other problem is that the spatial relationship at each time step may be unique. This could be something that can be addressed by future research. Tools presented here such as the d-dimensional cumulative marginal correlation periodogram ($CMCORR_p^d$) can be used to sample points randomly from fields and then compared with a reference. Even the outputs of existing weather generators can be tested for these properties by the various proposed measures.

The main contributions of this thesis, apart from the algorithms, are the newly introduced

objective functions and properties. Properties such as the d-dimensional cumulative periodogram ($CMPOW^d$) and d-dimensional cumulative mean correlation periodogram ($CMCORR^d$) and the d-asymmetries and its spectral derivatives. These could be used as standalone metrics similar to the Nash-Sutcliffe Efficiency to evaluate the results of any time series generator. It is not important that the simulated series have metrics exactly same as those of the observed but that they have negligible to little bias and show enough variance in their results. One point that the author wants to emphasize is that all the properties in the high dimensions (d-asymmetries) not be used as objective functions but only as final independent evaluation metrics. This is because as the number of dimensions grows larger while the number of points inside them remain the same, only a few points have to be moved around to produce a similar value to that of the reference leading users in to a false belief that the optimized time series are good. However, the various derivatives of $CMCORR^d$ can be used as they are not so sensitive. Another important point that is emphasized is that most metrics are based on copulas and show the performance on the whole and not just for the upper tails of the distributions which are a rather small (but very important) part of the entire story. Hence, they are more suited for the evaluation of model outputs where they may highlight any systematic problems that the model may have. Furthermore, using observed series, it was also shown how past time periods are different than recent ones i.e., how climate change might look like. It was also demonstrated that the spatial-sums of river discharge of catchments in the study area still showed an increasing behavior of extremes that my be expected in future as the recorded maxima do not show decorrelation in the upper tails of their distributions of sums. The same was the case for various groups of precipitation stations considered in the study area.

Coming to the methods more specifically, the variant of Phase Annealing presented here has one major drawback. That being the time it takes, to optimize a time series whose properties match those of the reference, is longer than the rest depending on the case. A couple of daily time series with a length of five years may take around an hour which is not bad but also not very useful for practical cases where extremes are required. If one were to simulate say a 50 years series for ten stations simultaneously, then the time to optimize goes into months. Such a setting is desirable but given the amount of time required, makes it impractical. Similarly, IAAFT-PSC variants introduced do produce series that have better cross-correlations but all of this at a loss of reduced accuracy of copula auto- and cross-asymmetries. However, the results showed that impact of these on the extremes are not significant and all of this within reasonable computational time because the number of evaluations required are much less comparatively. It was shown that even using simple algorithms, such as AAFT or MIAAFT results can be obtained that are acceptable. It shows that the problem may not be as complicated as it was thought to be in the beginning. Keeping a small subset of properties, e.g., $CMCORR_p^d$, seems to be enough. However, it should be kept in mind that the simple methods used, performed well because their annual and seasonal phases were locked. At least, it is one of the reasons why the simple ones also performed well. This is in contrast to a stationary process where such behaviors are generally not kept explicitly.

Regarding the problem of simulating extreme precipitation and discharges i.e., is it possible to simulate larger values of spatio-temporal sums as compared to observed cases?, which were among the main objectives of this thesis, the results are summarized as follows. For

precipitation, it was shown that the distributions of the sums of simulated series did surpass those of the observed in both space and time. The idea to test whether the sums in space-time are larger as compared to observed cases is important, mainly due to the fact that it is the larger values that cause destruction of property and life. For discharge, two approaches were shown. One was to simulate discharge at multiple locations directly and the other being simulating rainfall-runoff model inputs simultaneously and using them to obtain discharge via a model. The model being the HBV. Both approaches had problems. Discharge is a result of a complicated interaction of precipitation, temperature and the catchment. Simulating it directly through a statistical process is ill-advised as the time series generator does not take into account the various types of situations, say snow-melt or dry or wet catchment, into account. However, the spatial-sum distributions using methods that conserved cross-correlations showed very good results in their upper tails. Secondly, precipitation, temperature and potential evapotranspiration were simulated simultaneously together for various catchments. These were used as inputs for models that were calibrated with observed data already for computing discharge. This lead to the drawbacks of such an approach. The model output is not the same as observed. For one thing, almost all rainfall-runoff models suffer from the same problem that they never match the peak flows and here the case was no different. Only properties of model runoff could be compared for the simulated and observed inputs only and not observed discharge. As even the observed inputs when used for calibration do not produce an NSE of 1.0 or even close. Hence, only the properties of model discharges could be compared. This makes the results less useful because the extreme discharges created by models are consistently lower than those of the observed. But surprisingly, model discharges using simulated inputs did produce values that were higher than the observed. This was mainly due to the simulated precipitation where the spatial- and temporal-sum distributions did have values in the upper tails that were higher than those of the observed cases and all of this even when the distributions for all cases were exactly the same. This is good news because, resampling observed series to produce new scenarios can be much different than what was observed. However, it should not be seen as a replacement for what may take place in the future.

To summarize, in this thesis, propositions for new time series generators, objective functions and metrics that describe higher-order dependence were made. The main goal was to have simulations that mimic observed variables in their behavior of extremes in space-time. Measures were proposed to evaluate time series such as various $CMPOW$s, d-asymmetries, $CMPOW^d$ and $CMCORR^d$. It was shown that the higher frequencies of the $CMCORR^d$ spectra of given variables' time series can be used to differentiate between points whose extremes are more synchronized in time and those that do not. Consequently, this measure can be used to evaluate synthetic series for their properties compared to a reference. It was shown that even simple time series generators are not that bad to begin with, at least on the daily temporal scale. New algorithms (e.g., IAAFT-PSC1 and IAAFT-PSC2) were proposed to have better outputs. More performance metrics were proposed that can be used as objective functions to simulate and assess the quality of model outputs. For the daily resolution, all methods showed good performance for the extreme cases even when their auto- and cross-properties in space-time did not match those of the reference exactly. For the case of hourly precipitation, only the IAAFT-PSC2 algorithm worked satisfactorily.

In the future, attempts can be made to use generators that are even faster, if possible at all. The Fourier transform-based generators used here are mainly meant for linear time series generation but were adapted to do non-linear simulations. Signals that vary over long terms such as those of climate change were neither investigated nor dealt with in this study. Time series from whom such effects are removed should be compared to the original ones to assess how much the effect of these signals is on the final high-dimensional measures. A whole different field of study, *non-linear time series analysis*, exists to tackle problems in a more wholesome manner where measures such as the correlation dimension and the Lypunov's exponent with dynamical systems are the norm. An important feature of non-linear analysis is its use of *filters*. These allow for modification/handling of data for lower and higher quantiles in a different manner. This could be useful for variables such as precipitation and discharge where it is known that the dependence of smaller values is very much different compared to those of the larger ones. Another very important point to investigate is that of the high flows that are a result of the extrapolation of the rating curve and not actual measurements. All values of observed series are treated as *sacred* but this should not be the case for the extrapolation.

Bibliography

S. Ahmed and G. de Marsily. Comparison of geostatistical methods for estimating transmissivity using data transmissivity and specific capacity. *Water Resources Research*, 23: 1717–1737, 1987.

P. Ailliot, D. Allard, V. Monbet, and P. Naveau. Stochastic weather generators: an overview of weather type models. *Journal de la société française de statistique*, 156(1):101–113, 2015. URL http://www.numdam.org/item/JSFS_2015__156_1_101_0/.

A. Bárdossy and F. Anwar. Why do our rainfall–runoff models keep underestimating the peak flows? *Hydrology and Earth System Sciences*, 27(10):1987–2000, 2023. doi: 10.5194/hess-27-1987-2023. URL https://hess.copernicus.org/articles/27/1987/2023/.

A. Bárdossy and S. Hörning. Process-Driven Direction-Dependent Asymmetry: Identification and Quantification of Directional Dependence in Spatial Fields. *Mathematical Geosciences*, 49:871 – 891, 2017.

F. Beck. *Generation of Spatially Correlated Synthetic Rainfall Time Series in High Temporal Resolution - A Data Driven Approach*. PhD thesis, University of Stuttgart, 2013.

S. Bergström. *The HBV Model: Its Structure and Applications*. SMHI Reports Hydrology. SMHI, 1992. URL https://books.google.de/books?id=u7F7mwEACAAJ.

K. J. Beven. A history of the concept of time of concentration. *Hydrology and Earth System Sciences*, 24(5):2655–2670, 2020. doi: 10.5194/hess-24-2655-2020. URL https://hess.copernicus.org/articles/24/2655/2020/.

J. Bliefernicht and A. Bárdossy. Probabilistic forecast of daily areal precipitation focusing on extreme events. *Natural Hazards and Earth System Sciences*, 7(2):263–269, 2007. doi: 10.5194/nhess-7-263-2007. URL https://nhess.copernicus.org/articles/7/263/2007/.

G. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, 1976.

M. I. Brunner and E. Gilleland. Stochastic simulation of streamflow and spatial extremes: a continuous, wavelet-based approach. *Hydrology and Earth System Sciences*, 24(8):3967–3982, 2020. doi: 10.5194/hess-24-3967-2020. URL https://hess.copernicus.org/articles/24/3967/2020/.

A. Bárdossy and S. Hörning. Gaussian and non-gaussian inverse modeling of groundwater flow using copulas and random mixing. *Water Resources Research*, 52(6):4504–4526, 2016. doi: https://doi.org/10.1002/2014WR016820.

A. Bárdossy and E. J. Plate. Space-time model for daily rainfall using atmospheric circulation patterns. *Water Resources Research*, 28(5):1247–1259, 1992. doi: https://doi.org/10.1029/91WR02589.

A. Bárdossy, H. Giese, B. Haller, and J. Ruf. Erzeugung synthetischer Niederschlagsreihen in hoher zeitlicher Auflösung für Baden Württemberg. *Wasserwirtschaft*, 90(11):548–553, 2020.

C. Chatfield. *The Analysis of Time Series: An Introduction, Sixth Edition*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, 2016. ISBN 9780203491683. URL https://books.google.de/books?id=qKzyAbdaDFAC.

J. Chen, F. P. Brissette, and R. Leconte. Downscaling of weather generator parameters to quantify hydrological impacts of climate change. *Climate Research*, 51(3):185–200, 2012.

T. M. Cover and J. A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, USA, 2006. ISBN 0471241954.

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

I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992. doi: 10.1137/1.9781611970104. URL https://epubs.siam.org/doi/abs/10.1137/1.9781611970104.

L. P. de Figueiredo, D. Grana, and M. Le Ravalec. Revisited formulation and applications of FFT moving average. *Mathematical Geosciences*, 52(6):801–816, 2020.

M. Donatelli, G. Bellocchi, E. Habyarimana, S. Bregaglio, R. Confalonieri, and B. Baruth. *CLIMA: a weather generator framework*. Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, Australia, 2009. URL http://www.mssanz.org.au/modsim09/C3/donatelli_C3a.pdf.

DWD. https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/, 2019 2019.

J. I. R. Fernández. *High Order Interactions among environmental variables: Diagnostics and initial steps towards modeling*. PhD thesis, University of Stuttgart, 2013.

G. Flato, J. Marotzke, B. Abiodun, P. Braconnot, S. C. Chou, W. J. Collins, P. Cox, F. Driouech, S. Emori, V. Eyring, C. Forest, P. Gleckler, E. Guilyardi, C. Jakob, V. Kattsov, C. Reason, and M. Rummukaines. Evaluation of Climate Models. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. volume 5 of *Assessment Reports of IPCC*, pages 741–866. Cambridge University Press, 2013. URL https://elib.dlr.de/95697/.

J. Fourier. *Théorie analytique de la chaleur*. Chez Firmin Didot, père et fils, 1822. URL https://books.google.de/books?id=TDQJAAAAIAAJ.

A. Galka. *Topics in Nonlinear Time Series Analysis - With Implications for EEG Analysis*. World Scientific Publishing Company, Singapore, 2000. ISBN 981-02-4148-8.

C. Gao, M. J. Booij, and Y.-P. Xu. Development and hydrometeorological evaluation of a new stochastic daily rainfall model: Coupling Markov chain with rainfall event model. *Journal of Hydrology*, 589:125337, 2020. ISSN 0022-1694. doi: https://doi.org/10.1016/j.jhydrol.2020.125337.

H. Gebelein. Das statistische problem der korrelation als variations- und eigenwertproblem und sein zusammenhang mit der ausgleichsrechnung. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 21(6):364–379, 1941. doi: https://doi.org/10.1002/zamm.19410210604. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/zamm.19410210604.

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

P. Guthke. *Non-multi Gaussian spatial structures: Process-driven natural genesis, manifestation, modeling approaches, and influences on dependent processes.* Eigenverlag des Instituts Wasserbau, 2013.

Z. Hao and V. P. Singh. Review of dependence modeling in hydrology and water resources. *Progress in Physical Geography: Earth and Environment*, 40(4):549–578, 2016. doi: 10.1177/0309133316632460. URL https://doi.org/10.1177/0309133316632460.

G. Hargreaves and Z. Samani. Estimating potential evapotranspiration. *Journal of the Irrigation and Drainage Division*, 108(3):225–230, 1982.

S. Hörning and A. Bárdossy. Phase annealing for the conditional simulation of spatial random fields. *Computers and Geosciences*, 112:101–111, 2018. ISSN 0098-3004. doi: https://doi.org/10.1016/j.cageo.2017.12.008.

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

C. J. Keylock. A resampling method for generating synthetic hydrological time series with preservation of cross-correlative structure and higher-order properties. *Water Resources Research*, 48(12), 2012. doi: https://doi.org/10.1029/2012WR011923. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012WR011923.

C. Kilsby, P. Jones, A. Burton, A. Ford, H. Fowler, C. Harpham, P. James, A. Smith, and R. Wilby. A daily weather generator for use in climate change studies. *Environmental Modelling & Software*, 22(12):1705–1719, 2007. ISSN 1364-8152. doi: https://doi.org/10.1016/j.envsoft.2007.02.005.

S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983. ISSN 0036-8075. doi: 10.1126/science.220.4598.671. URL https://science.sciencemag.org/content/220/4598/671.

P. Krupskii. Copula-based measures of reflection and permutation asymmetry and statistical tests. *Statistical Papers*, 58(4):1165–1187, December 2017. doi: 10.1007/s00362-016-0743-1.

S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951. ISSN 00034851. URL http://www.jstor.org/stable/2236703.

G. Lancaster, D. Iatsenko, A. Pidde, V. Ticcinelli, and A. Stefanovska. Surrogate data for hypothesis testing of physical systems. *Physics Reports*, 748:1–60, July 2018. ISSN 0370-1573. doi: 10.1016/j.physrep.2018.06.001.

LUBW. https://udo.lubw.baden-wuerttemberg.de/public/, 2020.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953. doi: 10.1063/1.1699114. URL https://doi.org/10.1063/1.1699114.

T. Müller, T. Mosthaf, S. Gunzenhauser, J. Seidel, and A. Bárdossy. Grundlagenbericht Niederschlags-Simulator (NiedSim3), 2017.

T. Nakamura, M. Small, and Y. Hirata. Testing for nonlinearity in irregular fluctuations with long-term trends. *Phys. Rev. E*, 74:026205, Aug 2006. doi: 10.1103/PhysRevE.74.026205. URL https://link.aps.org/doi/10.1103/PhysRevE.74.026205.

J. Nash and J. Sutcliffe. River flow forecasting through conceptual models. 1. a discussion of principles. *Journal of Hydrology*, 10:282–290, 1970.

R. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer New York, 2007. ISBN 9780387286785. URL https://books.google.de/books?id=yexFAAAAQBAJ.

D. T. Nguyen and S.-T. Chen. Generating continuous rainfall time series with high temporal resolution by using a stochastic rainfall generator with a copula and modified huff rainfall curves. *Water*, 14(13), 2022. ISSN 2073-4441. doi: 10.3390/w14132123. URL https://www.mdpi.com/2073-4441/14/13/2123.

A. D. Nicks and G. A. Gander. CLIGEN: A weather generator for climate inputs to water resource and other models. In *Proc. Fifth Int. Conf. on Computers in Agriculture*, pages 903–909, 1994.

N. Peleg, S. Fatichi, A. Paschalis, P. Molnar, and P. Burlando. An advanced stochastic weather generator for simulating 2-D high-resolution climate variables. *Journal of Advances in Modeling Earth Systems*, 9(3):1595–1627, 2017. doi: https://doi.org/10.1002/2016MS000854.

P. Racsko, L. Szeidl, and M. Semenov. A serial approach to local stochastic weather models. *Ecological Modelling*, 57(1):27–41, 1991. ISSN 0304-3800. doi: https://doi.org/10.1016/0304-3800(91)90053-4.

M. L. Ravalec, B. Noetinger, and L. Y. Hu. The FFT moving average (FFT-MA) generator: An efficient numerical method for generating and conditioning gaussian simulations. *Mathematical Geology*, 32(6):701–723, 2000.

A. Rényi et al. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California, 1961.

C. W. Richardson. Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, 17(1):182–190, 1981. doi: https://doi.org/10.1029/WR017i001p00182.

C. W. Richardson. WGEN: A Model for Generating Daily Weather Variables. U.S. Dept. of Agriculture Agricultural Research Service ; National Technical Information Service (NTIS) Distributor, 1984.

J. Rosco and H. Joe. Measures of tail asymmetry for bivariate copulas. *Statistical Papers*, 54 (3):709–726, August 2013. doi: 10.1007/s00362-012-0457-y.

F. H. Ruymgaart and M. C. A. van Zuijlen. Asymptotic Normality of Multivariate Linear Rank Statistics in the Non-I.I.D. Case. *The Annals of Statistics*, 6(3):588 – 602, 1978. doi: 10.1214/aos/1176344203. URL https://doi.org/10.1214/aos/1176344203.

D. Schlabing, M. Frassl, M. Eder, K. Rinke, and A. Bárdossy. Use of a weather generator for simulating climate change effects on ecosystems: A case study on Lake Constance. *Environmental Modelling & Software*, 61:326–338, 2014. ISSN 1364-8152. doi: https://doi.org/10.1016/j.envsoft.2014.06.028.

F. Schmid and R. Schmidt. Multivariate conditional versions of spearman's rho and related measures of tail dependence. *Journal of Multivariate Analysis*, 98(6):1123–1140, 2007. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2006.05.005. URL https://www.sciencedirect.com/science/article/pii/S0047259X06000662.

T. Schreiber and A. Schmitz. Improved surrogate data for nonlinearity tests. *Phys. Rev. Lett.*, 77:635–638, Jul 1996. doi: 10.1103/PhysRevLett.77.635. URL https://link.aps.org/doi/10.1103/PhysRevLett.77.635.

M. Semenov and E. Barrow. LARS-WG a stochastic weather generator for use in climate impact studies. User's manual, Version3.0, 2002.

C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948. doi: https://doi.org/10.1002/j.1538-7305.1948.tb01338.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/j.1538-7305.1948.tb01338.x.

A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de ÍInstitut Statistique de ÍUniversitè de Paris*, 8:229–231, 1959.

C. Stöckle, G. S. Campbell, and R. Nelson. ClimGen manual. *Biological Systems Engineering Department, Washington State University, Pullman, WA*, page 28, 1999.

R. Storn and K. Price. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359, 1997.

J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. Doyne Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1):77–94, 1992. ISSN 0167-2789. doi: https://doi.org/10.1016/0167-2789(92)90102-S. URL https://www.sciencedirect.com/science/article/pii/016727899290102S.

C. Tsallis. Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52:479–487, 1988.

S. Vandenberghe, N. E. C. Verhoest, E. Buyse, and B. De Baets. A stochastic design rainfall generator based on copulas and mass curves. *Hydrology and Earth System Sciences*, 14(12): 2429–2442, 2010. doi: 10.5194/hess-14-2429-2010. URL https://hess.copernicus.org/articles/14/2429/2010/.

V. Černý. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of Optimization Theory and Applications*, 45(1):41–51, Jan. 1985. ISSN 0022-3239. URL http://dx.doi.org/10.1007/bf00940812.

V. Venema, F. Ament, and C. Simmer. A stochastic iterative amplitude adjusted fourier transform algorithm with improved accuracy. *Nonlinear Processes in Geophysics*, 13(3): 321–328, 2006. doi: 10.5194/npg-13-321-2006. URL https://npg.copernicus.org/articles/13/321/2006/.

E. F. Wolff. N-dimensional measures of dependence. *Stochastica*, 4(3):175–188, 1980. URL http://eudml.org/doc/38838.

Institut für Wasser- und Umweltsystemmodellierung
Universität Stuttgart

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

Direktoren
Prof. Dr.-Ing. Rainer Helmig
Prof. Dr.-Ing. Wolfgang Nowak
Prof. Dr.-Ing. Silke Wieprecht

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

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

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

Lehrstuhl für Stochastische Simulation und Sicherheitsforschung für Hydrosysteme
Leiter: Prof. Dr.-Ing. Wolfgang Nowak
Stellv.: apl. Prof. Dr.-Ing. Sergey Oladyshkin
Hydrogeophysik der Vadosen Zone
(mit Forschungszentrum Jülich)
Leiter: Prof. Dr. J.A. Sander Huisman

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

Verzeichnis der Mitteilungshefte

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

58 Kirchenbaur, Klaus: *Mikroprozessorgesteuerte Erfassung instationärer Druckfelder am Beispiel seegangsbelasteter Baukörper*, 1985, ISBN 3-921694-58-2

59 Kobus, Helmut (Hrsg.): *Modellierung des großräumigen Wärme- und Schadstofftransports im Grundwasser*, Tätigkeitsbericht 1984/85 (DFG-Forschergruppe an den Universitäten Hohenheim, Karlsruhe und Stuttgart), 1985, ISBN 3-921694-59-0, *vergriffen*

60 Spitz, Karlheinz: *Dispersion in porösen Medien: Einfluß von Inhomogenitäten und Dichteunterschieden*, 1985, ISBN 3-921694-60-4, *vergriffen*

61 Kobus, Helmut: *An Introduction to Air-Water Flows in Hydraulics*, 1985, ISBN 3-921694-61-2

62 Kaleris, Vassilios: *Erfassung des Austausches von Oberflächen- und Grundwasser in horizontalebenen Grundwassermodellen*, 1986, ISBN 3-921694-62-0

63 Herr, Michael: *Grundlagen der hydraulischen Sanierung verunreinigter Porengrundwasserleiter*, 1987, ISBN 3-921694-63-9

64 Marx, Walter: *Berechnung von Temperatur und Spannung in Massenbeton infolge Hydratation*, 1987, ISBN 3-921694-64-7

65 Koschitzky, Hans-Peter: *Dimensionierungskonzept für Sohlbelüfter in Schußrinnen zur Vermeidung von Kavitationsschäden*, 1987, ISBN 3-921694-65-5

66 Kobus, Helmut (Hrsg.): *Modellierung des großräumigen Wärme- und Schadstofftransports im Grundwasser*, Tätigkeitsbericht 1986/87 (DFG-Forschergruppe an den Universitäten Hohenheim, Karlsruhe und Stuttgart) 1987, ISBN 3-921694-66-3

67 Söll, Thomas: *Berechnungsverfahren zur Abschätzung anthropogener Temperaturanomalien im Grundwasser*, 1988, ISBN 3-921694-67-1

68 Dittrich, Andreas; Westrich, Bernd: *Bodenseeufererosion, Bestandsaufnahme und Bewertung*, 1988, ISBN 3-921694-68-X, *vergriffen*

69 Huwe, Bernd; van der Ploeg, Rienk R.: *Modelle zur Simulation des Stickstoffhaushaltes von Standorten mit unterschiedlicher landwirtschaftlicher Nutzung*, 1988, ISBN 3-921694-69-8, *vergriffen*

70 Stephan, Karl: *Integration elliptischer Funktionen*, 1988, ISBN 3-921694-70-1

71 Kobus, Helmut; Zilliox, Lothaire (Hrsg.): *Nitratbelastung des Grundwassers, Auswirkungen der Landwirtschaft auf die Grundwasser- und Rohwasserbeschaffenheit und Maßnahmen zum Schutz des Grundwassers*. Vorträge des deutsch-französischen Kolloquiums am 6. Oktober 1988, Universitäten Stuttgart und Louis Pasteur Strasbourg (Vorträge in deutsch oder französisch, Kurzfassungen zweisprachig), 1988, ISBN 3-921694-71-X

72 Soyeaux, Renald: *Unterströmung von Stauanlagen auf klüftigem Untergrund unter Berücksichtigung laminarer und turbulenter Fließzustände*,1991, ISBN 3-921694-72-8

73 Kohane, Roberto: *Berechnungsmethoden für Hochwasserabfluß in Fließgewässern mit überströmten Vorländern*, 1991, ISBN 3-921694-73-6

74 Hassinger, Reinhard: *Beitrag zur Hydraulik und Bemessung von Blocksteinrampen in flexibler Bauweise*, 1991, ISBN 3-921694-74-4, *vergriffen*

75 Schäfer, Gerhard: *Einfluß von Schichtenstrukturen und lokalen Einlagerungen auf die Längsdispersion in Porengrundwasserleitern*, 1991, ISBN 3-921694-75-2

76 Giesecke, Jürgen: *Vorträge, Wasserwirtschaft in stark besiedelten Regionen; Umweltforschung mit Schwerpunkt Wasserwirtschaft*, 1991, ISBN 3-921694-76-0

77 Huwe, Bernd: *Deterministische und stochastische Ansätze zur Modellierung des Stickstoffhaushalts landwirtschaftlich genutzter Flächen auf unterschiedlichem Skalenniveau*, 1992, ISBN 3-921694-77-9, *vergriffen*

78 Rommel, Michael: *Verwendung von Kluftdaten zur realitätsnahen Generierung von Kluftnetzen mit anschließender laminar-turbulenter Strömungsberechnung*, 1993, ISBN 3-92 1694-78-7

79 Marschall, Paul: *Die Ermittlung lokaler Stofffrachten im Grundwasser mit Hilfe von Einbohrloch-Meßverfahren*, 1993, ISBN 3-921694-79-5, *vergriffen*

80 Ptak, Thomas: *Stofftransport in heterogenen Porenaquiferen: Felduntersuchungen und stochastische Modellierung*, 1993, ISBN 3-921694-80-9, *vergriffen*

81 Haakh, Frieder: *Transientes Strömungsverhalten in Wirbelkammern,* 1993, ISBN 3-921694-81-7

82 Kobus, Helmut; Cirpka, Olaf; Barczewski, Baldur; Koschitzky, Hans-Peter: *Versuchseinrichtung zur Grundwasser- und Altlastensanierung VEGAS, Konzeption und Programmrahmen*, 1993, ISBN 3-921694-82-5

83 Zang, Weidong: *Optimaler Echtzeit-Betrieb eines Speichers mit aktueller Abflußregenerierung*, 1994, ISBN 3-921694-83-3, *vergriffen*

84 Franke, Hans-Jörg: *Stochastische Modellierung eines flächenhaften Stoffeintrages und Transports in Grundwasser am Beispiel der Pflanzenschutzmittelproblematik*, 1995, ISBN 3-921694-84-1

85 Lang, Ulrich: *Simulation regionaler Strömungs- und Transportvorgänge in Karstaquiferen mit Hilfe des Doppelkontinuum-Ansatzes: Methodenentwicklung und Parameteridenti-fikation*, 1995, ISBN 3-921694-85-X, *vergriffen*

86 Helmig, Rainer: *Einführung in die Numerischen Methoden der Hydromechanik*, 1996, ISBN 3-921694-86-8, *vergriffen*

87 Cirpka, Olaf: *CONTRACT: A Numerical Tool for Contaminant Transport and Chemical Transformations - Theory and Program Documentation -*, 1996, ISBN 3-921694-87-6

88 Haberlandt, Uwe: *Stochastische Synthese und Regionalisierung des Niederschlages für Schmutzfrachtberechnungen*, 1996, ISBN 3-921694-88-4

89 Croisé, Jean: *Extraktion von flüchtigen Chemikalien aus natürlichen Lockergesteinen mittels erzwungener Luftströmung*, 1996, ISBN 3-921694-89-2, *vergriffen*

90 Jorde, Klaus: *Ökologisch begründete, dynamische Mindestwasserregelungen bei Ausleitungskraftwerken*, 1997, ISBN 3-921694-90-6, *vergriffen*

91 Helmig, Rainer: *Gekoppelte Strömungs- und Transportprozesse im Untergrund - Ein Beitrag zur Hydrosystemmodellierung-*, 1998, ISBN 3-921694-91-4, *vergriffen*

92 Emmert, Martin: *Numerische Modellierung nichtisothermer Gas-Wasser Systeme in porösen Medien,* 1997, ISBN 3-921694-92-2

93 Kern, Ulrich: *Transport von Schweb- und Schadstoffen in staugeregelten Fließgewässern am Beispiel des Neckars*, 1997, ISBN 3-921694-93-0, *vergriffen*

94 Förster, Georg: *Druckstoßdämpfung durch große Luftblasen in Hochpunkten von Rohrleitungen 1997*, ISBN 3-921694-94-9

95 Cirpka, Olaf: *Numerische Methoden zur Simulation des reaktiven Mehrkomponententransports im Grundwasser*, 1997, ISBN 3-921694-95-7, *vergriffen*

96 Färber, Arne: *Wärmetransport in der ungesättigten Bodenzone: Entwicklung einer thermischen In-situ-Sanierungstechnologie*, 1997, ISBN 3-921694-96-5

97 Betz, Christoph: *Wasserdampfdestillation von Schadstoffen im porösen Medium: Entwicklung einer thermischen In-situ-Sanierungstechnologie*, 1998, SBN 3-921694-97-3

98 Xu, Yichun: *Numerical Modeling of Suspended Sediment Transport in Rivers*, 1998, ISBN 3-921694-98-1, *vergriffen*

99 Wüst, Wolfgang: *Geochemische Untersuchungen zur Sanierung CKW-kontaminierter Aquifere mit Fe(0)-Reaktionswänden,* 2000, ISBN 3-933761-02-2

100 Sheta, Hussam: *Simulation von Mehrphasenvorgängen in porösen Medien unter Einbeziehung von Hysterese-Effekten*, 2000, ISBN 3-933761-03-4

101 Ayros, Edwin: *Regionalisierung extremer Abflüsse auf der Grundlage statistischer Verfahren*, 2000, ISBN 3-933761-04-2, *vergriffen*

102 Huber, Ralf: *Compositional Multiphase Flow and Transport in Heterogeneous Porous Media*, 2000, ISBN 3-933761-05-0

103 Braun, Christopherus: *Ein Upscaling-Verfahren für Mehrphasenströmungen in porösen Medien*, 2000, ISBN 3-933761-06-9

104 Hofmann, Bernd: *Entwicklung eines rechnergestützten Managementsystems zur Beurteilung von Grundwasserschadensfällen*, 2000, ISBN 3-933761-07-7

105 Class, Holger: *Theorie und numerische Modellierung nichtisothermer Mehrphasenprozesse in NAPL-kontaminierten porösen Medien,* 2001, ISBN 3-933761-08-5

106 Schmidt, Reinhard: *Wasserdampf- und Heißluftinjektion zur thermischen Sanierung kontaminierter Standorte*, 2001, ISBN 3-933761-09-3

107 Josef, Reinhold: *Schadstoffextraktion mit hydraulischen Sanierungsverfahren unter Anwendung von grenzflächenaktiven Stoffen*, 2001, ISBN 3-933761-10-7

108 Schneider, Matthias: *Habitat- und Abflussmodellierung für Fließgewässer mit unscharfen Berechnungsansätzen*, 2001, ISBN 3-933761-11-5

109 Rathgeb, Andreas: *Hydrodynamische Bemessungsgrundlagen für Lockerdeckwerke an überströmbaren Erddämmen*, 2001, ISBN 3-933761-12-3

110 Lang, Stefan: *Parallele numerische Simulation instätionärer Probleme mit adaptiven Methoden auf unstrukturierten Gittern*, 2001, ISBN 3-933761-13-1

111 Appt, Jochen; Stumpp Simone: *Die Bodensee-Messkampagne 2001, IWS/CWR Lake Constance Measurement Program 2001*, 2002, ISBN 3-933761-14-X

112 Heimerl, Stephan: *Systematische Beurteilung von Wasserkraftprojekten,* 2002, ISBN 3-933761-15-8, *vergriffen*

113 Iqbal, Amin: *On the Management and Salinity Control of Drip Irrigation,* 2002, ISBN 3-933761-16-6

114 Silberhorn-Hemminger, Annette: *Modellierung von Kluftaquifersystemen: Geostatistische Analyse und deterministisch-stochastische Kluftgenerierung,* 2002, ISBN 3-933761-17-4

115 Winkler, Angela: *Prozesse des Wärme- und Stofftransports bei der In-situ-Sanierung mit festen Wärmequellen,* 2003, ISBN 3-933761-18-2

116 Marx, Walter: *Wasserkraft, Bewässerung, Umwelt - Planungs- und Bewertungsschwerpunkte der Wasserbewirtschaftung,* 2003, ISBN 3-933761-19-0

117 Hinkelmann, Reinhard: *Efficient Numerical Methods and Information-Processing Techniques in Environment Water,* 2003, ISBN 3-933761-20-4

118 Samaniego-Eguiguren, Luis Eduardo: *Hydrological Consequences of Land Use / Land Cover and Climatic Changes in Mesoscale Catchments*, 2003, ISBN 3-933761-21-2

119 Neunhäuserer, Lina: *Diskretisierungsansätze zur Modellierung von Strömungs- und Transportprozessen in geklüftet-porösen Medien*, 2003, ISBN 3-933761-22-0

120 Paul, Maren: *Simulation of Two-Phase Flow in Heterogeneous Poros Media with Adaptive Methods*, 2003, ISBN 3-933761-23-9

121 Ehret, Uwe: *Rainfall and Flood Nowcasting in Small Catchments using Weather Radar*, 2003, ISBN 3-933761-24-7

122 Haag, Ingo: *Der Sauerstoffhaushalt staugeregelter Flüsse am Beispiel des Neckars - Analysen, Experimente, Simulationen -,* 2003, ISBN 3-933761-25-5

123 Appt, Jochen: *Analysis of Basin-Scale Internal Waves in Upper Lake Constance,* 2003, ISBN 3-933761-26-3

124 Hrsg.: Schrenk, Volker; Batereau, Katrin; Barczewski, Baldur; Weber, Karolin und Koschitzky, Hans-Peter: *Symposium Ressource Fläche und VEGAS - Statuskolloquium 2003, 30. September und 1. Oktober 2003,* 2003, ISBN 3-933761-27-1

125 Omar Khalil Ouda: *Optimisation of Agricultural Water Use: A Decision Support System for the Gaza Strip,* 2003, ISBN 3-933761-28-0

126 Batereau, Katrin: *Sensorbasierte Bodenluftmessung zur Vor-Ort-Erkundung von Schadensherden im Untergrund*, 2004, ISBN 3-933761-29-8

127 Witt, Oliver: *Erosionsstabilität von Gewässersedimenten mit Auswirkung auf den Stofftransport bei Hochwasser am Beispiel ausgewählter Stauhaltungen des Oberrheins*, 2004, ISBN 3-933761-30-1

128 Jakobs, Hartmut: *Simulation nicht-isothermer Gas-Wasser-Prozesse in komplexen Kluft-Matrix-Systemen*, 2004, ISBN 3-933761-31-X

129 Li, Chen-Chien: *Deterministisch-stochastisches Berechnungskonzept zur Beurteilung der Auswirkungen erosiver Hochwasserereignisse in Flussstauhaltungen,* 2004, ISBN 3-933761-32-8

130 Reichenberger, Volker; Helmig, Rainer; Jakobs, Hartmut; Bastian, Peter; Niessner, Jennifer: *Complex Gas-Water Processes in Discrete Fracture-Matrix Systems: Up-scaling, Mass-Conservative Discretization and Efficient Multilevel Solution,* 2004, ISBN 3-933761-33-6

131 Hrsg.: Barczewski, Baldur; Koschitzky, Hans-Peter; Weber, Karolin; Wege, Ralf: *VEGAS - Statuskolloquium 2004,* Tagungsband zur Veranstaltung am 05. Oktober 2004 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2004, ISBN 3-933761-34-4

132 Asie, Kemal Jabir: *Finite Volume Models for Multiphase Multicomponent Flow through Porous Media*. 2005, ISBN 3-933761-35-2

133 Jacoub, George: *Development of a 2-D Numerical Module for Particulate Contaminant Transport in Flood Retention Reservoirs and Impounded Rivers,* 2004, ISBN 3-933761-36-0

134 Nowak, Wolfgang: *Geostatistical Methods for the Identification of Flow and Transport Parameters in the Subsurface,* 2005, ISBN 3-933761-37-9

135 Süß, Mia: *Analysis of the influence of structures and boundaries on flow and transport processes in fractured porous media*, 2005, ISBN 3-933761-38-7

136 Jose, Surabhin Chackiath: *Experimental Investigations on Longitudinal Dispersive Mixing in Heterogeneous Aquifers,* 2005, ISBN: 3-933761-39-5

137 Filiz, Fulya: *Linking Large-Scale Meteorological Conditions to Floods in Mesoscale Catchments,* 2005, ISBN 3-933761-40-9

138 Qin, Minghao: *Wirklichkeitsnahe und recheneffiziente Ermittlung von Temperatur und Spannungen bei großen RCC-Staumauern*, 2005, ISBN 3-933761-41-7

139 Kobayashi, Kenichiro: *Optimization Methods for Multiphase Systems in the Subsurface - Application to Methane Migration in Coal Mining Areas*, 2005, ISBN 3-933761-42-5

140 Rahman, Md. Arifur: *Experimental Investigations on Transverse Dispersive Mixing in Heterogeneous Porous Media,* 2005, ISBN 3-933761-43-3

141 Schrenk, Volker: *Ökobilanzen zur Bewertung von Altlastensanierungsmaßnahmen,* 2005, ISBN 3-933761-44-1

142 Hundecha, Hirpa Yeshewatesfa: *Regionalization of Parameters of a Conceptual Rainfall-Runoff Model*, 2005, ISBN: 3-933761-45-X

143 Wege, Ralf: *Untersuchungs- und Überwachungsmethoden für die Beurteilung natürlicher Selbstreinigungsprozesse im Grundwasser,* 2005, ISBN 3-933761-46-8

144 Breiting, Thomas: *Techniken und Methoden der Hydroinformatik - Modellierung von komplexen Hydrosystemen im Untergrund*, 2006, ISBN 3-933761-47-6

145 Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Müller, Martin: *Ressource Untergrund: 10 Jahre VEGAS: Forschung und Technologieentwicklung zum Schutz von Grundwasser und Boden,* Tagungsband zur Veranstaltung am 28. und 29. September 2005 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2005, ISBN 3-933761-48-4

146 Rojanschi, Vlad: *Abflusskonzentration in mesoskaligen Einzugsgebieten unter Berücksichtigung des Sickerraumes,* 2006, ISBN 3-933761-49-2

147 Winkler, Nina Simone: *Optimierung der Steuerung von Hochwasserrückhaltebeckensystemen,* 2006, ISBN 3-933761-50-6

148 Wolf, Jens: *Räumlich differenzierte Modellierung der Grundwasserströmung alluvialer Aquifere für mesoskalige Einzugsgebiete*, 2006, ISBN: 3-933761-51-4

149 Kohler, Beate: *Externe Effekte der Laufwasserkraftnutzung*, 2006, ISBN 3-933761-52-2

150 Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Stuhrmann, Matthias: *VEGAS-Statuskolloquium 2006,* Tagungsband zur Veranstaltung am 28. September 2006 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2006, ISBN 3-933761-53-0

151 Niessner, Jennifer: *Multi-Scale Modeling of Multi-Phase - Multi-Component Processes in Heterogeneous Porous Media*, 2006, ISBN 3-933761-54-9

152 Fischer, Markus: *Beanspruchung eingeerdeter Rohrleitungen infolge Austrocknung bindiger Böden,* 2006, ISBN 3-933761-55-7

153 Schneck, Alexander: *Optimierung der Grundwasserbewirtschaftung unter Berücksichtigung der Belange der Wasserversorgung, der Landwirtschaft und des Naturschutzes*, 2006, ISBN 3-933761-56-5

154 Das, Tapash: *The Impact of Spatial Variability of Precipitation on the Predictive Uncertainty of Hydrological Models,* 2006, ISBN 3-33761-57-3

155 Bielinski, Andreas: *Numerical Simulation of CO_2 sequestration in geological formations*, 2007, ISBN 3-933761-58-1

156 Mödinger, Jens: *Entwicklung eines Bewertungs- und Entscheidungsunterstützungssystems für eine nachhaltige regionale Grundwasserbewirtschaftung*, 2006, ISBN 3-933761-60-3

157 Manthey, Sabine: *Two-phase flow processes with dynamic effects in porous media - parameter estimation and simulation,* 2007, ISBN 3-933761-61-1

158 Pozos Estrada, Oscar: *Investigation on the Effects of Entrained Air in Pipelines,* 2007, ISBN 3-933761-62-X

159 Ochs, Steffen Oliver: *Steam injection into saturated porous media – process analysis including experimental and numerical investigations*, 2007, ISBN 3-933761-63-8

160 Marx, Andreas: *Einsatz gekoppelter Modelle und Wetterradar zur Abschätzung von Niederschlagsintensitäten und zur Abflussvorhersage, 2007,* ISBN 3-933761-64-6

161 Hartmann, Gabriele Maria: *Investigation of Evapotranspiration Concepts in Hydrological Modelling for Climate Change Impact Assessment*, 2007, ISBN 3-933761-65-4

162 Kebede Gurmessa, Tesfaye: *Numerical Investigation on Flow and Transport Characteristics to Improve Long-Term Simulation of Reservoir Sedimentation*, 2007, ISBN 3-933761-66-2

163 Trifković, Aleksandar: *Multi-objective and Risk-based Modelling Methodology for Planning, Design and Operation of Water Supply Systems*, 2007, ISBN 3-933761-67-0

164 Götzinger, Jens: *Distributed Conceptual Hydrological Modelling - Simulation of Climate, Land Use Change Impact and Uncertainty Analysis*, 2007, ISBN 3-933761-68-9

165 Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Stuhrmann, Matthias: *VEGAS – Kolloquium 2007,* Tagungsband zur Veranstaltung am 26. September 2007 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2007, ISBN 3-933761-69-7

166 Freeman, Beau: *Modernization Criteria Assessment for Water Resources Planning; Klamath Irrigation Project, U.S.*, 2008, ISBN 3-933761-70-0

167 Dreher, Thomas: *Selektive Sedimentation von Feinstschwebstoffen in Wechselwirkung mit wandnahen turbulenten Strömungsbedingungen*, 2008, ISBN 3-933761-71-9

168 Yang, Wei: *Discrete-Continuous Downscaling Model for Generating Daily Precipitation Time Series*, 2008, ISBN 3-933761-72-7

169 Kopecki, Ianina: *Calculational Approach to FST-Hemispheres for Multiparametrical Benthos Habitat Modelling,* 2008*,* ISBN 3-933761-73-5

170 Brommundt, Jürgen: *Stochastische Generierung räumlich zusammenhängender Niederschlagszeitreihen*, 2008, ISBN 3-933761-74-3

171 Papafotiou, *Alexandros: Numerical Investigations of the Role of Hysteresis in Heterogeneous Two-Phase Flow Systems*, 2008, ISBN 3-933761-75-1

172 He, Yi: *Application of a Non-Parametric Classification Scheme to Catchment Hydrology*, 2008, ISBN 978-3-933761-76-7

173 Wagner, Sven: *Water Balance in a Poorly Gauged Basin in West Africa Using Atmospheric Modelling and Remote Sensing Information*, 2008, ISBN 978-3-933761-77-4

174 Hrsg.: Braun, Jürgen; Koschitzky, Hans-Peter; Stuhrmann, Matthias; Schrenk, Volker: *VEGAS-Kolloquium 2008 Ressource Fläche III*, Tagungsband zur Veranstaltung am 01. Oktober 2008 an der Universität Stuttgart, Campus Stuttgart-Vaihingen, 2008, ISBN 978-3-933761-78-1

175 Patil, Sachin: *Regionalization of an Event Based Nash Cascade Model for Flood Predictions in Ungauged Basins*, 2008, ISBN 978-3-933761-79-8

176 Assteerawatt, Anongnart: *Flow and Transport Modelling of Fractured Aquifers based on a Geostatistical Approach,* 2008, ISBN 978-3-933761-80-4

177 Karnahl, Joachim Alexander: *2D numerische Modellierung von multifraktionalem Schwebstoff- und Schadstofftransport in Flüssen*, 2008, ISBN 978-3-933761-81-1

178 Hiester, Uwe: *Technologieentwicklung zur In-situ-Sanierung der ungesättigten Bodenzone mit festen Wärmequellen,* 2009, ISBN 978-3-933761-82-8

179 Laux, Patrick: *Statistical Modeling of Precipitation for Agricultural Planning in the Volta Basin of West Africa,* 2009, ISBN 978-3-933761-83-5

180 Ehsan, Saqib: *Evaluation of Life Safety Risks Related to Severe Flooding,* 2009, ISBN 978-3-933761-84-2

181 Prohaska, Sandra: *Development and Application of a 1D Multi-Strip Fine Sediment Transport Model for Regulated Rivers,* 2009, ISBN 978-3-933761-85-9

182 Kopp, Andreas: *Evaluation of CO_2 Injection Processes in Geological Formations for Site Screening,* 2009, ISBN 978-3-933761-86-6

183 Ebigbo, Anozie: *Modelling of biofilm growth and its influence on CO_2 and water (two-phase) flow in porous media,* 2009, ISBN 978-3-933761-87-3

184 Freiboth, Sandra: *A phenomenological model for the numerical simulation of multiphase multicomponent processes considering structural alterations of porous media*, 2009, ISBN 978-3-933761-88-0

185 Zöllner, Frank: *Implementierung und Anwendung netzfreier Methoden im Konstruktiven Wasserbau und in der Hydromechanik*, 2009, ISBN 978-3-933761-89-7

186 Vasin, Milos: *Influence of the soil structure and property contrast on flow and transport in the unsaturated zone*, 2010, ISBN 978-3-933761-90-3

187 Li, Jing: *Application of Copulas as a New Geostatistical Tool*, 2010, ISBN 978-3-933761-91-0

188 AghaKouchak, Amir: *Simulation of Remotely Sensed Rainfall Fields Using Copulas*, 2010, ISBN 978-3-933761-92-7

189 Thapa, Pawan Kumar: *Physically-based spatially distributed rainfall runoff modelling for soil erosion estimation*, 2010, ISBN 978-3-933761-93-4

190 Wurms, Sven: *Numerische Modellierung der Sedimentationsprozesse in Retentionsanlagen zur Steuerung von Stoffströmen bei extremen Hochwasserabflussereignissen*, 2011, ISBN 978-3-933761-94-1

191 Merkel, Uwe: *Unsicherheitsanalyse hydraulischer Einwirkungen auf Hochwasserschutzdeiche und Steigerung der Leistungsfähigkeit durch adaptive Strömungsmodellierung*, 2011, ISBN 978-3-933761-95-8

192 Fritz, Jochen: *A Decoupled Model for Compositional Non-Isothermal Multiphase Flow in Porous Media and Multiphysics Approaches for Two-Phase Flow*, 2010, ISBN 978-3-933761-96-5

193 Weber, Karolin (Hrsg.): *12. Treffen junger WissenschaftlerInnen an Wasserbauinstituten*, 2010, ISBN 978-3-933761-97-2

194 Bliefernicht, Jan-Geert: *Probability Forecasts of Daily Areal Precipitation for Small River Basins,* 2011, ISBN 978-3-933761-98-9

195 Hrsg.: Koschitzky, Hans-Peter; Braun, Jürgen: *VEGAS-Kolloquium 2010 In-situ-Sanierung - Stand und Entwicklung Nano und ISCO -,* Tagungsband zur Veranstaltung am 07. Oktober 2010 an der Universität Stuttgart, Campus Stuttgart-Vaihingen*,* 2010*,* ISBN 978-3-933761-99-6

196 Gafurov, Abror: *Water Balance Modeling Using Remote Sensing Information - Focus on Central Asia,* 2010, ISBN 978-3-942036-00-9

197 Mackenberg, Sylvia: *Die Quellstärke in der Sickerwasserprognose: Möglichkeiten und Grenzen von Labor- und Freilanduntersuchungen,* 2010, ISBN 978-3-942036-01-6

198 Singh, Shailesh Kumar: *Robust Parameter Estimation in Gauged and Ungauged Basins,* 2010, ISBN 978-3-942036-02-3

199 Doğan, Mehmet Onur: *Coupling of porous media flow with pipe flow,* 2011, ISBN 978-3-942036-03-0

200 Liu, Min: *Study of Topographic Effects on Hydrological Patterns and the Implication on Hydrological Modeling and Data Interpolation,* 2011, ISBN 978-3-942036-04-7

201 Geleta, Habtamu Itefa: *Watershed Sediment Yield Modeling for Data Scarce Areas,* 2011, ISBN 978-3-942036-05-4

202 Franke, Jörg: *Einfluss der Überwachung auf die Versagenswahrscheinlichkeit von Staustufen,* 2011, ISBN 978-3-942036-06-1

203 Bakimchandra, Oinam: *Integrated Fuzzy-GIS approach for assessing regional soil erosion risks,* 2011, ISBN 978-3-942036-07-8

204 Alam, Muhammad Mahboob: *Statistical Downscaling of Extremes of Precipitation in Mesoscale Catchments from Different RCMs and Their Effects on Local Hydrology,* 2011, ISBN 978-3-942036-08-5

205 Hrsg.: Koschitzky, Hans-Peter; Braun, Jürgen: *VEGAS-Kolloquium 2011 Flache Geothermie - Perspektiven und Risiken,* Tagungsband zur Veranstaltung am 06. Oktober 2011 an der Universität Stuttgart, Campus Stuttgart-Vaihingen*,* 2011*,* ISBN 978-3-933761-09-2

206 Haslauer, Claus: *Analysis of Real-World Spatial Dependence of Subsurface Hydraulic Properties Using Copulas with a Focus on Solute Transport Behaviour,* 2011, ISBN 978-3-942036-10-8

207 Dung, Nguyen Viet: *Multi-objective automatic calibration of hydrodynamic models – development of the concept and an application in the Mekong Delta,* 2011, ISBN 978-3-942036-11-5

208 Hung, Nguyen Nghia: *Sediment dynamics in the floodplain of the Mekong Delta, Vietnam,* 2011, ISBN 978-3-942036-12-2

209 Kuhlmann, Anna: *Influence of soil structure and root water uptake on flow in the unsaturated zone,* 2012, ISBN 978-3-942036-13-9

210 Tuhtan, Jeffrey Andrew: *Including the Second Law Inequality in Aquatic Ecodynamics: A Modeling Approach for Alpine Rivers Impacted by Hydropeaking,* 2012, ISBN 978-3-942036-14-6

211 Tolossa, Habtamu: *Sediment Transport Computation Using a Data-Driven Adaptive Neuro-Fuzzy Modelling Approach,* 2012, ISBN 978-3-942036-15-3

212 Tatomir, Alexandru-Bodgan: *From Discrete to Continuum Concepts of Flow in Fractured Porous Media,* 2012, ISBN 978-3-942036-16-0

213 Erbertseder, Karin: *A Multi-Scale Model for Describing Cancer-Therapeutic Transport in the Human Lung,* 2012, ISBN 978-3-942036-17-7

214 Noack, Markus: *Modelling Approach for Interstitial Sediment Dynamics and Reproduction of Gravel Spawning Fish,* 2012, ISBN 978-3-942036-18-4

215 De Boer, Cjestmir Volkert: *Transport of Nano Sized Zero Valent Iron Colloids during Injection into the Subsurface*, 2012, ISBN 978-3-942036-19-1

216 Pfaff, Thomas: *Processing and Analysis of Weather Radar Data for Use in Hydrology*, 2013, ISBN 978-3-942036-20-7

217 Lebrenz, Hans-Henning: *Addressing the Input Uncertainty for Hydrological Modeling by a New Geostatistical Method*, 2013, ISBN 978-3-942036-21-4

218 Darcis, Melanie Yvonne: *Coupling Models of Different Complexity for the Simulation of CO_2 Storage in Deep Saline Aquifers*, 2013, ISBN 978-3-942036-22-1

219 Beck, Ferdinand: *Generation of Spatially Correlated Synthetic Rainfall Time Series in High Temporal Resolution - A Data Driven Approach*, 2013, ISBN 978-3-942036-23-8

220 Guthke, Philipp: *Non-multi-Gaussian spatial structures: Process-driven natural genesis, manifestation, modeling approaches, and influences on dependent processes*, 2013, ISBN 978-3-942036-24-5

221 Walter, Lena: *Uncertainty studies and risk assessment for CO_2 storage in geological formations*, 2013, ISBN 978-3-942036-25-2

222 Wolff, Markus: *Multi-scale modeling of two-phase flow in porous media including capillary pressure effects,* 2013, ISBN 978-3-942036-26-9

223 Mosthaf, Klaus Roland: *Modeling and analysis of coupled porous-medium and free flow with application to evaporation processes,* 2014, ISBN 978-3-942036-27-6

224 Leube, Philipp Christoph: *Methods for Physically-Based Model Reduction in Time: Analysis, Comparison of Methods and Application*, 2013, ISBN 978-3-942036-28-3

225 Rodríguez Fernández, Jhan Ignacio: *High Order Interactions among environmental variables: Diagnostics and initial steps towards modeling*, 2013, ISBN 978-3-942036-29-0

226 Eder, Maria Magdalena: *Climate Sensitivity of a Large Lake*, 2013, ISBN 978-3-942036-30-6

227 Greiner, Philipp: *Alkoholinjektion zur In-situ-Sanierung von CKW Schadensherden in Grundwasserleitern: Charakterisierung der relevanten Prozesse auf unterschiedlichen Skalen*, 2014, ISBN 978-3-942036-31-3

228 Lauser, Andreas: *Theory and Numerical Applications of Compositional Multi-Phase Flow in Porous Media*, 2014, ISBN 978-3-942036-32-0

229 Enzenhöfer, Rainer: *Risk Quantification and Management in Water Production and Supply Systems*, 2014, ISBN 978-3-942036-33-7

230 Faigle, Benjamin: *Adaptive modelling of compositional multi-phase flow with capillary pressure*, 2014, ISBN 978-3-942036-34-4

231 Oladyshkin, Sergey: *Efficient modeling of environmental systems in the face of complexity and uncertainty*, 2014, ISBN 978-3-942036-35-1

232 Sugimoto, Takayuki: *Copula based Stochastic Analysis of Discharge Time Series*, 2014, ISBN 978-3-942036-36-8

233 Koch, Jonas: *Simulation, Identification and Characterization of Contaminant Source Architectures in the Subsurface*, 2014, ISBN 978-3-942036-37-5

234 Zhang, Jin: *Investigations on Urban River Regulation and Ecological Rehabilitation Measures, Case of Shenzhen in China*, 2014, ISBN 978-3-942036-38-2

235 Siebel, Rüdiger: *Experimentelle Untersuchungen zur hydrodynamischen Belastung und Standsicherheit von Deckwerken an überströmbaren Erddämmen,* 2014, ISBN 978-3-942036-39-9

236 Baber, Katherina: *Coupling free flow and flow in porous media in biological and technical applications: From a simple to a complex interface description,* 2014, ISBN 978-3-942036-40-5

237 Nuske, Klaus Philipp: *Beyond Local Equilibrium — Relaxing local equilibrium assumptions in multiphase flow in porous media,* 2014, ISBN 978-3-942036-41-2

238 Geiges, Andreas: *Efficient concepts for optimal experimental design in nonlinear environmental systems*, 2014, ISBN 978-3-942036-42-9

239 Schwenck, Nicolas: *An XFEM-Based Model for Fluid Flow in Fractured Porous Media*, 2014, ISBN 978-3-942036-43-6

240 Chamorro Chávez, Alejandro: *Stochastic and hydrological modelling for climate change prediction in the Lima region, Peru*, 2015, ISBN 978-3-942036-44-3

241 Yulizar: *Investigation of Changes in Hydro-Meteorological Time Series Using a Depth-Based Approach*, 2015, ISBN 978-3-942036-45-0

242 Kretschmer, Nicole: *Impacts of the existing water allocation scheme on the Limarí watershed – Chile, an integrative approach*, 2015, ISBN 978-3-942036-46-7

243 Kramer, Matthias: *Luftbedarf von Freistrahlturbinen im Gegendruckbetrieb*, 2015, ISBN 978-3-942036-47-4

244 Hommel, Johannes: *Modeling biogeochemical and mass transport processes in the subsurface: Investigation of microbially induced calcite precipitation*, 2016, ISBN 978-3-942036-48-1

245 Germer, Kai: *Wasserinfiltration in die ungesättigte Zone eines makroporösen Hanges und deren Einfluss auf die Hangstabilität*, 2016, ISBN 978-3-942036-49-8

246 Hörning, Sebastian: *Process-oriented modeling of spatial random fields using copulas*, 2016, ISBN 978-3-942036-50-4

247 Jambhekar, Vishal: *Numerical modeling and analysis of evaporative salinization in a coupled free-flow porous-media system,* 2016, ISBN 978-3-942036-51-1

248 Huang, Yingchun: *Study on the spatial and temporal transferability of conceptual hydrological models*, 2016, ISBN 978-3-942036-52-8

249 Kleinknecht, Simon Matthias: *Migration and retention of a heavy NAPL vapor and remediation of the unsaturated zone*, 2016, ISBN 978-3-942036-53-5

250 Kwakye, Stephen Oppong: *Study on the effects of climate change on the hydrology of the West African sub-region*, 2016, ISBN 978-3-942036-54-2

251 Kissinger, Alexander: *Basin-Scale Site Screening and Investigation of Possible Impacts of CO_2 Storage on Subsurface Hydrosystems*, 2016, ISBN 978-3-942036-55-9

252 Müller, Thomas: *Generation of a Realistic Temporal Structure of Synthetic Precipitation Time Series for Sewer Applications*, 2017, ISBN 978-3-942036-56-6

253 Grüninger, Christoph: *Numerical Coupling of Navier-Stokes and Darcy Flow for Soil-Water Evaporation*, 2017, ISBN 978-3-942036-57-3

254 Suroso: *Asymmetric Dependence Based Spatial Copula Models: Empirical Investigations and Consequences on Precipitation Fields*, 2017, ISBN 978-3-942036-58-0

255 Müller, Thomas; Mosthaf, Tobias; Gunzenhauser, Sarah; Seidel, Jochen; Bárdossy, András: *Grundlagenbericht Niederschlags-Simulator (NiedSim3)*, 2017, ISBN 978-3-942036-59-7

256 Mosthaf, Tobias: *New Concepts for Regionalizing Temporal Distributions of Precipitation and for its Application in Spatial Rainfall Simulation,* 2017, ISBN 978-3-942036-60-3

257 Fenrich, Eva Katrin: *Entwicklung eines ökologisch-ökonomischen Vernetzungsmodells für Wasserkraftanlagen und Mehrzweckspeicher,* 2018, ISBN 978-3-942036-61-0

258 Schmidt, Holger: *Microbial stabilization of lotic fine sediments*, 2018, ISBN 978-3-942036-62-7

259 Fetzer, Thomas: *Coupled Free and Porous-Medium Flow Processes Affected by Turbulence and Roughness–Models, Concepts and Analysis*, 2018, ISBN 978-3-942036-63-4

260 Schröder, Hans Christoph: *Large-scale High Head Pico Hydropower Potential Assessment*, 2018, ISBN 978-3-942036-64-1

261 Bode, Felix: *Early-Warning Monitoring Systems for Improved Drinking Water Resource Protection*, 2018, ISBN 978-3-942036-65-8

262 Gebler, Tobias: *Statistische Auswertung von simulierten Talsperrenüberwachungsdaten zur Identifikation von Schadensprozessen an Gewichtsstaumauern*, 2018, ISBN 978-3-942036-66-5

263 Harten, Matthias von: *Analyse des Zuppinger-Wasserrades – Hydraulische Optimierungen unter Berücksichtigung ökologischer Aspekte*, 2018, ISBN 978-3-942036-67-2

264 Yan, Jieru: *Nonlinear estimation of short time precipitation using weather radar and surface observations*, 2018, ISBN 978-3-942036-68-9

265 Beck, Martin: *Conceptual approaches for the analysis of coupled hydraulic and geomechanical processes*, 2019, ISBN 978-3-942036-69-6

266 Haas, Jannik: *Optimal planning of hydropower and energy storage technologies for fully renewable power systems*, 2019, ISBN 978-3-942036-70-2

267 Schneider, Martin: *Nonlinear Finite Volume Schemes for Complex Flow Processes and Challenging Grids*, 2019, ISBN 978-3-942036-71-9

268 Most, Sebastian Christopher: *Analysis and Simulation of Anomalous Transport in Porous Media*, 2019, ISBN 978-3-942036-72-6

269 Buchta, Rocco: *Entwicklung eines Ziel- und Bewertungssystems zur Schaffung nachhaltiger naturnaher Strukturen in großen sandgeprägten Flüssen des norddeutschen Tieflandes*, 2019, ISBN 978-3-942036-73-3

270 Thom, Moritz: *Towards a Better Understanding of the Biostabilization Mechanisms of Sediment Beds*, 2019, ISBN 978-3-942036-74-0

271 Stolz, Daniel: *Die Nullspannungstemperatur in Gewichtsstaumauern unter Berücksichtigung der Festigkeitsentwicklung des Betons*, 2019, ISBN 978-3-942036-75-7

272 Rodriguez Pretelin, Abelardo: *Integrating transient flow conditions into groundwater well protection*, 2020, ISBN: 978-3-942036-76-4

273 Weishaupt, Kilian: *Model Concepts for Coupling Free Flow with Porous Medium Flow at the Pore-Network Scale: From Single-Phase Flow to Compositional Non-Isothermal Two-Phase Flow*, 2020, ISBN: 978-3-942036-77-1

274 Koch, Timo: *Mixed-dimension models for flow and transport processes in porous media with embedded tubular network systems*, 2020, ISBN: 978-3-942036-78-8

275 Gläser, Dennis: *Discrete fracture modeling of multi-phase flow and deformation in fractured poroelastic media*, 2020, ISBN: 978-3-942036-79-5

276 Seitz, Lydia: *Development of new methods to apply a multi-parameter approach – A first step towards the determination of colmation*, 2020, ISBN: 978-3-942036-80-1

277 Ebrahim Bakhshipour, Amin: *Optimizing hybrid decentralized systems for sustainable urban drainage infrastructures planning*, 2021, ISBN: 978-3-942036-81-8

278 Seitz, Gabriele: *Modeling Fixed-Bed Reactors for Thermochemical Heat Storage with the Reaction System CaO/Ca(OH)$_2$*, 2021, ISBN: 978-3-942036-82-5

279 Emmert, Simon: *Developing and Calibrating a Numerical Model for Microbially Enhanced Coal-Bed Methane Production*, 2021, ISBN: 978-3-942036-83-2

280 Heck, Katharina Klara: Modelling and analysis of multicomponent transport at the interface between free- and porous-medium flow - influenced by radiation and roughness, 2021, ISBN: 978-3-942036-84-9

281 Ackermann, Sina: *A multi-scale approach for drop/porous-medium interaction*, 2021, ISBN: 978-3-942036-85-6

282 Beckers, Felix: *Investigations on Functional Relationships between Cohesive Sediment Erosion and Sediment Characteristics*, 2021, ISBN: 978-3-942036-86-3

283 Schlabing, Dirk: *Generating Weather for Climate Impact Assessment on Lakes*, 2021, ISBN: 978-3-942036-87-0

284 Becker, Beatrix: *Efficient multiscale multiphysics models accounting for reversible flow at various subsurface energy storage sites*, 2021, ISBN: 978-3-942036-88-7

285 Reuschen, Sebastian: *Bayesian Inversion and Model Selection of Heterogeneities in Geo-statistical Subsurface Modeling*, 2021, ISBN: 978-3-942036-89-4

286 Michalkowski, Cynthia: *Modeling water transport at the interface between porous GDL and gas distributor of a PEM fuel cell cathode*, 2022, ISBN: 978-3-942036-90-0

287 Koca, Kaan: *Advanced experimental methods for investigating flow-biofilm-sediment interactions*, 2022, ISBN: 978-3-942036-91-7

288 Modiri, Ehsan: *Clustering simultaneous occurrences of extreme floods in the Neckar catchment*, 2022, ISBN: 978-3-942036-92-4

289 Mayar, Mohammad Assem: *High-resolution spatio-temporal measurements of the colmation phenomenon under laboratory conditions*, 2022, ISBN: 978-3-942036-93-1

290 Schäfer Rodrigues Silva, Aline: *Quantifying and Visualizing Model Similarities for Multi-Model Methods*, 2022, ISBN: 978-3-942036-94-8

291 Moreno Leiva, Simón: *Optimal planning of water and renewable energy systems for copper production processes with sector coupling and demand flexibility*, 2022, ISBN 978-3-942036-95-5

292 Schönau, Steffen: *Modellierung von Bodenerosion und Sedimentaustrag bei Hochwasserereignissen am Beispiel des Einzugsgebiets der Rems*, 2022, ISBN 978-3-942036-96-2

293 Glatz, Kumiko: *Upscaling of Nanoparticle Transport in Porous Media,* 2022, ISBN 978-3-942036-97-9

294 Pavía Santolamazza, Daniela: *Event-based flood estimation using a random forest algorithm for the regionalization in small catchments*, 2022, ISBN 978-3-942036-98-6

295 Haun, Stefan: *Advanced Methods for a Sustainable Sediment Management of Reservoirs*, 2022, ISBN 978-3-942036-99-3

296 Herma, Felix: *Data Processing and Model Choice for Flood Prediction*, 2022, ISBN 978-3-910293-00-7

297 Weinhardt, Felix: *Porosity and permeability alterations in processes of biomineralization in porous media - microfluidic investigations and their interpretation*, 2022, ISBN 978-3-910293-01-4

298 Sadid, Najibullah: *Bedload Transport Estimation in Mountainous Intermittent Rivers and Streams*, 2023, ISBN 978-3-910293-02-1

299 Mohammadi, Farid: *A Surrogate-Assisted Bayesian Framework for Uncertainty-Aware Validation Benchmarks*, 2023, ISBN 978-3-910293-03-8

300 Praditia, Timothy: *Physics-informed Neural Networks for Learning Dynamic, Distributed and Uncertain Systems*, 2023, ISBN 978-3-910293-04-5

301 Gyawali, Dhiraj Raj: *Development and parameter estimation of conceptual snow-melt models using MODIS snow-cover distribution*, 2023, ISBN 978-3-910293-05-2

302 Görtz, Jan: *Coupled modeling approach for physico-chemical processes during the deterioration of cement-based structures*, 2023, ISBN 978-3-910293-06-9

303 Veyskarami, Maziar: *Coupled free-flow–porous media flow processes including drop formation*, 2023, ISBN 978-3-910293-07-6

304 El Hachem, Abbas: *Spatial Extent of Precipitation Extremes in Hydrology*, 2023, ISBN 978-3-910293-08-3

305 Banerjee, Ishani: *Stochastic Model Comparison and Refinement Strategies for Gas Migration in the Subsurface*, 2023, ISBN 978-3-910293-09-0

306 Anwar, Faizan: *Spatial aspects of hydrological extremes: Description and simulation*, 2024, ISBN 978-3-910293-10-6

Die Mitteilungshefte ab der Nr. 134 (Jg. 2005) stehen als pdf-Datei über die Homepage des Instituts: www.iws.uni-stuttgart.de zur Verfügung.