HENRY
Hydraulic Engineering Repository

Ein Service der Bundesanstalt für Wasserbau

Doctoral Thesis, Periodical Part, Published Version

Schlabing, Dirk

Generating weather for climate impact assessment on lakes

Mitteilungen. Institut für Wasser- und Umweltsystemmodellierung, Universität Stuttgart

Zur Verfügung gestellt in Kooperation mit/Provided in Cooperation with:
Universität Stuttgart

Verfügbar unter/Available at: https://hdl.handle.net/20.500.11970/108873

Vorgeschlagene Zitierweise/Suggested citation:
Schlabing, Dirk (2021): Generating weather for climate impact assessment on lakes. Stuttgart: Universität Stuttgart, Institut für Wasser- und Umweltsystemmodellierung (Mitteilungen. Institut für Wasser- und Umweltsystemmodellierung, Universität Stuttgart, 282). http://dx.doi.org/10.18419/opus-12051.

Standardnutzungsbedingungen/Terms of Use:

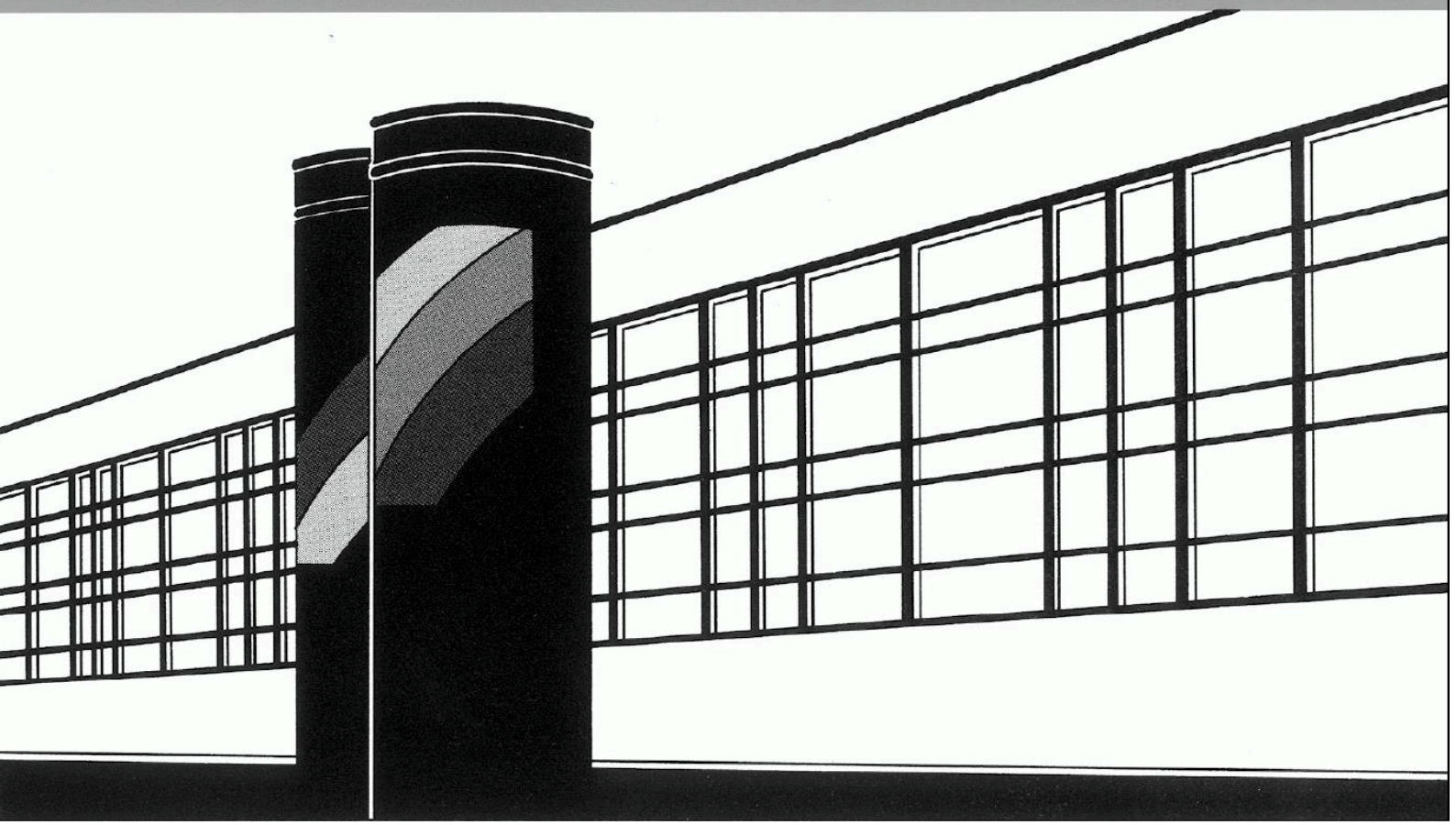
Die Dokumente in HENRY stehen unter der Creative Commons Lizenz CC BY 4.0, sofern keine abweichenden Nutzungsbedingungen getroffen wurden. Damit ist sowohl die kommerzielle Nutzung als auch das Teilen, die Weiterbearbeitung und Speicherung erlaubt. Das Verwenden und das Bearbeiten stehen unter der Bedingung der Namensnennung. Im Einzelfall kann eine restriktivere Lizenz gelten; dann gelten abweichend von den obigen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Documents in HENRY are made available under the Creative Commons License CC BY 4.0, if no other license is applicable. Under CC BY 4.0 commercial use and sharing, remixing, transforming, and building upon the material of the work is permitted. In some cases a different, more restrictive license may apply; if applicable the terms of the restrictive license will be binding.

Universität Stuttgart

Institut für Wasser- und Umweltsystemmodellierung

Mitteilungen



Heft 283 Dirk Schlabing

Generating Weather for Climate Impact Assessment on Lakes

Generating Weather for Climate Impact Assessment on Lakes

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

vorgelegt von

Dirk Schlabing

aus Darmstadt, Deutschland

Hauptberichter:	Prof. Dr.-rer. nat. Dr.-Ing. András Bárdossy
Mitberichter:	Univ.-Prof. Dr. Henning Rust

Tag der mündlichen Prüfung: 26.03.2021

Institut für Wasser- und Umweltsystemmodellierung
der Universität Stuttgart
2021

Heft 283

Generating Weather for Climate Impact Assessment on Lakes

von
Dr.-Ing.
Dirk Schlabing

Eigenverlag des Instituts für Wasser- und Umweltsystemmodellierung
der Universität Stuttgart

D93 Generating Weather for Climate Impact Assessment on Lakes

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

Schlabing, Dirk:
Generating Weather for Climate Impact Assessment on Lakes, Universität Stuttgart. - Stuttgart: Institut für Wasser- und Umweltsystemmodellierung, 2021

(Mitteilungen Institut für Wasser- und Umweltsystemmodellierung, Universität Stuttgart: H. 283)
Zugl.: Stuttgart, Univ., Diss., 2021
ISBN 978-3-942036-87-0

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

Gegen Vervielfältigung und Übersetzung bestehen keine Einwände, es wird lediglich um Quellenangabe gebeten.

Herausgegeben 2021 vom Eigenverlag des Instituts für Wasser- und Umweltsystem-modellierung

Druck: DCC Kästl e.K., Ostfildern

Danksagungen

Ich danke:

Prof. András Bárdossy für das mir entgegengebrachte Vertrauen und vorsichtig aufgebrachten Druck. Eine Arbeitsgruppe mit einer angenehmen und fehlerbereinigenden Diskussionskultur in der ständig neue Ideen frei zirkulieren und verfeinert werden, kommt nicht aus dem Nichts. Und für Pörkölt!

Prof. Henning Rust für aufmerksames Lesen des Manuskriptes – insbesondere der Formeln! Außerdem für den anregenden Wettergeneratoren-Workshop 2017 in Berlin.

Prof. Christian Moormann für das Einspringen als Prüfungsvorsitzender in einer ungewöhnlichen Zeit und seine Geduld bei der technischen Umsetzung der Prüfung.

Jochen Seidel für das Kümmern darum, dass das soziale Miteinander in der Abteilung nicht zu kurz kommt und Bürokratienavigation für mich Verwaltungsorientierungslosen.

Astrid Lemp Jochens Mitstreiterin/Machete im Formular-Djungel, die mich immer wieder beruhigen konnte, wenn ich gerade wieder Angst hatte etwas vergessen oder alles falsch ausgefüllt zu haben. Und für vielfaches Korrekturlesen – mit fachlichem Interesse! Ohne Astrid hätten meine Chilipflanzen im Büro die Pandemie nicht überlebt.

Magdalena Eder und Marieke Frassl für die intensive und freundschaftliche Zusammenarbeit, die aus den Wettergeneratoren das gemacht hat was sie sind. Ein weiterer Dank geht an Magdalenas Onkel, auf dessen Bio-Bauernhof wir drei in Klausur gegangen sind.

Meinen "alten" Kollegen Ferdinand Beck, Thomas Pfaff und Claus Haslauer von denen ich unter anderem vieles über Statistik gelernt habe. Wenn etwas falsch war, haben sie darauf hingewiesen und das war richtig und hilfreich. Philipp Guthke ebenso und außerdem für das Verwirklichen von schrägen Ideen[1]. Takayuki Sugomoto für interessante wie leckere Menüs. Felix Herma dafür dass er mir den Vortritt beim Abgeben gelassen hat.

Meinen "mittelalten" Kollegen Thomas Müller für Brettspieldiskussionen und die Zusammenarbeit an dem Code der Wetterstationsseite. Tobias Mosthaf für Beistand bei der Betreuung fachfremder Doktoranden. Sebastian Hörning für Random Mixing Code. Micha Eisele für gemeinsames Windmessen auf der Kormoran.

Meinen "jungen" Kollegen Faizan Anwar, dem Kondensationskeim für angeregte Diskussionen. Meinen Bürokollegen Dhiraj Gyawali und Ehsan Modiri für oft kulinarischen Kulturdialog. Abbas El Hachem, Claudia Teutsch, Masoud Mehrvand für die Kafferunden- und Mittagspausen-Gespräche.

[1] Und eine Bitte um Entschuldigung: Hätte ich nicht gebremst, wären wir heute Bitcoin-Millionäre.

Sarah Hemer für ihre Liebe, notwendige moralische Unterstützung und dem Beharren darauf, dass ich es verständlich mache, wenn ich ihr enthusiastisch von multivariaten Abhängigkeiten, Zirkulation in Seen oder Metaprogrammierung erzähle. Eigentlich braucht jeder Doktorand so jemanden.

Meinen Eltern und meinem Bruder

Contents

List of Figures

List of Tables

Glossary

AR: Auto-regressive. 6

AWE-GEN-2d: **A**dvanced **WE**ather **GEN**erator for a two-dimensional grid. 4

CAEDYM: Computational Aquatic Ecosystem DYnamics Model. 100, 102, 104, 107, VIII

canonical vine: An R-Vine containing a central node (maximum possible degree) in each tree.. 11, 55, IV canonical vine. 11, 14–17, 55, 56, 59, 60, IV.

cdf: cumulative distribution function. xv, 12, 29, 32, 33, 35, 37, 46, 56, 59, 61, 82, 83, 91, 92, 94, 96, IV, VII–IX

connected graph: A graph in which there is a path between every pair of nodes. 10

degree: A degree of a node in a graph is the number of edges attached to it. 11

doy: day of year. The day number of the year (e.g. January 3rd 1981 = 3, December 16th 2013 = 350). 29–31, 35, 37, 44, 46, 60, 63, 64, 75, 77, 94, 101, VI

drawable vine: An R-Vine with a maximum degree of 2 in the first tree.. 11, 59, IV drawable vine. 11, 13–15, 59, IV.

DWD: German Meteorological Service. 27, 62, 100

DYRESM: DYnamic REservoir Simulation Model. 100, 104, 106, 107, VIII

ELCOM: Estuary, Lake and Coastal Ocean Model. 102, 106

FFT: Fast Fourier Transform. 24, 30, 60, 100, 103

FT: Fourier Transform. 21–24

GCM: general circulation model. 1, 4, 45, 97

GLM: General Lake Model. 107, 110, VIII

GOTM: General Ocean Turbulence Model. 107, 109

impact model: A model used to estimate climate impact. Also sometimes called climate impact model. In this work those are Estuary, Lake and Coastal Ocean Model (ELCOM) and DYRESM. 4, 99

kernel density estimate: A non-parametric probability density function (*pdf*) estimation method that replaces each data point in a sample with a kernel.. kernel density estimation. 18, 27, 29–31, 49, 62, 63.

KLL: Kinneret Limnological Lab. 107

KNN: K-Nearest Neighbors. xvii, 3–5, 18, 19, 74, 75, 77–79, 86, 89, 91, 93, 97, 98, II, VI, VII, X, XIV, XVIII

MSB: Mediterranean Sea Breeze. 106

NHMM: Non-Homogeneous Hidden Markov Model. 3

pdf: probability density function. 12

RCM: regional climate model. 45, 48

RCP: Representative Concentration Pathway. Scenarios for the evolution of greenhouse gas trajectories as defined in the IPCC in its fifth Assessment Report. 101

tree: A connected graph with n nodes and $n - 1$ edges. 10

VAR: Vector-Autoregressive. xvii, 5–9, 25–27, 29, 31, 32, 34, 38, 40–42, 45, 47, 48, 53–55, 74, 75, 83, 85, 86, 88–94, 98, 103, V, VII, VIII, X

VG: Vector-Autoregressive Weather Generator. xvii, 1, 2, 5, 25–29, 32, 34, 37, 38, 45, 49, 53, 55, 57, 73–75, 80, 83, 85, 86, 88, 89, 91, 92, 94, 99–102, 106–108, 110, 111, I, IV, V, VII, IX, X, XIII, XIV, XVI, XVIII

WeatherCop: **Weather** Generator based on Phase Randomization and Vine **Cop**ulas. xvii, 2, 5, 15, 55, 57–62, 66, 68, 73–75, 80, 83, 85, 88, 89, 91–93, 99, 112, V, VII, X, XIV, XVII, XVIII

WG: Weather Generator. 1–4, 6, 15, 25, 26, 28, 29, 34, 53–55, 57, 73–75, 80, 99, 110, 112, 113, IV, V

WGEN: Richardson's "Weather Generator". 3, 4

Abstract

Lakes are driven in part by weather and they are affected by climate. Short-term events can initiate mixture, but its reach into the deep is bounded by the vertical thermal distribution, which reflects the longer-term meteorological situation. Intertwined in the physical processes are biological ones, like phytoplankton communities that produce turbidity-altering blooms under the right conditions of light availability, water temperature and stable stratification.

In other words, lakes are complex ecosystems and their reactions to changed climatic conditions can hardly be determined in a trivial manner.

Hydrodynamic and combined ecological lake models reflect the complexity of natural lakes to a certain extent and make it possible to conduct experiments with hypothetical input following envisioned climatic changes. The research of lakes with such models can be enriched by recognizing the influence of the variability within the lake's boundary conditions. Building stochastic abstractions of weather, so called weather generators, allows to repeatedly run the same experiment, changing only what is deemed random from run to run and distilling an aggregate lake response complemented with bounds resulting from the modelled input uncertainty.

The tools presented in this work are the culmination of years of cooperation with lake modellers. The intent of the limnologists was not to be supplied with best-guess values according to a greenhouse gas emission scenario, as done when downscaling output of climate models, but to have a tool, that allows them to define climate scenarios in accordance to a given increase in air temperature. Also, changes in variability should be configurable and defined in terms of air temperature. Air temperature is not the only variable having an influence on a lake and is correlated with other meteorological variables as sunshine, precipitation, relative humidity and more. In order to keep such freely defined temperature-based scenarios plausible, existing weather generation methods had to be adapted. In order to propagate changes in air temperature to the rest of the generated meteorological variables, the statistical dependencies in measured data using linear and non-linear models were exploited.

The weather generator **Vector-Autoregressive Weather Generator (VG)** relies on such linear dependencies and generates daily, single-site time series. It has been used in a few published studies, meeting the above requirements. Its central model is a vector-autoregressive process, a simple extension of the autoregressive process to the multivariate case. In order to improve long-term variability, it was combined with phase randomization, a method that uses the Fourier transform to generate "surrogate data". As vector-autoregressive models generate time series with normal marginals, transformations of the input and output are

used to traverse from measurement to standard-normal domain and back after simulation. To propagate changes in one guiding variable to the rest of the generated variables, the properties of the parameters of the vector-autoregressive model were exploited and complemented with linear regression.

The multi-site, daily **Weather Generator based on Phase Randomization and Vine Copulas (WeatherCop)** uses copulas to model possibly non-linear dependencies between the modelled variables in source data. In order to reduce the number of parameters, WeatherCop uses phase randomization as well. Here it maintains the inter-site and temporal dependencies. Its use was predicated on formulating an inverse sampling algorithm that is used to decorrelate observations. Copulas are models of multivariate dependence with uniform marginals, so transformations of the input are used for this weather generator as well. As copulas are models of dependence, they lend themselves ideally for propagating changes in one guiding variable to other variables. The type of copula approach chosen here is the pair construction method, commonly referred to as Vine copulas. They are flexible tools as they allow building a multivariate dependence structure from pairs of variables, using the vast amount of available bivariate copula families.

With these two parametric weather generators, a **novel way to generate precipitation** was explored. Both weather generators are multivariate models and as such have more variables available than just precipitation. The dependency between precipitation and other variables is usually thought of as implying a direction of causation in which the presence of rain determines the values of the other simulated variables. This direction of dependency was reversed in order to extract information about a dryness probability from the non-precipitation variables. Properly mapped, this dryness probability can fill the dry gaps in the transformed input time series for each of the weather generators. An upside to this treatment of precipitation is that it does not require a precipitation occurrence model and no different parameterizations for wet and dry states for the non-precipitation variables.

Testing the weather generators was done in several ways. First, by **direct validation** of statistical properties in observed and generated time series. The second, harder **extrapolation test** for their applicability in climate change scenarios was on how well they could project a change in the guiding variable temperature to the rest of the generated variables. This was done by calibrating the models on a cold period and driving them with the temperature difference to a warmer period in the measured data. The ability to extrapolate lies then not primarily in how well the guiding variable is agreeing with measurements, but in the other generated variables that are supposed to change accordingly. Extrapolation was more accurate using the more complex WeatherCop than VG. Still, results of the extrapolation tests show, that only relying on statistical relationships can mis-project the changes in dependent variables that accompany temperature increases. The last kind of test was only done for VG, namely an **indirect validation**, using multiple lake models in cooperation with lake modellers. Indirect validation tests whether the relevant properties of a measured time series is still present in generated time series by comparing the output of the lake model run with generated and measured data.

The two parametric weather generators VG and WeatherCop are contrasted with a **non-parametric K-Nearest Neighbors (KNN) resampler**, highlighting differences between those

approaches and underlining specific weaknesses. Resampling existing records is deemed unsuitable as it lacks abstraction and with it variability, but the parametric methods simplify dependencies and overestimate spread. The resampling method proved to have lower variance dependent on the autocorrelation of the source data, which was underlined with an experiment where time series of varying autocorrelation were generated. A general tendency towards central values was observed in both a uni- and multivariate sense and entropy comparisons showed a higher than measured order in resampled and too low order in time series generated with the studied parametric models.

Zusammenfassung

Seen werden zum Teil durch das Wetter angetrieben und sie werden über das Klima beeinflusst. Kurzfristige Ereignisse können eine Vermischung der Wasserschichten auslösen, aber ihre Reichweite in die Tiefe wird durch die vertikale Wärmeverteilung begrenzt, die die längerfristige meteorologische Situation widerspiegelt. In die physikalischen Prozesse verflochten sind biologische Prozesse, wie Phytoplankton-Gemeinschaften, die bei passenden Bedingungen, wie ausreichender Lichtverfügbarkeit, Wassertemperatur und stabiler Schichtung, trübungsverändernde Blüten hervorbringen.

Mit anderen Worten: Seen sind komplexe Ökosysteme und ihre Reaktionen auf veränderte klimatische Bedingungen lassen sich kaum in einer trivialen Art und Weise abschätzen.

Hydrodynamische und kombinierte ökologische Seenmodelle spiegeln die Komplexität der natürlichen Seen bis zu einem gewissen Grad wider und machen es möglich Experimente mit hypothetischen Eingaben durchzuführen, die angenommenen klimatischen Veränderungen folgen. Die Erforschung von Seen mit Hilfe solcher Modelle kann durch das Anerkennen des Einflusses der Variabilität der Randbedingungen des Sees bereichert werden. Das Erstellen stochastischer Abstraktionen von Wetter, so genannter Wettergeneratoren, erlaubt es dasselbe Experiment wiederholt durchzuführen, wobei von Durchlauf zu Durchlauf nur das geändert wird, was als zufällig angesehen wird. Neben der Destillation einer aggregierten Reaktion des Sees kann das Ergebnis durch Ausgabebereiche, die sich aus der modellierten Eingangsunsicherheit ergeben, ergänzt werden.

Die in dieser Arbeit vorgestellten Werkzeuge sind das Ergebnis jahrelanger Zusammenarbeit mit Seenmodellierern. Die Absicht der Limnologen war es nicht mit Werten beliefert zu werden, die eine Abschätzung entsprechend eines Klimagasemissionsszenarios darstellen, was im Rahmen eines Downscalings von Klimamodellausgaben geschieht, sondern ein Werkzeug zu erhalten, das es ihnen erlaubt, Klimaszenarien in Übereinstimmung mit einem gegebenen Anstieg der Lufttemperatur zu definieren. Außerdem sollten Änderungen der Variabilität konfigurierbar sein und in Form von Lufttemperaturänderungen bestimmt werden. Die Lufttemperatur ist nicht die einzige Variable, die einen Einfluss auf einen See besitzt und sie korreliert mit anderen meteorologischen Variablen wie Sonnenschein, Niederschlag, relativer Feuchtigkeit und mehr. Um solche frei definierten temperaturbasierten Szenarien plausibel zu halten, mussten vorhandene Methoden zur Wettergenerierung angepasst werden. Um Veränderungen in der Lufttemperatur auf den Rest der generierten meteorologischen Variablen zu übertragen, wurden die statistischen Abhängigkeiten in den Messdaten unter Verwendung linearer und nicht-lineare Modelle ausgenutzt.

Der Wettergenerator **Vector-Autoregressive Weather Generator (VG)** stützt sich auf solche linearen Abhängigkeiten und generiert tägliche Zeitreihen für einen Ort. Er wurde in ei-

nigen wenigen veröffentlichten Studien verwendet, wobei er den oben genannten Anforderungen entsprach. Sein zentrales Modell ist ein vektor-autoregressiver Prozess, eine einfache Erweiterung des autoregressiven Prozesses für den multivariaten Fall. Um die Variabilität auf längeren Zeitskalen zu verbessern, wurde er mit Phasenrandomisierung kombiniert, einem Verfahren, das die Fourier-Transformation zur Erzeugung von Surrogatdaten verwendet. Da vektor-autoregressive Modelle Zeitreihen mit normalen Randverteilungen erzeugen, werden Transformationen der Eingangs- und Ausgabedaten verwendet, um von der Messdomäne zum Standard-Normalen und zurück zu gelangen. Um Änderungen in einer Leitvariablen an die übrigen generierten Variablen weiterzugeben, wurden Eigenschaften der Parameter des vektor-autoregressiven Models ausgenutzt und mit linearer Regression ergänzt.

Der für mehrere Orte gleichzeitig generierende, tägliche Wettergenerator **Weather Generator based on Phase Randomization and Vine Copulas (WeatherCop)** verwendet Copulas um eventuell nicht-lineare Abhängigkeiten zwischen den Variablen in den Quelldaten zu modellieren. Um die Anzahl an Parametern zu reduzieren, verwendet WeatherCop auch Phasenrandomisierung. Diese wird hier verwendet um die räumliche und zeitliche Abhängigkeit aufrecht zu erhalten. Ihre Verwendung basiert auf der Formulierung des inversen Sampling-Algorithmus, der für die Dekorrelation der Beobachtungen verwendet wird. Copulas sind Modelle multivariater Abhängigkeit mit gleichverteilten Randverteilungen, weswegen auch für diesen Wettergenerator Transformationen der Eingangsdaten verwendet werden. Da Copulas Abhängigkeitsmodelle sind, eignen sie sich ideal für die Weitergabe von Änderungen einer Leitvariable an andere Variablen. Die hier gewählte Art des Copula-Ansatzes ist die Paar-Konstruktionsmethode, die üblicherweise auch als Vine Copula bezeichnet wird. Vines sind flexible Werkzeuge, da sie es erlauben, die multivariate Abhängigkeitsstruktur aus Paaren von Variablen aufzubauen, wobei die große Menge an verfügbaren bivariaten Copulafamilien verwendet wird.

Mit diesen beiden parametrischen Wettergeneratoren wurde ein **neuartiger Weg Niederschlag zu erzeugen** untersucht. Beide Wettergeneratoren sind multivariate Modelle und haben als solche mehr Variablen zur Verfügung als nur Niederschlag. Die Abhängigkeit zwischen Niederschlag und den anderen Variablen wird gewöhnlich so verstanden, dass sie eine Richtung der Kausalität impliziert, in der das Vorhandensein von Regen die Werte der anderen simulierten Variablen beeinflusst. Diese Abhängigkeitsrichtung wurde umgekehrt, um Informationen über eine Trockenheitswahrscheinlichkeit aus den Nichtniederschlagsvariablen zu extrahieren. Richtig abgebildet kann diese Trockenheitswahrscheinlichkeit die Trockenlücken in der transformierten Eingabezeitreihe für beide Wettergeneratoren füllen. Ein Vorteil dieser Behandlung von Niederschlag ist, dass kein Niederschlagsvorkommensmodell und keine unterschiedlichen Parametrisierungen für nasse und trockene Zustände der Nicht-Niederschlagsvariablen benötigt werden.

Das Testen der Wettergeneratoren wurde auf mehrere Weisen durchgeführt. Erstens durch **direkte Validierung** der statistischen Eigenschaften in beobachteten und generierten Zeitreihen. Der zweite, schwierigere **Extrapolationstest** für ihre Anwendbarkeit in Klimawandelszenarien bestand darin, wie gut sie eine Änderung der leitenden Variablen Temperatur auf den Rest der generierten Variablen projizieren konnten. Dies wurde erreicht

indem die Modelle auf eine kalte Periode kalibriert und mit dem Temperaturunterschied zu einer wärmeren Periode in den gemessenen Daten angetrieben wurden. Die Extrapolationsfähigkeit liegt dann nicht primär darin, wie gut die Leitvariable mit den Messungen übereinstimmt, sondern in den anderen generierten Variablen, die sich entsprechend ändern sollen. Die Extrapolationsfähigkeit war mit dem komplexeren WeatherCop besser als mit VG. Die Ergebnisse der Extrapolationstests zeigen jedoch, dass die alleinige Verwendung statistischer Beziehungen die Änderungen der abhängigen Variablen, die mit Temperaturerhöhungen einhergehen, falsch projizieren kann. Die letzte Art von Test wurde nur für VG durchgeführt, nämlich eine **indirekte Validierung** unter Verwendung mehrerer Seenmodelle in Zusammenarbeit mit Seenmodellierern. Die indirekte Validierung testet ob die relevanten Eigenschaften einer gemessenen Zeitreihe in generierten Zeitreihen noch vorhanden sind, indem die Ausgabe des Seenmodelllaufs mit generierten und gemessenen Daten verglichen wird.

Die beiden parametrischen Wettergeneratoren VG und WeatherCop wurden einem **nichtparametrischen K-Nearest Neighbors (KNN)-Resampler** gegenübergestellt, was Unterschiede zwischen diesen Ansätzen aufzeigt und spezifische Schwächen hervorhebt. Das Resampling vorhandener Datensätze wird als ungeeignet angesehen, da es ihm an Abstraktion und damit an Variabilität mangelt. Die parametrischen Methoden vereinfachen jedoch Abhängigkeiten und überschätzen die Streuung. Das Resampling-Verfahren zeigte eine geringere Varianz in Abhängigkeit der Autokorrelation der Quelldaten, was durch ein Experiment unterstrichen wurde, bei dem Zeitreihen unterschiedlicher Autokorrelation erzeugt wurden. Eine allgemeine Tendenz zu zentralen Werten wurde sowohl im uni- als auch im multivariaten Sinne beobachtet, und Entropievergleiche zeigten eine höhere als die gemessene Ordnung in Datenreihen aus Resampling und eine zu niedrige Ordnung in Zeitreihen, die mit den untersuchten parametrischen Modellen erzeugt wurden.

1 Introduction

Weather Generators (WGs) provide stochastic abstractions of meteorological data. With these abstractions, synthetic data in the likeness of the source data can be created. The synthetic data resembles the source data, when compared with the help of descriptive statistics, while being different in "all the details". Use-cases for and implementations of weather generators are as numerous as this concept is broad. WGs can be used to generate data where it is scarce or incomplete. By spatial interpolation of WG parameters, time series can be generated for location where no measurements exist. They have first been used for hydrological applications (Caskey, 1962), but have been applied in other fields as well. The ability to produce time series of indefinite lengths allows setting up experiments to study the occurrence of extreme or otherwise unusual situations. There exist publications of the use of WGs for crop modelling (Mavromatis and Hansen, 2001), modelling terrestrial ecosystems (Kucharik, 2003), modelling water resource systems (Fowler, Kilsby, and O'Connell, 2000) and flood-risk assessment (Te Linde et al., 2010).

Beyond the use of WGs to generate more data where it is of insufficient length are estimations of possible climate impacts. This can be implemented for example by disturbing their parameters or conditioning on a large-scale predictor variable from a general circulation model (GCM). An advantage of using a WG for climate impact assessment is that the inherently unpredictable local short-term variability is modelled stochastically. Hence, the unpredictable is replaced by the plausible. In a stochastic context, the plausible is not a singular thing, but a multitude of outcomes which makes it possible to do so-called Monte Carlo experiments in which a high number of WG output realizations are generated and evaluated.

There is no, and there never will be, "one true" weather generator that is closer to reality than every other one. In contrast to physically-based modelling, there is no equivalent to energy and mass balance equations that can be adhered to. Instead of converging towards a more precise estimation of known physical laws, a stochastic mechanism that is able to produce data with similar statistical properties as the source data is sought. WGs differ in the properties considered important to mimic and the methods used to achieve that similarity.

1.1 Aims of the Weather Generators in this Work

The task of the WGs I designed[1] and which I describe here is mainly to provide input time-series for hydrodynamic and ecological modeling of lakes under changed climatic condi-

[1] My former colleague Magdalena Eder made contributions to the WG Vector-Autoregressive Weather Generator (VG).

tions. The focus here is not to project the most likely climatic change, but to provide a tool to freely define scenarios independent of climate model output. This helps impact modellers to better understand if, how and when their studied system might react to changes. As this work presents new WGs with these applications in mind, a few aims are provided in the following:

- Balanced "what if" climate scenario output in which a change in one guiding variable leads to changes in the other variables by exploiting statistical dependencies in the source data. The type of change should encompass:
  - increasing the mean
  - increasing variability
  - increasing mean and variability simultaneously
  - a trend in mean
  - enhancing specific local weather patterns
- Flexibility in the choice of generated variables. This enables the generation of variables that are needed as input for simulation of lakes as well as adopting further variables for future works in possibly different fields.
- Continuously changing characteristics throughout the year. This should include not only aspects of marginal distributions, but also dependencies. Important possible impacts of climatic change on ecological systems are the changes in the timing of plant and animal life cycles. Thus, I want to avoid sudden regime-shifts in the generated time series.

At the time of writing, these aims together cannot be fulfilled using existing WGs. VG was the first WG I developed to address the above aims. It employs traditional methods from the field of time series analysis. In order to contrast its output with widely used and very different methods, I implemented a resampling approach. Finally, in an effort to innovate in terms of more modern methods, I developed **Weather** Generator based on Phase Randomization and Vine **Cop**ulas (WeatherCop) which employs modern techniques and has a wider spectrum of possible applications.

1.2 Existing Weather Generators

Central to stochastic weather generation is precipitation. Precipitation occurrence has an influence on environmental processes and can thus be seen as "controlling" other meteorological variables (Daniel S. Wilks and R. L. Wilby, 1999). A large number of publications deal mostly with rain generation (Harrold, 2003; Mehrotra, Srikanthan, and Sharma, 2006; Bárdossy and Pegram, 2016).

Since the conception of WGs by C. W. C. Richardson (1981), most WGs follow a three-step process: the generation of rain occurrence, rain amount and non-precipitation variables. Before generating precipitation values, an occurrence model has to decide on a rain state. After

the decision of rain occurrence, the rain value and values of the other simulated variables are generated. Richardson's "Weather Generator" (WGEN) (C. W. Richardson and Wright, 1984) is the first implementation of Richardson's approach. Rain occurrence in Richardson-type WGs is modelled with a two-state Markov-Chain. This means, that, dependent on the previous time-step's rain state and transition probabilities, the model decides randomly on the current time step's wet or dry state. Racsko, Szeidl, and M. Semenov (1991) suggested a serial approach to occurrence modelling in order to improve the frequency of longer dry and wet spells as these are underestimated by short Markov Chains. In Racsko's model, whole wet and dry period lengths are drawn from distributions fitted on period lengths within observed data. The widely used weather generator LARS-WG (M. A. Semenov, 2002) is an implementation that is close to the Racsko approach. Both Racsko's and Richardson's approach use separate parameter sets for wet and dry state for non-precipitation variables. LARS-WG assumes serial independence of precipitation amounts during individual wet spells.

WGs that first generate precipitation and then the rest of the variables are unable to operate in the "balanced scenario"-mode mentioned above if the guiding variable is not precipitation. While it would be easy to change aspects of precipitation and have the other variables reflect the change, the other way around is hard, because of the flow of information goes from precipitation to the rest of the variables.

Another method used in stochastic weather generation is based on the Non-Homogeneous Hidden Markov Model (NHMM) (Hughes, Guttorp, and Charles, 1999). It is "non-homogeneous" as precipitation occurrence is not only modelled dependent on precipitation on previous time steps, but on an additional, external variable. The "hidden" refers to an unobserved weather state that is modelled by a Markov Chain and which determines rainfall occurrences. At the same time, this state provides a link between a large scale atmospheric circulation pattern and local measurements. Similarly to the C. W. Richardson and Wright (1984) and Racsko, Szeidl, and M. Semenov (1991) type of WGs, the NHMM is unable to be used in guided scenarios, as all variables are dependent on the hidden state which enforces the flow of information.

Poisson cluster processes combine simulation of precipitation occurrence and amount (Onof et al., 2000). "Storms" are generated at random times with random duration. These storms are made up of a random number of rectangular pulse "cells" of random duration and intensity which are generated by another Poisson process. For every point in time within a storm, precipitation intensity is obtained by adding the intensities of every cell. Two prominent versions of this method are the Bartlett-Lewis and the Neyman-Scott model, which differ in the way cells are placed in a storm (Rodriguez-Iturbe, 1987). Again, this kind of generation is hard to be guided by an arbitrarily chosen variable.

Another group of WGs use data resampling methods instead of a parametric models. These resampling methods reorder (i.e. draw with replacement) measured time series while conserving important statistics like auto- and cross-correlations. The most commonly used model is a K-Nearest Neighbors (KNN) model that randomly selects values for the current day from a candidate set of "K" nearest neighbors, that are near in terms of similarity of values from previous time steps (Lall and Sharma, 1996).

The Schaake-Shuffle (Clark et al., 2004) is a method to recreate temporal and spatial variability in meteorological ensemble forecasts. Ensemble feature vectors (i.e. a vector of all variables at a time step) are replaced by historical feature vectors by matching the ranks of forecast values and historical values. Daniel S Wilks (2015) argues that this method can be interpreted as a simple form of a an empirical copula. Despite its aim to better represent temporal and spatial variability, the replacement of simulated with observed feature vectors limits possible values and value combinations to ones that occurred historically.

WGs have also been used as a tool to bridge the resolution gap between GCMs and local impact models, a process that has been coined as "stochastic downscaling" (R. Wilby and Wigley, 1997). Daniel S. Wilks (1992) first described how WGEN-parameters can be changed so that precipitation and temperature means and variances follow those projected by a climate model. The changes rely on large sample size statistics of the distributions of precipitation and temperature in order to change daily distribution parameters according to monthly output from the GCM[2].

There is also extensive literature on generating weather not only at one point in space (single-site) but at many points simultaneously (multi-site). An early approach by D. Wilks (1998) employs spatially correlated residuals to drive Richardson-type single-station WGs. Later came latent Gaussian fields for generating spatially correlated precipitation occurrence (Kleiber, Katz, and Rajagopalan, 2012) or for occurrence and precipitation amount combined (Baxevani and Lennartsson, 2015). The latter addresses the "edge-effect" which consists of high precipitation amounts appearing on the boundary of wet areas. Resampling approaches can easily be made to function as multi-site WGs as well, given that they can be made to select the same time step from a multi-station data set across multiple measurement stations. Steinschneider and C. Brown (2013) is an example of such a WG.

Another development step further from single-site to multi-site WGs are gridded WGs which achieve spatially gridded output by interpolation of WG-parameters to unobserved locations. Their conception followed quickly after multi-site WGs (Daniel S. Wilks, 1999). A more recent example is the **A**dvanced **WE**ather **GEN**erator for a two-dimensional grid (AWE-GEN-2d) (Peleg et al., 2017) which combines a total of eight models for different variables and relies in part on physical relationships instead of a purely statistic treatment prevalent in most WGs.

A common problem with WGs is their tendency to underestimate long-term variability. A symptom is a lack of inter-annual variability. This issue is termed "overdispersion". Various techniques have been proposed to tackle overdispersion (Kim et al., 2012). Steinschneider and C. Brown (2013) use autoregressive models of wavelet components to introduce low-frequency signals into a KNN scheme. Multiple authors include observed time series or time-series from circulation-models as covariates which contain information on longer-term behaviour (Mehrotra and Sharma, 2007). The inclusion of a covariate constricts the use of WGs especially when used to generate scenarios as the covariate might not be available for the targeted change. This is less of a problem when using a WG for stochastic downscaling of climate model output, but more so if it used within "what-if"-type experiments as the ones in the application chapter 6.

[2]GCM output at the beginning of the 1990s was usually at monthly resolution.

1.3 Outline of the Thesis

The following chapter 2 explains the methods from existing literature for generating multivariate time series that are relevant for the remainder of the work. They can be seen as building blocks for the weather generators presented later. The methods are Vector-Autoregressive (VAR), Copulas, KNN resampling and phase randomization.

The next chapter 3 presents the first novel weather generator of this thesis, the single-site, daily VG. This chapter introduces novel or adapted methods used throughout the remainder of the thesis, like the variable transformation used and the concept of dryness probability estimation. VG uses transformed dryness probability to treat precipitation as an ordinary variable during VAR generation. Also VAR generation is enhanced by phase randomization of residuals.

Following is chapter 4, which gives a description of WeatherCop, a daily, multi-site weather generator based on copulas and phase randomization.

Chapter 5 starts with a description of a KNN resampling scheme adapted to be directly comparable to VG and WeatherCop. Then, output of this non-parametric approach is compared to that of the parametric models, highlighting strengths and deficiencies of all methods.

The penultimate chapter 6 summarizes published applications of the weather generator VG for Lake Constance at the southern German border to Switzerland and Austria and Lake Kinneret in Israel.

The last chapter 7 gives some ideas to how the models could be developed and used further.

The chapters are in a meaningful reading order as one builds upon the other. If familiar with the used methods, chapter 2 might be omitted, even though the presentation is given as such that important theoretical aspects are highlighted that have consequences in implementation and application. I included back-references in later chapters to concrete locations in the methods chapter to emphasize this. Chapter 4 should not be read without having read chapter 3, because WeatherCop heavily builds upon VG. If only interested in the application, chapter 6 should be understandable to a large extent without familiarity of the rest of the thesis.

2 Methods: Generation of Synthetic Multivariate Time Series

The scope of multivariate time series simulation is vast. In the following, only a few approaches are described in detail. These variable-agnostic methods lack an approach for handling precipitation generation explicitly and thus have to be adapted for being able to generate weather.

This chapter serves as a presentation of the building blocks used to construct the Weather Generators (WGs) in chapters 3 to 5. The methods described here summarize the relevant literature and in some details differ from it in notation in order to achieve consistency in presentation. While not providing a rigorous mathematical treatment with derivations from first principles and proofs, sketches of derivations are shown in order to highlight critical assumptions. Also while derivations are not outlined for every case, every method is still accompanied by a short description of its assumptions, which serves to highlight its limitations from a purely theoretical stand-point.

2.1 Vector-Autoregressive Processes

The Vector-Autoregressive (VAR) process is a multivariate extension of the univariate Autoregressive (AR) process:

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + u_t \tag{2.1}$$

where
- y_t – value at time step t
- p – auto-regressive order
- a_i – parameter of the AR process
- u_t – residual

This means, that the value y_t of every time step is linearly dependent on the values of p previous time steps, leaving the residual u_t to close the identity. AR processes are simple models capable of reproducing linear autocorrelations of low separation length. "Vector" in VAR refers to the fact that for every time-step t there is a vector of values (one value per variable).

Following the notation in (Lütkepohl, 2006, p. 13), the VAR process is defined recursively as follows:

$$\mathbf{y}_t = \sum_{i=1}^{p}(A_i \mathbf{y}_{t-i}) + \mathbf{m} + \mathbf{u}_t \tag{2.2}$$

where
- $\mathbf{y}_t$ – $(k \times 1)$ vector of values at time-step t
- k – number of variables
- p – auto-regressive order
- A_i – $(k \times k)$ matrix containing parameters of the VAR process
- $\mathbf{m}$ – $(k \times 1)$ vector of intercept terms
- $\mathbf{u}_t$ – $(k \times 1)$ vector of residuals or error terms

Thus, the deterministic part $\sum_{i=1}^{p}(A_i \mathbf{y}_{t-1})$ of each value in the time series is constructed as a multiple linear regression of its own and the other variables' values of the past p time steps. If the vector $\mathbf{m}$ contains non-zero elements, $\mathbf{y}_t$ will have a mean vector with non-zero elements as well. The last term, $\mathbf{u}_t$, contains all that is neither attributable to the deterministic part nor $\mathbf{m}$.

Assumptions

Linearity: Values at each time step are linearly dependent on the values of the p previous time steps.

Normal marginals: Common to linear regression, the error terms $\mathbf{u}_t$ are assumed to be Gaussian and serially independent, i.e. "white":

$$\begin{aligned} E(\mathbf{u}_t) &= \mathbf{0}_k \\ E(\mathbf{u}_t \mathbf{u}_s') &= 0_{k \times k} \quad \forall t \neq s \end{aligned} \tag{2.3}$$

where
- $\mathbf{0}_k$ – zero-filled $(k \times 1)$ vector
- $\mathbf{0}_{k \times k}$ – zero-filled $(k \times k)$ matrix

Stability: Reading equation (2.2), it is apparently possible that for certain values of A_i that $|\lim_{t \to \infty} \mathbf{y}_t| = \infty$. These unstable processes are of little practical use and not of interest here. Conversely, only stable processes are regarded in this work. Formally, a VAR process is stable if the following polynomial has no roots in or on the complex unit circle (Lütkepohl, 2006, p. 16):

$$\det\left(I_k - \sum_{i=1}^{p} A_i \mathbf{z}^i\right) \neq 0 \quad \text{for } |\mathbf{z}| \leq 1 \tag{2.4}$$

Weak stationarity: The first and second statistical moment (mean and variance) of the process do not change with time. This is a weaker form than strong stationarity, which means that the joint distribution of the vectors $(\mathbf{y}_t, \ldots, \mathbf{y}_{t+n})$ remains the same irregardless of time shifts for all $n \in \{\mathbb{N} | n > 0\}$. Formulated as such, the time-invariance includes correlations, auto-correlations and cross-correlations (correlations with time-shift). Roughly speaking, this means that the statistical properties of the data remain

the same when looking at different time steps. Weak stationarity is defined in a more formal way in the following.

Equation (2.5) states that irrespective of the place in the time series, the expected vector of means is μ.

$$E(\mathbf{y}_t) = \mu \quad \forall t \in \{\mathbb{N} | t \leq T\} \tag{2.5}$$

where T – number of time steps
μ – $(k \times 1)$ process mean vector

It can be shown that weak stationarity follows from the stability condition given in equation (2.4).

Estimation

In order to simulate time series with the help of a VAR-process, one needs values for the parameters A_i and the covariance matrix of u for the drawing of random residuals. Suitable parameter values can be calculated from data with the help of an estimator. Assembling observed data in matrices and formulating an expression for the error terms u, yields a minimization problem that can be solved with linear algebra. Following Lütkepohl (2006, p. 70):

$$\begin{aligned}
Y &:= (\mathbf{y}_1, \ldots, \mathbf{y}_T) && (k \times T), \\
B &:= (\mathbf{m}, A_1, \ldots, A_p) && (k \times (pk+1)), \\
Z_t &:= \begin{bmatrix} 1 \\ \mathbf{y}_t \\ \vdots \\ \mathbf{y}_{t-p+1} \end{bmatrix} && ((pk+1) \times 1), \\
Z &:= (Z_0, \ldots, Z_{T-1}) && ((pk+1) \times T), \\
U &:= (\mathbf{u}_1, \ldots, \mathbf{u}_T) && (k \times T)
\end{aligned} \tag{2.6}$$

equation (2.2) can be written in a very compact form:

$$Y = BZ + U \tag{2.7}$$

One has to keep in mind, that despite the fact that the terms A_i are related to the covariance matrix $\mathrm{Cov}(\mathbf{y}_t, \mathbf{y}_t)$, the noise still has a covariance matrix with non-zero off-diagonal elements:

$$\Sigma_u = E(\mathbf{u}_t \mathbf{u}_t') \tag{2.8}$$

Minimizing the determinant of the error covariance matrix, the estimators for $\hat{B}$ and $\hat{\Sigma}_u$ (the covariance matrix of the error terms) become:

$$\hat{B} = YZ'(ZZ')^{-1} \tag{2.9}$$

$$\hat{\Sigma}_u = \frac{1}{T - pk - 1}\left(YY' - \hat{B}ZY'\right) \tag{2.10}$$

The denominator $T-pk-1$ arises due to an adjustment for degrees of freedom (Lütkepohl, 2006, p. 75).

Simulation

A prerequisite for simulation is that it is possible to generate random numbers that resemble the residuals $\mathbf{u}$. It is possible to draw serially independent random vectors from a multivariate Gaussian distribution with the estimated covariance matrix from equation (2.9) if the residuals are "white", meaning that there is no correlation between the residuals at different time steps.

In principle, one can fill the first p time-steps of the generated time-series $\hat{\mathbf{y}}$ with the process mean μ. This would, however, bias the beginning of $\hat{\mathbf{y}}$ towards the mean and thus reduce variability. In practice one generates more time steps than needed and discards the ones at the beginning which solves this problem.

In terms of implementation, U can be drawn first from a multivariate Gaussian distribution with mean $\mu = \mathbf{m}$ and covariance matrix $\Sigma_{\mathbf{u}}$. After that, the products $A_i\mathbf{y}_{t-i}$ can be applied for every time step t. Due to the inherently recursive nature of the VAR-process, this can only be implemented as a loop over the time domain.

2.2 Copula Approaches

VAR processes model linear dependencies and generate normal distributions if their parameterizations are stable and Gaussian error terms ($\mathbf{u}_t$) are used for simulation. However, linear dependencies and normal marginals are not necessarily good approximations to the data at hand. Copula models are models of dependence and thus enable an explicit shaping of possibly non-linear, asymmetric dependencies. Their theoretic appeal lies in part in their unambiguous nature. Under the mild assumption of the existence of continuous marginal distribution functions there exists one unique copula. Thus the copula is the expression of pure dependence.

Concretely, the copula is a multivariate distribution function with uniform univariate marginals between 0 and 1. The copula can be linked to the multivariate distribution function with Sklar's theorem (1959):

$$H(x_1, \ldots, x_k) = C\left(F_1(x_1), \ldots, F_k(x_k)\right) \tag{2.11}$$

where H – multivariate distribution function
k – number of dimensions
C – copula: maps $[0,1]^k \rightarrow [0,1]$
F_i – marginal distribution function of the ith variable

Furthermore, the multivariate density function $h(x_1, \ldots, x_k)$ can be formulated with the copula density $c(F(x_1), \ldots, F(x_k))$ and the univariate densities $f_1(x_1), \ldots, f_k(x_k)$:

$$h(x_1, \ldots, x_k) = \frac{\partial^k H(x_1, \ldots, x_k)}{\partial x_1 \cdots \partial x_k} \tag{2.12}$$

$$= c(F(x_1), \ldots, F(x_k)) \cdot f_1(x_1) \cdots f_k(x_k) \tag{2.13}$$

As given in Sklar's theorem (equation (2.11)) the inputs to the copula are the non-exceedance probabilities of x_i (c.f. $F_i(x_i)$). Hence, the copula has no connection to the marginal distributions of the data, which enables a separate treatment of dependencies and marginals. This also makes the joint modelling of variables with different marginals possible, which stands in contrast to multivariate parametric distributions as the marginals of the latter all share the same distribution family.

In practice, the effectiveness of copula modelling relies on the availability of suitable parametric copula families or enough data for non-parametric approximations. While there exist a multitude of copula families exhibiting a lot of desirable properties in the bivariate case (asymmetry, tail-dependence, variable strength of correlation), finding equally able copula families for higher-dimensional data is a topic of current research (D. Schirmacher and E. Schirmacher, 2008; Hao and Singh, 2016).

Vines

As long as flexible higher-dimensional parametric copulas are missing, the abundance of bivariate copulas can still be exploited to construct higher-dimensional copulas. A method proposed by Joe (1994) models a multivariate dependence structure by building pair-wise relationships with bivariate copulas. These pairs are first assembled from the original variables in such a way that every variable is associated with every other variable directly or indirectly with $k-1$ pair relationships (with k being the number of variables). Seeing the variables as nodes and the bivariate copulas as edges, this network is a tree in graph theory terminology. A tree is a connected graph with k nodes and $k-1$ edges. A connected graph is a network in which every node can be reached from every other node, i.e. there exists a path between each pair of nodes. In order to account for higher order dependencies, additional pairings are made from so-called virtual-observations derived from conditional distributions based on previously fitted bivariate copulas. The pairings have to be constructed in such a way that the variables involved from previous trees differ in exactly two elements

(proximity condition). These pairings again result in trees. Each tree has one fewer node as the one before because a tree has $k-1$ edges and going from one tree to the next results in edges from the previous tree becoming nodes in the next, resulting in $k-1$ trees.

A structure that fits the previous criteria (collection of $k-1$ trees, nodes in one tree become edges in the next and the proximity condition), is called a regular vine (R-Vine). More specific structures can be defined around the definition of the degree of a node, which is the number of edges attached to it. A canonical vine (C-Vine) is a an R-Vine which contains one central node with the maximum degree possible in each tree. A drawable vine (D-Vine) is an R-Vine in which the maximum degree in the first tree is 2, resulting in a string-like structure. The appearance of these nested trees (c.f. figure 2.1) motivated the name "vine" (Cooke, 1997).

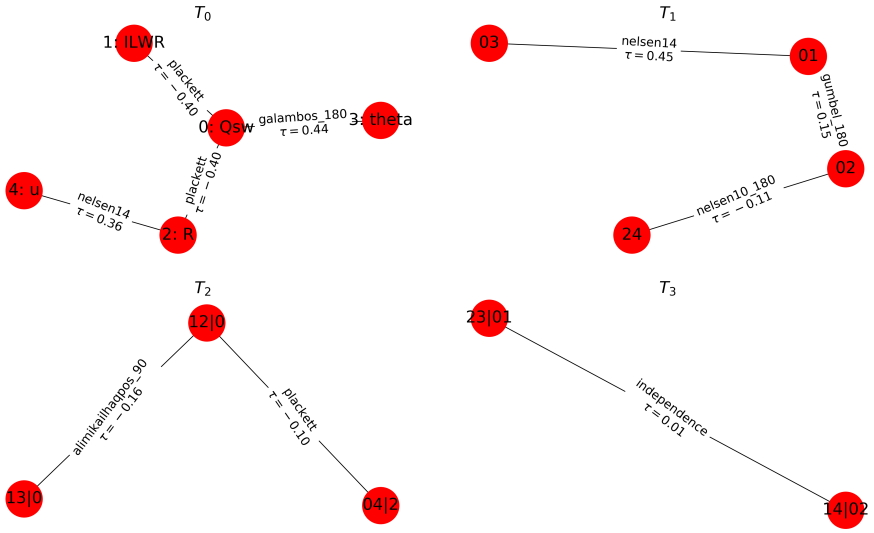


Figure 2.1: Vine structure of a vine fitted to a 5-dimensional meteorological dataset. $\mathcal{T}_1$ is the first vine tree which shows which variable is related to which other variable via bivariate copulas. τ are Kendall rank coefficients which are used to decide on the tree structure. $\mathcal{T}_2, \ldots, \mathcal{T}_4$ are trees constructed from conditioned ranks of data inserted in the previous trees. $\mathcal{T}_2$ is a drawable vine (D-Vine), whereas $\mathcal{T}_3$ is a drawable as well as a canonical vine (C-Vine) with $12|0$ as a central node. $\mathcal{T}_1$ is an example for regular vine structure (R-vine) that is neither a D- or a C-Vine.

Vine constructions are possible because multivariate probability densities can be decomposed into products of pair-wise relationships. In order to achieve this, the conditional probability rule has to be extended to the product rule:

$$P(A,B) = P(A)P(B|A) \tag{2.14}$$

$$P(A,B,C) = P(A,B)P(C|A,B) = P(A)P(B|A)P(C|A,B) \tag{2.15}$$

This means that the joint probability of event A and B is given by the probability of

event A (regardless of B) times the probability of event B given that A appeared. Equation (2.14) can be applied recursively to achieve the joint probability of 3 and more events (equation (2.15)).

In the following, a 3-dimensional probability density function is decomposed using two bivariate copulas. Ultimately, a third bivariate copula will appear and complete the pair-construction method for three variables. This mathematical excursion helps to understand not only how vines arise, but more importantly, where simplifications are made. Using the product rule equation (2.15), the density can be written as:

$$f(x_1, x_2, x_3) = f_1(x_1) f_{2|1}(x_2|x_1) f_{3|12}(x_3|x_1, x_2) \tag{2.16}$$

The decomposition will result in lengthy equations, so in order to improve readability, parameters of the probability density functions (*pdf*s) and cumulative distribution functions (*cdf*s) will be omitted in the following text, as they are apparent from the subscripts (e.g. $f_{3|12} \equiv f_{3|12}(x_3|x_1, x_2)$ and $F_1 \equiv F(x_1)$).

For the term $f_{2|1}$ rearranging equation (2.14) and using equation (2.12) for the bivariate unconditional density f_{12} yields:

$$f_{2|1} = \frac{f_{12}}{f_1} = \frac{c_{12}(F_1, F_2) f_1 f_2}{f_1} = c_{12}(F_1, F_2)\ f_2 \tag{2.17}$$

Equation (2.17) will be used in the following in this general form:

$$f_{j|i} = c_{ij}(F_i, F_j)\ f_j \tag{2.18}$$

The term $f_{3|12}$ can be decomposed in two ways ((I) and (II)), depending on which of the variables x_1 or x_2 is used to condition on first:

$$\text{(I)}: \qquad f_{3|12} = c_{13|2}(F_{1|2}, F_{3|2}, \mathbf{x_2})\ f_{3|2} \tag{2.19}$$

$$\text{(II)}: \qquad f_{3|12} = c_{23|1}(F_{2|1}, F_{3|1}, \mathbf{x_1})\ f_{3|1} \tag{2.20}$$

where $F_{i|j}$ – conditioned ("virtual") rank given by $C_{i|j}(x_i|x_j)$

It has to be stressed that, in general, the copulas $c_{13|2}$ and $c_{23|1}$ depend on the values of their conditioning variable $\mathbf{x_2}$ and $\mathbf{x_1}$, respectively. This dependency is consciously neglected in practical applications and referred to as the *simplifying assumption*. It is assumed that the influence of the conditioning variable (e.g. x_2) is removed by forming the conditioned ranks (e.g. $F_{1|2}$ and $F_{3|2}$).

The bivariate conditional densities in equation (2.19) and equation (2.20) ($f_{3|2}$ and $f_{3|1}$) can be replaced using equation (2.18) again:

$$\text{(I)}: \qquad f_{3|2} = c_{23}(F_2, F_3)\ f_3 \tag{2.21}$$

$$\text{(II)}: \qquad f_{3|1} = c_{13}(F_1, F_3)\ f_3 \tag{2.22}$$

At this point equation (2.16) can be expressed with only using bivariate copulas and univariate densities:

$$\begin{aligned} \text{(I)}: \quad f =& f_1 \\ & c_{12}(F_1, F_2)\ f_2 \\ & c_{13|2}(F_{1|2}, F_{3|2}, x_2)\ c_{23}(F_2, F_3)\ f_3 \end{aligned} \tag{2.23}$$

$$\begin{aligned} \text{(II)}: \quad f =& f_1 \\ & c_{12}(F_1, F_2)\ f_2 \\ & c_{23|1}(F_{2|1}, F_{3|1}, x_1)\ c_{13}(F_1, F_3)\ f_3 \end{aligned} \tag{2.24}$$

Both (I) and (II) connect x_1 and x_2 via c_{12}. However, (I) models the dependence between pair (x_1, x_2) and x_3 by the direct bivariate copula density c_{23} and the indirect copula density $c_{13|2}$, where (II) uses c_{13} and $c_{23|1}$ respectively. At this point, it becomes convenient to visualize these dependency structures. Figure 2.2 shows a graph for each vine structure (I) and (II).

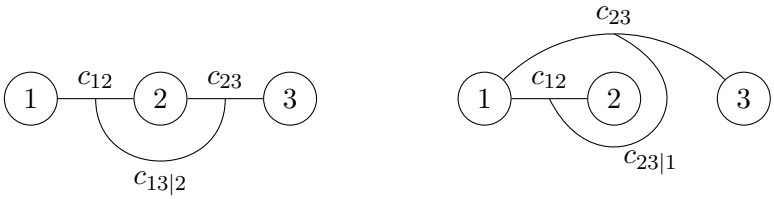


Figure 2.2: 3d-vines (I) and (II) using two trees each.

For 4 variables, more options to decompose the full copula density into bivariate copula densities appear. Starting again with the product rule equation (2.15), the 4-dimensional density is:

$$f = f_1\ f_{2|1}\ f_{3|12}\ f_{4|123} \tag{2.25}$$

As in equations (2.19) and (2.19) one of the conditioning variables of $f_{4|123}$, can be chosen to decompose further:

$$\text{(I)}: \quad f_{4|123} = c_{14|23}(F_{1|23}, F_{4|23}, x_2, x_3)\ f_{4|23} \tag{2.26}$$

$$\text{(II)}: \quad f_{4|123} = c_{24|13}(F_{2|13}, F_{4|13}, x_1, x_3)\ f_{4|13} \tag{2.27}$$

$$\text{(III)}: \quad f_{4|123} = c_{34|12}(F_{3|12}, F_{4|12}, x_1, x_2)\ f_{4|12} \tag{2.28}$$

Because the last term of each of (I) – (III) can be decomposed in two ways (as in equations (2.19) and (2.20)), the number of possible vine structures increases by a factor of 3.

If decomposing in ascending order of variable index, one ends up with the so called D-Vine structure (also see figure 2.3):

$$\begin{aligned}\text{(I)}: f =& f_1 \\ & c_{12}(F_1, F_2)\ f_2 \\ & c_{13|2}(F_{1|2}, F_{3|2}, x_2)\ c_{23}(F_2, F_3)\ f_3 \\ & c_{14|23}(F_{1|23}, F_{4|23}, x_2, x_3)\ c_{24|3}(F_{2|3}, F_{4|3}, x_3)\ c_{34}(F_3, F_4)\ f_4\end{aligned} \tag{2.29}$$

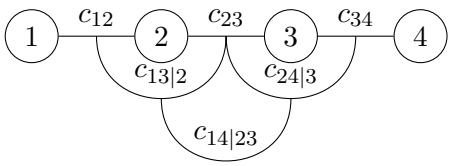


Figure 2.3: 4d-vine (I) – a D-Vine composed of three trees.

Another special structure is obtained when decomposing in the following way:

$$
\begin{aligned}
\text{(II)}: \qquad f =& f_1 \\
& c_{12}(F_1, F_2)\ f_2 \\
& c_{23|1}(F_{2|1}, F_{3|1}, x_1)\ c_{13}(F_1, F_3)\ f_3 \\
& c_{24|13}(F_{2|13}, F_{4|13}, x_1, x_3)\ c_{34|1}(F_{3|1}, F_{4|1}, x_1)\ c_{14}(F_1, F_4)\ f_4
\end{aligned} \tag{2.30}
$$

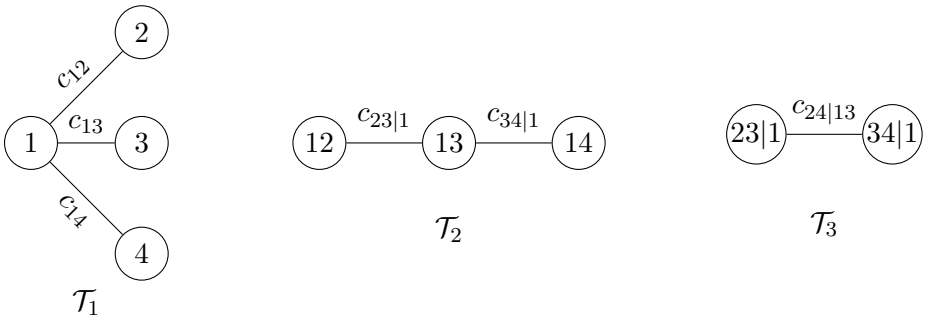


Figure 2.4: 4d-vine (II) – a C-Vine with the three trees $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_3$. The notable difference to the D-Vine structure is the existence of a central node in tree $\mathcal{T}_1$.

The tree figures of the 4d-vines reveal another, subtle property. While the two decompositions of 3d-vines in figure 2.2 can be made equivalent by relabelling the variables, a similar operation can not be done in four dimensions. In 5 and more dimensions, regular vines that are neither drawable nor canonical appear (see $\mathcal{T}_1$ in figure 2.1 for an example).

While the decomposition in general is mathematically valid, the practice of applying the simplifying assumption results in vines that are merely estimations of the full multivariate distribution.

Estimation

Fitting a vine copula to a data set is a non-trivial task. First, the vine structure represents only an approximation to the true dependence structure and many structures might lead to good fits. The number of possible R-vine structures grows extremely fast with the number of dimensions ($N_k = \frac{1}{2}k!2^{(k-3)(k-2)/2}$ for k dimensions (Morales-Nápoles, 2010)). For every possible vine structure, which consists of $k-1$ trees, $k(k-1)/2$ bivariate copulas have to be fitted. Dißmann et al. (2013) proposed an algorithm that aims to find a good R-vine structure by pairing variables in such a way that the sum of absolute values of correlations is maximized for each tree, starting with the largest one. For C- and D-Vines, the number of possible structures is much lower. Also, when fitting such a more specific vine, this limiting path is usually chosen where there is a known hierarchy in the dependence structure and thus, not all theoretically possible structures are necessary to consider. D-Vines are appropriate for time series who's temporal order suggests a likewise order of dependencies. C-Vine

structures are centered around one root node in each tree. This central node corresponds to a variable on which all others are directly dependent.

I concentrate on C-Vine estimation in the following, because this vine family is the one used in the WG **Weather** Generator based on Phase Randomization and Vine **Cop**ulas (WeatherCop) in section 4.2. While the number of C-Vine structures is low compared to R-vines given the same number of dimensions, it is still large. So finding the best structure under the $N_k = k!/2$ possible ones (Aas et al., 2009), can be intractable even for moderate number of dimensions k. This motivated the sequential algorithm proposed by Czado, Schepsmeier, and Min (2012), which further utilizes the heuristic that most dependence is captured by the bivariate copulas in the first few trees. Dißmann et al. (2013) uses a related line of reasoning for his sequential algorithm for selection of R-vines. Namely, the model fit depends mostly on the copulas used for the first tree.

The details are given in algorithm 1. For each tree, a root node has to be found that, when used as such, the sum of absolute Kendall taus[1] for the (conditioned) ranks that are used for fitting the bivariate copulas is maximal. Conditioned ranks used in a tree depend on the copulas fitted on ranks in the earlier tree. As described in the previous paragraph, it is argued that most dependence lies in the first trees. By deciding on the structure and copulas of the first tree first, this method ensures the biggest gains early, without changing these decisions later on. Therefore, this algorithm can be categorized as greedy and cannot guarantee finding the global optimum, which might involve weaker dependencies in earlier trees in favour of much stronger dependencies in later ones.

To the modeller it seems that all copulas have the same number of data points available for fitting and the same number of copula parameters to estimate. All bivariate copulas in a simplified vine are modelled as being a function of a pair of ranks and the parameter vector θ – and not the values of the ranks they are indirectly conditioned in earlier trees. This means that the simplifying assumption obscures the fact that, ideally, a lot of data is needed to robustly fit the deeply conditional bivariate copulas in the higher trees.

Simulation

While specific methods exist to sample from particular copula families, the most general method is based on sequences of bivariate conditional distributions. This technique is called conditional method or Rosenblatt's method and is an inverted formulation of the Rosenblatt transformation. The Rosenblatt transformation transforms a k-dimensional distribution into k independent variables (Rosenblatt, 1952). Equivalent to sampling from multivariate distributions, for sampling from the copula C one starts with a vector of values drawn from uniform distributions between 0 and 1: $p_1, \dots p_k$. By applying:

[1] This is the Kendall rank correlation coefficient, which is based on the number of pairs with concordant sort order of both variables. Values range from -1 (inverse ranking) to 1 (equal ranking).

Algorithm 1: Sequential selection of C-Vine and pair-copulas; adapted from Czado, Schepsmeier, and Min (2012) and Joe (2015, Algorithm 30, p. 304)

Data: Input data ranks $\mathbf{u}_1, \ldots, \mathbf{u}_k$.
Optionally: index of root node of first tree i

Compute empirical Kendall's tau $\hat{\tau}_{ij}$ for all pairs of variables.
`// First tree`
Let tree level $l \leftarrow 1$
If the root node i is not given, find it by maximizing $\sum_{j=1}^{k} |\hat{\tau}_{ij}|$.
Let tree 1 $\mathcal{T}_1 \leftarrow \{[i,j] \mid j \in \mathbb{N} \wedge j \leq k \wedge j \neq i\}$.
forall edges $\mathbf{e} = (i,j) \in \mathcal{T}_1$ **do**
    Select parametric bivariate copulas for the edge e and estimate the corresponding copula parameters $\hat{\theta}_{ij}$; choosing copulas by maximum likelihood.
    for $m \leftarrow 1$ **to** k **do**
        Save conditioned ranks $C_{i|j}(u_{mi}|u_{mj}; \hat{\theta}_{ij})$ and $C_{j|i}(u_{mj}|u_{mi}; \hat{\theta}_{ij})$.
`// Next trees`
for $l \leftarrow 2$ **to** $k-1$ **do**
    Compute empirical conditional Kendall taus $\hat{\tau}_{pq;S}$ from the conditioned ranks $C_{p|S}(u_{ip}|u_{i,S}; \hat{\theta}_{p\cup S})$ and $C_{q|S}(u_{iq}|u_{i,S}; \hat{\theta}_{q\cup S})$ over all edges $[pq|S]$ from tree $\mathcal{T}_{l-1}$. S is the set of conditioning variables from earlier trees.
    Find the root node $\{p\}$ by maximizing $\sum |\hat{\tau}_{pq;S}|$ where $\{p\} \cup S$ and $\{q\} \cup S$ are sets in the tree $l-1$.
    Let tree l $\mathcal{T}_l \leftarrow \{[p,q|S]|\, \{q\} \cup S \in \mathcal{T}_{l-1}\}$ (with $k-l$ edges).
    Select parametric bivariate copulas for the edges $e \in \mathcal{T}_l$ and estimate the copula parameters $\hat{\theta}_{\{pq\}\cup S}$.
    forall edges $[pq|S] \in \mathcal{T}_l$ **do**
        Save conditioned ranks $C_{p|q;S}(u_{ip}|u_{i,q\cup S}; \hat{\theta}_{\{pq\}\cup S})$ and $C_{q|p;S}(u_{iq}|u_{i,p\cup S}; \hat{\theta}_{\{pq\}\cup S})$.

$$
\begin{aligned}
u_1 &= p_1 \\
u_2 &= C_{2|1}^{-1}(p_2|u_1) \\
u_3 &= C_{3|12}^{-1}(p_3|u_1, u_2) \\
&\vdots \\
u_k &= C_{k|1:(k-1)}^{-1}\left(p_{k-1}|u_{1:(k-1)}\right)
\end{aligned} \tag{2.31}
$$

where $C_{2|1}^{-1}$ – inverse of the conditional distribution $C_{2|1}$

one obtains the vector $(u_1, \ldots, u_k)$ which follows the multivariate distribution of C. u_1 and u_2 can be obtained easily, but calculation of subsequent values of u_k require recursive evaluations of bivariate inverse conditional copulas. The order of the evaluations and the concrete copulas are determined by the vine structure.

The C-Vine sampling algorithm 2 reminds one of a pairwise decomposed equation (2.31). In fact, given a multivariate copula, one can sample from that in the same way (the necessary bivariate copulas can be derived from the full multivariate distribution). However, by setting the vine structure, one has lost the flexibility of choosing the order of variables to sample. The order of which bivariate copulas to sample from is fixed in the vine representation.

Algorithm 2: Sampling from a C-Vine; Joe (2015, Algorithm 15, p. 291)

Data: Generate $p_1, \ldots, p_d$ to be independent $U(1, 0)$ random variables.

Let $u_1 \leftarrow p_1$
Let $u_2 \leftarrow C_{2|1}^{-1}(p_2|p_1)$
for $j = 3, \ldots, d$: **do**
    let $q \leftarrow p_j$
    for $l = j - 1, j - 2, \ldots, 1$ **do**
        let $q \leftarrow C_{j|l;1:(l-1)}^{-1}(q|p_l)$
    let $u_j \leftarrow q$
return $(u_1, \ldots, u_d)$

Bivariate Archimedean Copulas

I present this class here only in its bivariate form, as higher-dimensional Archimedean copulas are not as flexible as vine constructions and bivariate Archimedean are useful as building blocks for pair-wise constructions. Higher-dimensional Archimedeans exist, but use one or two parameters to tune the strength of dependence between all variables, which is not flexible enough for multivariate meteorological data sets.

The Archimedean class of copulas has many members due to its simple construction method via a so-called generator function. This generator function ψ must enable formulating the sum of the function of the ranks as the function of the copula:

$$\psi\left(C(u_1, u_2)\right) = \psi(u_1) + \psi(u_2) \tag{2.32}$$

where ψ – generator function mapping from $[0, 1] \rightarrow [0, \infty]$

One then solves for $C(u_1, u_2)$ with the help of the pseudo-inverse $\psi^{[-1]}$:

$$\psi^{[-1]}(t) = \begin{cases} \psi^{-1}(t) & \text{if } 0 \leq t \leq \psi(0) \\ 0 & \text{if } \psi(0) \leq t \leq \infty \end{cases} \tag{2.33}$$

with which the Archimedean can be constructed as such:

$$C(u_1, u_2) = \psi^{[-1]}\left(\psi(u_1) + \psi(u_2)\right) \tag{2.34}$$

In order for $C(u_1, u_2)$ to be a copula, ψ has to be a continuous strictly decreasing convex function with $\psi(1) = 0$.

An example for such a copula is the Gumbel copula:

$$\psi(t) = \left(-\ln(t)\right)^{\theta} \tag{2.35}$$

$$C(u_1, u_2) = \exp\left(-((-\ln(u_1))^{\theta} + (-\ln(u_2))^{\theta})^{1/\theta}\right) \tag{2.36}$$

2.3 Resampling Methods

Resampling methods within the field of generation of multivariate time series offer simple ways to produce random samples from an observed data set. Samples from this source data are drawn in a serially dependent fashion in order to reproduce temporal dependencies. Inter-variate dependence is maintained by resampling time steps instead of individual variables.

Lall and Sharma (1996) introduced a resampling method based on a K-Nearest Neighbors (KNN) scheme. Candidates for individual time steps are identified by a weighted distance measure applied to the values of m previously selected time steps. The candidate is not chosen with equal probability, but based on a discrete kernel density estimation (KDE) of the k nearest neighbors.

Estimation

Compared to parametric methods, resampling schemes require no or fewer parameters to be estimated and are therefore classified as being non-parametric. The KNN scheme requires the number of considered neighbors (which refers to the "K" in KNN). Lall and Sharma (1996) suggest to set this number to the square root of the number of time steps in the source data.

If climate scenarios are considered, some method to produce biased time series is required. This necessitates the fitting of parameters to the source data. In section 5.2.2, such a method is described.

Simulation

The aim of the KNN method is to randomly choose time steps from the source data without destroying the temporal and inter-variate dependence. One begins by randomly choosing a chunk of p subsequent time steps from the multivariate source time series. Then, the squared euclidean distance between this and all other chunks in the data is computed:

$$d_s = \sum_{i=1}^{p} \sum_{j=1}^{k} \left(y_{t-i,j} - x_{s-i,j}\right)^2 \quad \forall s \in \{p+1, \ldots, T\} \tag{2.37}$$

where
- d_s – squared euclidean distance between chunk in the source data before time step s and resampled data at time step t
- s – time index in the source data
- p – number of previous time-steps considered
- k – number of variables
- t – time index in the resampled data
- $y_{t-i,j}$ – resampled value of variable j at time step t
- $x_{s-i,j}$ – source data value of variable j at time step $s-i$
- T – number of time steps in the source data set

The elements of the set of all d_s are put in ascending order. From that ordered set, one of the first q elements[2] is randomly chosen (say, time step τ). The candidates are not chosen with equal weight, but in order to maintain the distribution of $F(x_t|x_{t-1}, \ldots x_{t-p})$, candidates with a smaller distance d_s are given a higher chance of being selected. Following Lall and Sharma (1996), these probabilities are determined as follows:

$$P(s(l)) = \frac{1/l}{\sum_{i=1}^{q} 1/i} \tag{2.38}$$

Where l is the position of s in the ordered list of candidate time steps, so $d_{s(l)}$ is the l-th most similar chunk to the chunk preceding time step s.

[2]Note that this would be the "k" of KNN, but k is used throughout this chapter as the number of variables.

The current time step in the resampled series is then fixed as

$$\mathbf{y}_t = \mathbf{x}_\tau \tag{2.39}$$

The algorithm then proceeds with the next time step $t + 1$.

In essence, the vector $\mathbf{y}_t$ is chosen based on the similarity of the preceding days to chunks of data in the observations. Because of this, the auto-covariance matrix $\Gamma(h) = \mathrm{Cov}(\mathbf{y}_t, \mathbf{y}_{t-h})$ should largely stay the same, subject to the choices of p and q. Further, due to the fact that for each time step, all variables from the source data are sampled, the covariance matrix is close to the one of the source data $\mathrm{Cov}(\mathbf{y}_t, \mathbf{y}_t) = \mathrm{Cov}(\mathbf{x}_t, \mathbf{x}_t)$. The reason it is not identical is due to the fact that the time steps from the source data do not appear with equal frequency as in the source data. Section 5.4.1 explores this aspect.

Notice, that despite differences in notation, all "generated" values y appear in x, i.e. arguably, y should not be referred to as "generated" but rather as "resampled" values.

The euclidean distance does not distinguish different scales which is problematic when generating different meteorological variables at once. This issue has been approached by using a weighted euclidean distance (e.g. Rajagopalan and Lall, 1999) or using a different distance metric like the Mahalanobis distance (Yates, 2003; T. a. Buishand and Wójcik, 2003) which is scale-invariant as it uses distances along the principle components of the data.

Assumptions

As equation (2.37) is a distance measure that combines more than one variable, it is necessary to deal with the different scale of the involved variables. In other words, each dimension has to be given equal weight. Lall and Sharma (1996) address this by introducing a weight for each variables' squared difference. Another way to treat the assumption of equal distances in all dimension is to convert all variables to the same distribution, which has the desirable effect of putting all variables on the same scale.

The most notable assumption, however, consists in the treatment of the source data as its own abstraction. A consequence is not only that the generated extrema do not exceed the ones in the source data, but that only the combinations of values (the vectors $\mathbf{y}_t$) present in the source data can be produced. This topic is further explored in chapter 5.

2.4 Phase Randomization

Phase randomization was introduced by J. Theiler et al. (1992) as a method to generate "surrogate data" for testing the null hypothesis that a given time series was generated by a linear process. This works because phase randomization is able to generate time series according to *any linear* autocorrelation function. Since phase randomization can be used to generate time series that share the autocorrelation function of a measured time series, it can provide

a Monte-Carlo estimate of the population the measured time series might have been drawn from. By comparing a test statistic that is related to non-linear properties with the empirical distribution of statistics from the sample of "surrogate data", a decision on whether to accept or reject the null hypothesis can be reached.

Apart from its intended use for null hypothesis testing, phase randomization can be used as a technique to generate random time series (or fields in two or more dimensions) with a given covariance structure. Its appeal lies in the fact that it is a non-parametric approach and that it allows for maintaining linear dependencies with all possible time lags.

In the following, the steps to separate a signal into an amplitude and a phase spectrum will be sketched. These steps are necessary as, at its core, phase randomization consists of keeping the amplitude spectrum and randomizing the phase spectrum of a time series.

The main tool for phase randomization is the Fourier Transform (FT). With the FT, *any* continuous function $f(t)$ in the time domain can be expressed as an infinite series of harmonics in the frequency domain:

$$F(u) = \int_{-\infty}^{\infty} f(t) e^{-2\pi i u t} dt \tag{2.40}$$

where $F(u)$ – contribution of frequency u
t – time

It is possible to transform $F(u)$ back to the time domain by integrating $F(u)$ over the frequency dimension:

$$f(t) = \int_{-\infty}^{\infty} F(u) e^{2\pi i u t} du \tag{2.41}$$

If these transformations exist for $F(u)$ and $f(t)$, $F(u)$ and $f(t)$ are said to be *transform pairs*.

$F(u)$ is a complex-valued function with real input ($f : \mathbb{R} \to \mathbb{C}$) and its results contain information on the amplitude and phase angle of the signal at frequency u. It can be reformulated with the help of Eulers' formula:

$$e^{ix} = \cos(x) + i \sin(x) \tag{2.42}$$

Leading to:

$$F(u) = r(u) \left(\cos(\theta(u)) + i \sin(\theta(u))\right) = r(u) e^{i\theta(u)} \tag{2.43}$$

where $\theta(u)$ – phase angle for frequency u
$r(u)$ – amplitude of frequency u given by the complex radius of $F(u)$

With the complex radius r being the absolute value of $F(u)$, equation (2.43) can be reformulated as:

$$F(u) = |F(u)|\, e^{i\theta(u)} \tag{2.44}$$

In which $|F(u)|$ is said to be the amplitude spectrum and $\theta(u)$ the phase spectrum.

A convenient consequence of the integral formulation in equation (2.40) is that the autocorrelation function can be expressed in the frequency domain. This stems from the fact that the autocorrelation is a convolution of a function with its own complex conjugate:

$$\mathrm{Corr}(f, f)(\tau) = \int_{-\infty}^{\infty} f(t) f^*(t+\tau) dt = (f * f^*)(\tau) \tag{2.45}$$

where f^*– complex conjugate of f
τ – time shift
$*$ – convolution

And it can be shown that for f and g being continuous functions in the time domain with F and G being their FTs, $f * g^*$ and FG^* form a transform pair. This is called the Correlation Theorem (see e.g. Olson (2017, p. 131)[3]). With it the autocorrelation in the time domain consists of a simple multiplication in the frequency domain.

$$\mathrm{Corr}(f, f)(\tau) = \int_{-\infty}^{\infty} F(u) F^*(u) e^{2\pi i \tau u} du \tag{2.46}$$

where $F^*(u)$ – complex conjugate of $F(u)$

The complex conjugate of $F(u)$ can be expressed in terms of equation (2.44):

$$F^*(u) = |F(u)|\, e^{-i\theta(u)} \tag{2.47}$$

Looking at the product $F(u)F^*(u)$ from the right-hand side of equation (2.46) and combining equation equation (2.47) again with equation (2.44) one immediately sees that the phase spectrum $\theta(u)$ cancels out:

$$F(u)F^*(u) = |F(u)|^2 \frac{e^{i\theta(u)}}{e^{i\theta(u)}} = |F(u)|^2 \tag{2.48}$$

$|F(u)|^2$ is also called the spectral density $S(u)$:

[3]The proof consists of forming the FT of $f * g^*$, resulting in an expression with two integrals (one from the FT and one from the convolution). After reordering the order of integration, it can be seen that the result of the integration that contains τ does not depend on it. One can then reformulate the expression as a product of F and G.

$$S(u) = |F(u)|^2 \tag{2.49}$$

The spectral density $S(u)$ is of importance, because it is part of the Wiener-Kinchin-Theorem (Wiener, 1930), which connects it to the autocorrelation function in the time domain:

$$\Gamma(\tau) = \int_{-\infty}^{\infty} S(u) e^{2\pi i \tau u} du \tag{2.50}$$

where $\Gamma(\tau)$ – autocorrelation at time shift τ

In summary, the autocorrelation function depends on the spectral density which is obtained from the amplitude spectrum alone. Hence, the autocorrelation function is independent from the phase spectrum; it is "phase-blind".

In the context of multivariate time series it is necessary not only to reproduce the autocorrelation function, but also the crosscorrelation function. The crosscorrelation function between function f and g is also a convolution:

$$\mathrm{CrossCorr}(f,g)(\tau) = \int_{-\infty}^{\infty} f(t) g(t+\tau) dt = (f * g^*)(\tau) \tag{2.51}$$

But in contrast to equation (2.46), there are two different phase spectra involved, so they do not cancel out, which means that the cross-correlation still depends on the phases:

$$\mathrm{CrossCorr}(f,g)(\tau) = \int_{-\infty}^{\infty} |F(u)| \, e^{i\theta_f(u)} \, |G(u)| \, e^{-i\theta_g(u)} e^{2\pi i \tau u} du \tag{2.52}$$

where $\theta_f(u)$ – phase spectrum of f
$\theta_g(u)$ – phase spectrum of g

However, in the context of phase randomization, in which the phase spectrum is changed, it is possible to keep the cross-correlation intact by using the same random phases for both variables. This method was first described by Prichard and Theiler (1994).

In practice, one uses the discrete FT, which results in a replacement of the integral in equation (2.40) with a sum over all time steps. More concrete, a time series y_t can then be expressed in terms of cosine waves with distinct amplitudes and phase shifts:

$$y_t = \frac{a_0}{2} + \sum_{j=1}^{T} A_j \cos\left(j\frac{\pi t}{T} + \phi_j\right) \tag{2.53}$$

where a_0 – amplitude of the zero-frequency harmonic
T – length of the time series
A_j – amplitude of the jth harmonic
ϕ_j – phase angle of jth harmonic

Assumptions

The autocorrelation is assumed to be stationary (only dependent on the distance) and circular (distance wraps around).

Phase randomization only retains the linear correlation structure of the original data. This information is contained in the amplitude spectrum while the phase spectrum holds information about everything else. For example, temporal asymmetries manifest in subtle relationships of phases of waves of similar frequencies. By randomizing the phases, these aspects of the data are lost.

Also, in meteorological data, some frequencies relate to natural phenomena with a known phase, such as yearly and daily cycles. If no other efforts are made to filter out these frequencies before phase randomization (and introduce them later again), shifts in these phases cause unrealistic generated time series.

As randomizing the phases moves the marginal distribution closer to the normal distribution, it is advisable to use suitable transformations before and after applying it.

Contrary to other generation methods, correlations and the mean (represented by the "zero-frequency" amplitude) are not just reproduced asymptotically, but exactly.

Estimation

The amplitude spectrum can be obtained by applying the FT. The phases are the angles of the real and complex component of A_j.

Simulation

The approach for simulation follows naturally from the Wiener-Kinchin Theorem in conjunction with the assumptions (section 2.4):

1. The source time series is transformed via the Fast Fourier Transform (FFT) into the frequency domain:
$$F(u) = \int_{-\infty}^{\infty} f(t) e^{2\pi i t} dt \tag{2.54}$$
2. The amplitude spectrum $|F(u)|$ is calculated from the FFT parameters.
3. Phases are drawn from a uniform distribution: $\hat{\theta} \sim U(0, 2\pi)$
4. A new time series is obtained by inverse FT using the amplitude spectrum and the randomly generated phases:
$$\hat{F}(u) = |F(u)| \, e^{i\hat{\theta}} \tag{2.55}$$
$$\hat{f}(u) = \int_{-\infty}^{\infty} \hat{F}(u) e^{-2\pi i u} du \tag{2.56}$$

(Compare with equation (2.44) and equation (2.41))

3 VG: A Vector-Autoregressive Weather Generator

The Vector-Autoregressive Weather Generator (VG) was intended to be used to generate input time-series for climate impact analysis on lakes. This history motivated a simple design that is unlike most Weather Generators (WGs) in that it does not have an explicit rain occurrence model.

Schlabing et al. (2014) introduced an earlier version of VG. We used it as a tool to generate scenario data in a Monte Carlo study for climate impact assessment on Lake Constance, a large lake ecosystem. Eder (2013) also used an earlier version of VG for estimating the climate sensitivity of a large lake with a 3-dimensional hydrodynamic model. Kobler, Wüest, and Schmid (2018) used the weather generator for estimating the impact of pumped-storage operation on thermal structure and water quality under climatic change. Fenocchi et al. (2018) used it for estimating climate impacts on the mixing regime of Lake Maggiore. The latest publication presenting work based on VG is Gal, Gilboa, et al. (2020), in which we used VG to generate very specific climate scenarios for a small ensemble of 1-dimensional lake models simulating thermal conditions of a sub-tropical lake. The cases in which I was involved directly – Schlabing et al. (2014), Gal, Gilboa, et al. (2020), and Eder (2013) – are further described in chapter 6.

The most important difference to the published version of VG is that the version described here generates precipitation. Questions of water balance were not important in Schlabing et al. (2014) and thus precipitation was not generated. This enabled the initial omission of a rain occurrence model and the adoption of a single Vector-Autoregressive (VAR)-process as its core. Moving on-wards, as precipitation was intended to be added to broaden VGs scope of application, I realized that I could keep this simple structure if I could find a way to deal with zero-valued precipitation. Section 3.6 describes the idea and implementation.

Another difference of the version of VG that I describe to the ones used in previous studies is the use of phase randomization to generate residuals for use during simulation. This gives the VAR process a longer memory and reduces the number of fitted parameters.

The existence of a single VAR model makes the problem of propagating the change in one variable to others tractable using methods from Linear Algebra. Hence, there is no need for external information on the changes in precipitation statistics following temperature changes. If and how well this approach works will be analyzed in section 3.9.

The structure of this chapter is as follows. A motivation section shows the transition from pure methods to a WG while briefly explaining why other WGs fall short in meeting the stated aims. After a section providing an overview of VG, the sections following are descriptions of specific methods. The chapter ends with a summary.

3.1 Motivation

While it is possible to generate multivariate time-series with the methods described in chapter 2, these methods alone are arguably not sufficient enough to perform the task of a WG. The following points are reasons for adapting the methods from chapter 2:

- **Assumptions** necessary to apply the method to the given data might not be met (e.g. marginal distributions, distribution of residuals).
- Generating **scenarios** requires a change in the otherwise desired match of key statistics.
- Demands from lake modellers to generate "balanced 'what-if'-style scenarios", meaning that a change in one guiding variable is accompanied by equivalent changes in the other variables.
- Introducing periodic dependence throughout the year without relying on seasonally switching models. This would increase the number of fitted parameters and produce sudden regime shifts, which are undesirable in particular if the timing in impact models is to be studied (section 6.2.1 summarizes such an example).

WGs that rely on rain occurrence modelling as a first step are hard to adapt to the style of scenarios with a guiding variable that is not precipitation. This includes the widely used Richardson and Racsko approaches. Essentially, if precipitation is modelled as controlling the rest of the simulated variables, the direction of dependence is difficult to reverse directly. As an example, it is easier to change the transition probability of dry- to wet state and observe a change in temperature than the other way around in these models. Further, if relying on only one variable in scenario studies, it is arguably better to use air temperature than precipitation, as the confidence in its increase in climate projections is higher (Pachauri et al., 2014, p. 59 f). The same reason excludes approaches based on non-homogeneous hidden Markov models, where a hidden state variable controls precipitation and other variables. Resampling methods can be adapted to generate such scenarios, but their output becomes more limited when introducing such a scenario bias (see chapter 5).

3.2 Overview over the Weather Generator VG

Figure 3.1 gives an overview of the structure of VG. The scheme reveals the trans-Gaussian nature of the WG. As meteorological variables seldom follow normal distributions, they are converted to standard-normal, so that the corresponding assumption (see chapter 2) of the VAR process is not violated. The transformation makes a re-transformation after VAR generation necessary, for which the distributions, that were used for the transformation, are leveraged again.

All parameters except the autoregressive order p in VG are smoothly changing over the year. This includes the distribution parameters, the VAR-parameters and the parameters involved in estimating dryness probability.

The steps VG executes are as follows:

1. Annually changing parametric distributions or kernel density estimation (KDE) are fitted on all variables.
2. Using the fitted distributions, all marginals are transformed to standard-normal.
3. Dryness probability during dry spells is estimated with information from non-precipitation variables. That probability is transformed to its equivalent quantile in standard-normal.
4. A VAR-process is fitted on the transformed data.
5. A time series is generated by phase-randomizing the observed residuals (see section 2.4 for a description of phase randomization) and the recursive equation of the VAR process (equation (2.2)) is applied. Alternatively the residuals can be generated by drawing random vectors from a multivariate normal distribution fitted on residuals. If a scenario is to be run, the process is disturbed by non-zero values in the vector m before applying equation (2.2).
6. The generated time series is back-transformed using the fitted distributions from step 1.
7. Hourly information is added.

3.2.1 Observational Data

Wherever simulation results are presented in the present chapter, they are based on a data set of hourly measurements for the reference period of 1980 – 2000 from the Konstanz station in southern Germany by German Meteorological Service (DWD). Table 3.1 gives an overview.

Table 3.1: Variables and their distribution families used for transformation to standard-normal.

Variable	Symbol	Distribution
Air temperature	θ	Normal
Precipitation	R	Kumaraswamy with KDE for upper tail
Short-wave radiation	Q_{sw}	KDE
Incident long-wave radiation	$Q_{\text{lw(in.)}}$	Normal
Relative humidity	ϕ	Truncated Normal
Eastward wind speed	u	KDE
Northward wind speed	v	KDE

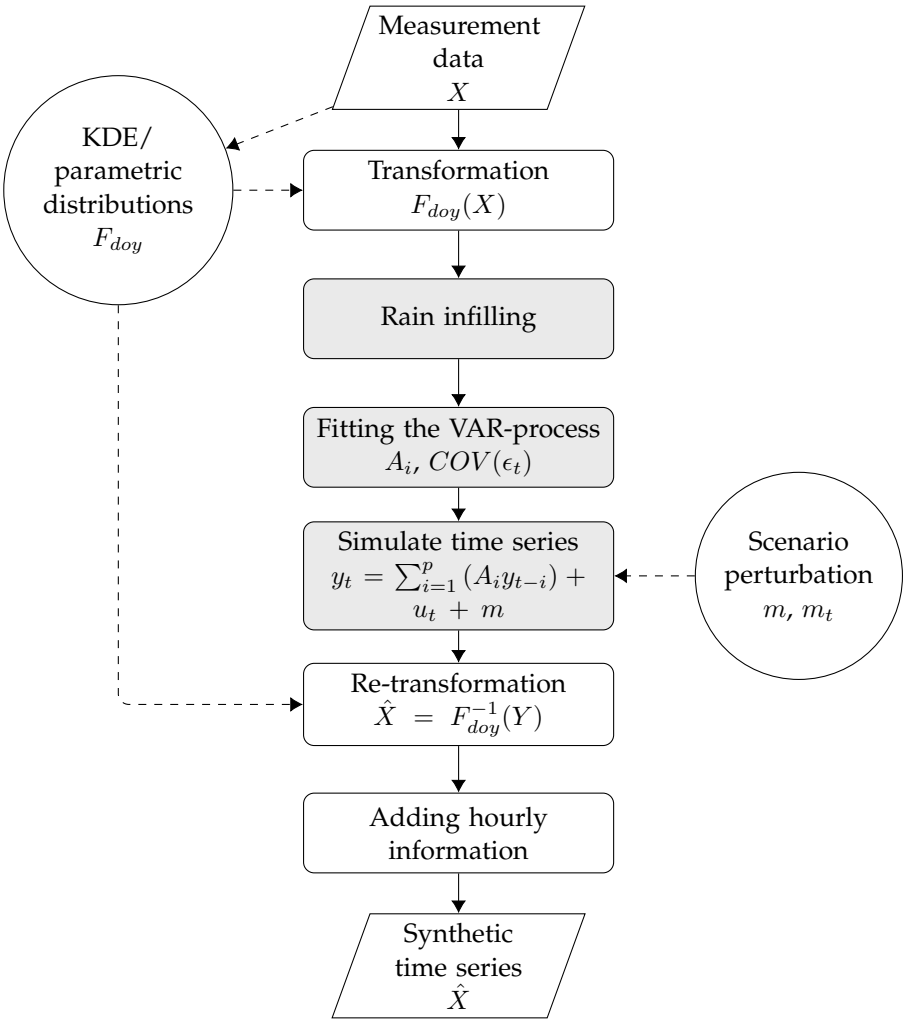


Figure 3.1: Structure of the WG VG. Boxes with grey background refer to computations in the stationary, standard-normal transformed domain.

3.3 Transformation by Quantile-Mapping

Not only do meteorological variables seldom follow normal distributions, their statistical moments change throughout the year. VG employs a VAR process which assumes and generates time series following multivariate normal distributions. This requires suitable transformations before fitting a VAR-process and re-transformations after VAR-based simulation.

Quantile-mapping (also called quantile-quantile transformation) exchanges values of one distribution with that of another distribution by matching the non-exceedance probability of these two distributions. Mathematically, this can be expressed as

$$Z = F_Z^{-1}\left(F_X(X)\right) \tag{3.1}$$

where Z – transformed variable
F_Z^{-1} – inverse cumulative distribution function (*cdf*) of the transformed variable
F_X – *cdf* of the untransformed variable X
X – untransformed variable

If the target distribution is the normal distribution, this transformation is called the normal-score transformation

$$Z = \Phi^{-1}\left(F_X(X)\right) \tag{3.2}$$

where Φ^{-1}– inverse *cdf* of the standard-normal distribution

While it is possible to estimate $F_X(X)$ with the empirical relative ranks of X, using the empirical distribution for re-transformation would limit the output of the WG. To avoid this, the empirical distribution of X is abstracted by fitting a parametric or kernel-based distribution F_X to X. Z then approximately resembles a normal marginal distribution depending on how well F_X has been fitted.

In order to remove annual cycles in the marginals, I did not assume F_X to be stationary, but a function of the *doy*. Two general approaches to achieve *doy*-specific *cdf*s ($F_{\mathrm{doy},X}$) are taken in this work: describing the annual cycle of distribution parameters with the help of the Fourier transform (section 3.3.1) and approximating $F_{\mathrm{doy},X}$ with a *doy*-specific numerically integrated KDE (section 3.3.2). Both methods aim to provide smooth changes of the distribution throughout the year while limiting the number of needed parameters.

3.3.1 Annual parametric Distributions

In a first step, a vector of distribution parameters θ_{doy} is obtained by fitting a distribution to the data of each *doy* separately. In order to have a reasonable number of observations for each fitting procedure and to achieve smoothly changing parameters, the data of the near *doy*-neighborhood is included. Note that *doy*-neighborhood does not generally imply neighborhood in the time series. For these distributional aspects it is assumed that values

on, say 1st of January have a very similar distribution as those from the 24th of December. As an example, for the Constance data-set (see section 6.2), which contains 21 years of data, I chose this neighborhood to consist of 30 *doy*s. This means, that the estimation of each parameter set originated from a fit to 21 * 30 = 630 data points.

The result of this operation are 365 distribution parameter sets θ_{doy} for each variable. Due to the fact that there is an intrinsic periodicity in meteorological variables and the inclusion of *doy*-neighboring values, each *doy*-series shows annual patterns. I captured these patterns with the help of a Fast Fourier Transform (FFT) based approximation and thus effectively reduced the number of actual used parameters. Equation (3.3) gives the resulting estimation for each parameter.

$$p_{j,\text{doy}} = \frac{a_{j,0}}{2} + \sum_{n=1}^{N} \left[a_{j,n} \cos\left(n \frac{2\pi \text{doy}}{365.25} \right) + b_{j,n} \sin\left(n \frac{2\pi \text{doy}}{365.25} \right) \right] \tag{3.3}$$

where $j \in \{\mathbb{N} | j \leq m\}$ – with m as the number of distribution parameters
$p_{j,\text{doy}}$ – j^{th} distribution parameter for a given *doy*
N – number of harmonics used to approximate $P_{j,\text{doy}}$
$a_{j,n}$, $b_{j,n}$ – n^{th} Fourier coefficients, amplitudes of n^{th} harmonic

N in equation (3.3) controls the number of effective parameters for distribution fitting and the smoothness of the series $p_{j,\text{doy}}$. Throughout this chapter, N is set to 4.

The estimated distribution for each variable is then

$$\hat{F}_{\text{doy}}(X) = F_X\left(X, p_{1,\text{doy}}, \ldots, p_{L,\text{doy}}\right) = F_X(X, \theta_{\text{doy}}) \tag{3.4}$$

where L – number of distribution parameters
θ_{doy} – fitted parameter set for a given *doy*

3.3.2 Annual Kernel Density Estimation

Not all measured variables exhibit a marginal that is easily fit with parametric distribution families. An alternative are KDEs which do not require the assumption of an underlying parametric distribution. As KDE is a *density* estimation technique, arriving at non-exceedance probabilities needed for quantile mapping requires a numerical integration.

I used the following formulation of KDE, which involves the *doy* as a second dimension in order to have a smooth transition of the distribution estimate over the year:

$$\hat{f}_{\text{doy}}(x) = \frac{\sum^{i \in \left\{ |\text{doy}_x - \text{doy}_{x_i}| < 15 \right\}} K_x\left(\frac{x - x_i}{h_{\text{doy}}} \right) K_{\text{doy}}\left(\frac{\text{doy}_x - \text{doy}_{x_i}}{15} \right)}{h_{\text{doy}} \# \left\{ |\text{doy}_x - \text{doy}_{x_i}| < 15 \right\}} \tag{3.5}$$

where $\hat{f}_{\text{doy}}$ – kernel density at x for the *doy*
K_x – kernel for the x dimension
K_{doy} – kernel for the *doy* dimension
h_{doy} – *doy*-specific kernel bandwidth

As can be inferred from equation (3.5), the kernel bandwidth is fixed to 15 days in the temporal dimension. The bandwidth in the data dimension h_{doy} is obtained by maximizing the likelihood using the leave-one-out method (Duin, 1976). This bandwidth is found by using only the data in the window defined by $|\mathrm{doy}_x - \mathrm{doy}_{x_i}| < 15$ and therefore *doy*-dependent.

3.3.3 Mixed Distribution for Precipitation

For finding a good fit for precipitation, I used a Kumaraswamy distribution (Kumaraswamy, 1980) for the bulk and log-Normal kernels for the upper part of the distribution. The Kumaraswamy distribution is a variant of the Beta distribution without any transcendental function (like the Gamma function in the case of the Beta distribution). It shares with the Beta distribution the existence of a lower and an upper bound and two parameters to tune its shape. Log-Normal kernels have the advantage of being able to mimic right-skewed distributions. In contrast to Normal kernels, they also imply a higher probability of generating values greater than historical ones, which is of importance in the context of scenario generation. The quantile which defines the transition from parametric distribution to KDE is determined by maximum likelihood and changes throughout the year. As a whole this method of modelling the distribution of precipitation is flexible, yet involves more free parameters than using a non-mixed distribution.

3.4 Annually Changing Dependence

Periodic changes in the dependency structure can be captured by fitting VAR-parameters for each *doy* using a moving window. The resulting time series of parameters are then again smoothed with the help of harmonics obtained by Fourier transformation.

This is implemented by masking values as non-existent and using a slightly modified VAR-Estimator for data with gaps (c.f. equation (2.9) on page 9). Columns in Y and rows in Z that contain at least one masked entry are deleted. This reduces the size of available data which has to be reflected in the estimator of the error covariance matrix $\hat{\Sigma}_u$ (see equation (2.10) on page 9).

$$\hat{\Sigma}_u = \frac{T}{l}\frac{1}{T-pk-1}\left(YY' - YZ'(ZZ')^{-1}ZY'\right) \tag{3.6}$$

where l – number of masked time steps

3.5 Phase-Randomized Residuals

Simulating time series with a VAR-process requires synthetic residuals. Usually these are drawn from a multivariate normal distribution fitted on the residuals obtained from the observations after estimating the VAR parameters. The residuals can be found by applying a rearranged version of equation (2.2) for all time-steps:

$$\mathbf{u}_t = \mathbf{y}_t - \sum_{i=1}^{p} (A_i \mathbf{y_{t-i}}) - \mathbf{m} \tag{3.7}$$

$$= \mathbf{y}_t - \hat{\mathbf{y}}_t \tag{3.8}$$

where $\mathbf{y}_t$ – vector of observed values
$\hat{\mathbf{y}}_t$ – vector of predicted values

I opted for a different approach by phase-randomizing the residuals (see section 2.4 for phase randomization). I used the same change of phases for each variable's residuals in order to maintain not only the autocorrelation of the residuals but also the cross-correlation. This is important because the covariance matrix of the residuals usually has non-zero values on the off-diagonal elements. Over the traditional method of using serially independent gaussian residuals, this has the advantage of potentially reproducing longer-term auto- and cross-correlations that are not removed by VAR. VAR-processes fitted with parameter parsimony in mind usually do not account for long-term memory – the VAR processes in this study all have an autoregressive order of $p \leq 3$ days.

The autocorrelation coefficients of the residuals are generally low, but are amplified by generating a VAR time series. Figure 3.2 shows the empirical distribution of autocorrelation coefficients of the residuals, phase-randomized residuals and serially independently drawn gaussian vectors (which is the method for VAR-simulation described in section 2.1). It is apparent that the distributions shown are not very different for most variables. Where differences exist, however, phase randomization is able to maintain the correlation of the observed values.

Despite the small differences in the distribution of the autocorrelation coefficients of the residuals, the weakly correlated residuals make a difference after applying the VAR-process equation (2.2) on them. Figure 3.3 shows empirical *cdf*s of simulated time series in the standard-normal domain. Using phase-randomized residuals raises the distribution of simulated autocorrelations very close to the autocorrelations of the transformed observations.

With phase randomization being able to maintain all linear correlations in a multivariate time series and without introducing additional free parameters, it might seems advisable to not use autoregressive models at all. This can be a good idea if no further changes to statistical properties of the time series are to be made. As will be shown in section 3.8, VAR processes offer a way to generate "balanced scenarios", that cannot be produced by phase randomization alone.

Phase randomization also keeps the mean constant from realization to realization. This can be an advantage when only one realization is generated, but when an ensemble of many realizations is generated, such a behaviour is undesirable, as the mean is a naturally changing statistic. In order to achieve an abstraction of means in VG ensembles, a constant is drawn from a normal distribution with zero mean and standard deviation of 0.25 and added to the primary variable before calculating the dependent changes to the other variables as described in the later section section 3.8. This ensures that the variation in mean spread to the rest of the generated variables, keeping the realizations "balanced".

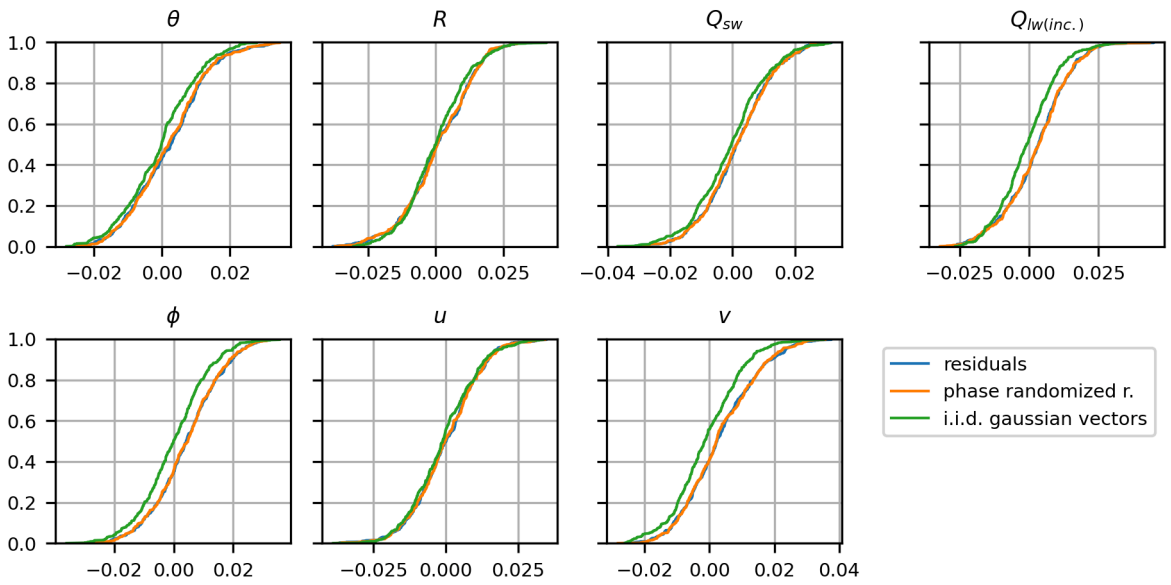


Figure 3.2: *cdf*s of autocorrelation coefficients of residuals with lags up to 365 days. Where the blue line (residuals) cannot be seen, it coincides with the orange one (phase-randomized residuals). The green line refers to serially independently drawn Gaussian vectors as described in section 2.1.

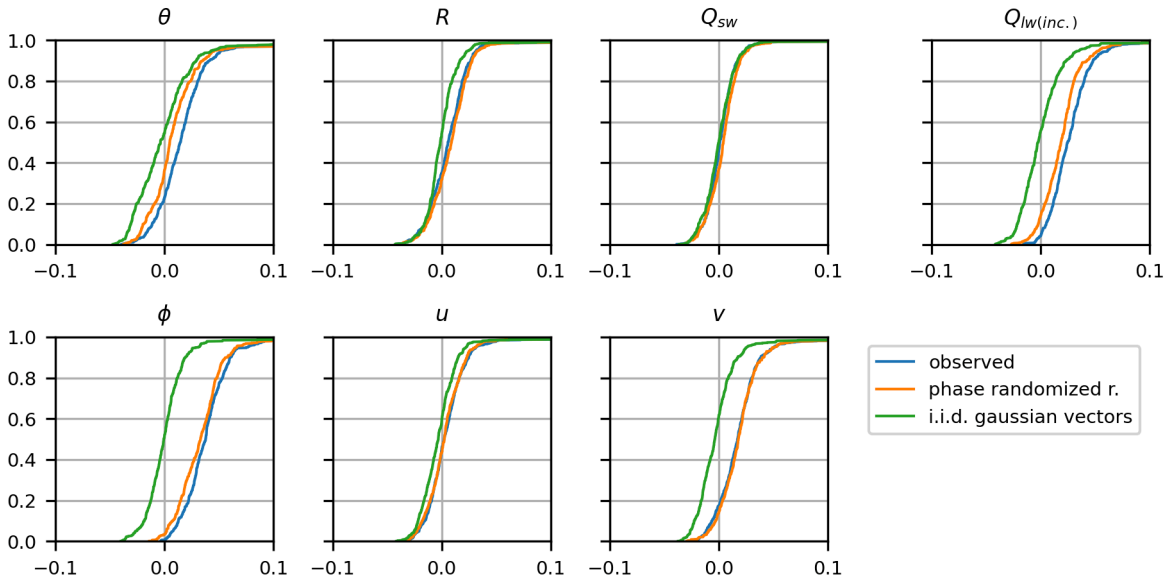


Figure 3.3: *cdf*s of autocorrelation coefficients with lags up to 365 days in the standard-normal domain. The blue line refers to observations, orange to simulation using phase randomized residuals and green to simulation using serially independently drawn Gaussian vectors as described section 2.1. The x-axis was fixed to [-0.1, 0.1] to highlight differences in the middle of the distribution.

3.6 Implicit Generation of Precipitation by Estimating Dryness Probability

The aim in this study of using any arbitrary variable as a guiding variable necessitates that precipitation is not given a special role preceding the auto-regressive time-series generation. This stands in contrast to established WGs, which are, beginning with the C. W. C. Richardson (1981) and continuing with the Racsko, Szeidl, and M. Semenov (1991) model, usually implemented in a doubly-parameterized fashion: one set of parameters apply to wet and one set of parameters to dry spells. It is generally agreed that the presence of rain has an influence on many non-precipitation variables (Daniel S. Wilks and R. L. Wilby, 1999). Vice versa, this effect is exploited in VG in order to estimate a probability of dryness *during dry spells*. p_0, the relative frequency of no rain occurring, would be a very crude estimate of this probability, given the possible values of non-rain variables. For a day with a high solar radiation, low humidity and a wind from the north-east, p_0 would be, as an average dryness probability, too low, as these conditions indicate a very "dry" situation[1]. Likewise, days that have a more wet appearance in their non-rain variables, should be assigned a lower dryness probability than p_0. These intuitive arguments motivate the approach described in the following, which consists of deriving a degree of dryness from the values of non-precipitation variables. The such estimated probability of dryness is scaled to the interval $[0, p_0]$, transformed to standard-normal and inserted in the gaps between rain events in the transformed precipitation time series. At this point, there exists one multivariate time series with approximate standard-normal margins, suitable to be modelled by a VAR process. This has the advantageous side-effect of a reduced number of free parameters, as only one VAR process, regardless of rain state, can be fitted.

I investigated two methods for estimating dryness probability. The first uses a simple multiple linear regression-based approach, while the second involves an euclidean distance measure between current dry state to typical wet state values.

3.6.1 Estimating Dryness Probability by Multiple Linear Regression

For the regression-based approach, all non-precipitation variables' marginals are transformed to stationary standard-normal, resulting in the matrix X (the details of the transformation are described in section 3.3). Precipitation values exceeding a threshold (i.e. $\mathbf{y}^{+}$[2] with time-steps $t^+ \equiv \{t \in \mathbb{N} | t \leq T \wedge \mathbf{y}_t \geq c\}$ and $t^- \equiv \{t \in \mathbb{N} | t \leq T \wedge t \notin t^+\}$) are mapped to their quantiles of the standard-normal distribution, taking the probability of dryness into account (see panel I and II in figure 3.4), resulting in $\mathbf{y}^+$. Then, the vector of parameters $\mathbf{b}$ of the regression is obtained by minimizing ϵ in the following equation:

$$\mathbf{y}^+ = X^+\mathbf{b} + \epsilon \tag{3.9}$$

[1] This is specific for at least southern Germany.

[2] In the following the "$^+$"-superscript will refer to rainy and the "$^-$" superscript to dry conditions.

Estimated values of "dry precipitation" $\hat{\mathbf{y}}^-$ are then extrapolated using the values of non-precipitation values during dry conditions:

$$\hat{\mathbf{y}}^- = X^- \mathbf{b} \tag{3.10}$$

The relative ranks of $\hat{\mathbf{y}}^-$ within the full time-series $\mathbf{y}$ can be interpreted as the dryness probability.

As $\hat{y}^-$ is not guaranteed to match the lower part of the standard-normal *cdf* (depicted in grey in panel II in figure 3.4), it is quantile-transformed to that part of the target distribution.

As the relation between precipitation and the non-precipitation variables changes throughout the year, $\mathbf{b}$ is evaluated dependent on the *doy* with a rolling window of width of 30 *doy*s. The resulting series of $\mathbf{b}_{\text{doy}}$ is smoothed by replacing it with the sum of its 2 longest harmonics.

3.6.2 Estimating Dryness Probability by Euclidean Distance

This method also starts with the matrix X, consisting of variables with standard-normal marginals, obtained by transforming a measured time series using the methods described in section 3.3. In the case of rain, the target distribution is truncated with the dryness probability as lower truncation point. The mean vector of non-rain variables $\overline{\mathbf{x}}^+$ during rain events is used as a reference point for wet conditions:

$$\overline{\mathbf{x}}_i^+ = \frac{1}{n_{t^+}} \sum^{t \in t^+} \mathbf{x}_{i,t} \tag{3.11}$$

where
- $\mathbf{x}_{i,t}$ – value of non-rain-variable i at time step t in the standard-normal domain
- $t^+ \in \{\mathbb{N} | t \leq T \wedge R_t \geq c\}$ – rainy time steps, i.e. time steps where precipitation is at least as high as a threshold c
- R_t – rain amount on time step t
- n_{t^+} – number of rainy time steps in the record

For every dry time step t^- the squared euclidean distance of the non-rain variables to that reference point $\overline{\mathbf{x}}^+$ is calculated:

$$d_{t^-} = \sum_{i=1}^{k} \left(x_{i,t^-} - \overline{\mathbf{x}}_i^+ \right)^2 \tag{3.12}$$

where
- $\overline{\mathbf{x}}_i^+$ – mean of variable i during wet conditions (see equation (3.11))
- $t^- \in \{\mathbb{N} | t \leq T \wedge R_t < c\}$ – dry time steps
- k – number of non-rain variables

As all $\mathbf{x}_i$ were transformed to standard-normal, there is no need to weigh the terms $x_{i,t} - \overline{\mathbf{x}}_i^+$ as they are on the same scale. This distance is then used to derive the probability of

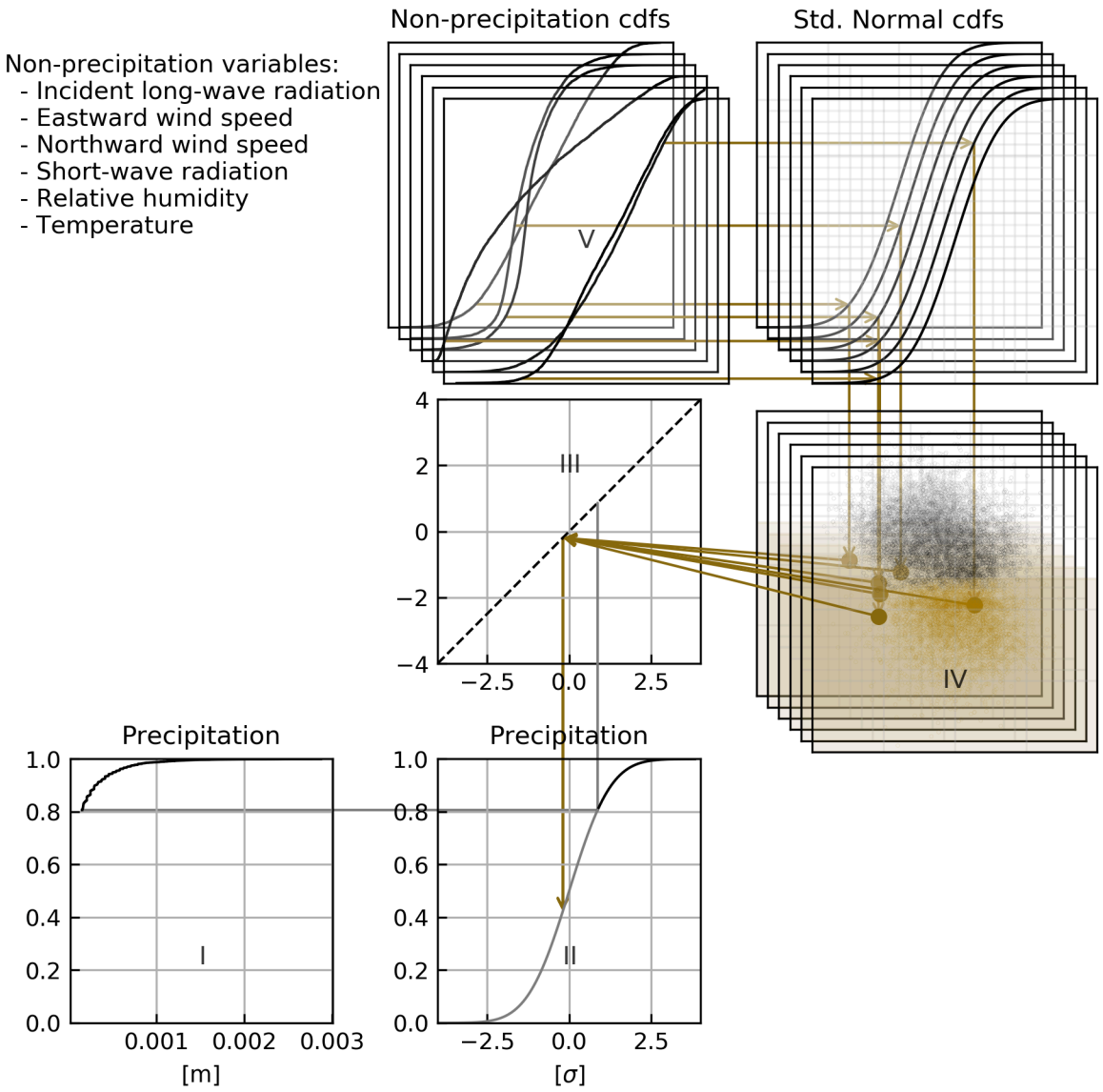


Figure 3.4: Schematic for precipitation infilling via multiple linear regression of standard-normal transformed variables. Precipitation values in I are matched with its upper standard normal quantile in II, resulting in the y-values of the black dots in IV. The x-values of the dots in IV are standard-normal transformed non-precipitation values from V. A multiple linear regression provides then the parameters to extrapolate into the left leg of standard-normal precipitation in II using standard-normal-transformed values of non-precipitation variables from V.

dryness. This probability has a corresponding value within the lower part of the truncated normal distribution fitted to the precipitation amounts.

Combining the lower truncated distribution of rain amounts and the upper truncated distribution of dry distances, an infilled time series without gaps and with a continuous standard-normal marginal distribution is achieved as follows:

$$\mathbf{y}_t = \begin{cases} \Phi^{-1}\left(p_0 + p_1 F_R(R_t)\right) & \text{if } t \in t^+ \\ \Phi^{-1}\left(p_0(1 - F_d(d_t))\right) & \text{if } t \in t^- \end{cases} \tag{3.13}$$

where $\mathbf{y}_t$ – standard-normal transformed rain without gaps at time-step t
Φ^{-1} – inverse *cdf* of the standard-normal distribution
p_0 – relative frequency of dry time steps
p_1 – relative frequency of rainy time steps
F_R – *cdf* of rain amount
F_d – *cdf* of dry distances (approximated by relative ranks)

The distribution of the dry distances (F_d) is approximated by relative ranks. Equation (3.13) yields low values for $\mathbf{y}_t$ if the distance d_t is high, so weather that is unlike wet conditions has a high dryness probability.

In order to account for annual periodicity in the relation between precipitation and the other variables throughout the year, the reference point $\overline{x}^+$ is evaluated per *doy* with a rolling data window of the width of 30 *doy*s. To further smooth this series of parameters, it is replaced by the sum of its 2 longest harmonics.

Remarks on Dryness Probability Estimation Differences

While both methods use non-precipitation variables to estimate a dryness probability, there are notable differences. The regression introduces new parameters which have to be fitted, which somewhat weakens the claim to have fewer parameters by not having a model parameter set for each of wet and dry state. Further, equation (3.10) represents a "best guess" value, as it is drops the residual ϵ of the regression in the estimation. The distance method needs just the wet-state mean vector $\overline{\mathbf{x}}^+$ as parameters and separates no residual, hence is likely to contain more noise.

3.6.3 Separation of Distributions Based on Precipitation State

While exploiting the fact that precipitation occurrence has an effect on non-precipitation variables, it stands to test whether VG is able to produce separated distributions of non-precipitation variables, despite the fact that it does not have different model parameters for wet and dry state. For the following short analysis, I generated a 100-year time series with each dryness probability estimation method.

Figure 3.5 reveals that precipitation state indeed has an impact on non-precipitation variables as all variables show markedly distinct conditional histograms for both observed and

simulated values. Both methods show good agreement in the frequencies regarding dry states, but slightly less good fit for the wet state. Notable differences between simulated and observed data occur in the upper tail of long-wave radiation during wet spells, where both methods yield an overestimation. The wet-conditional distributions of the wind-speed components are shifted towards higher values in the simulations.

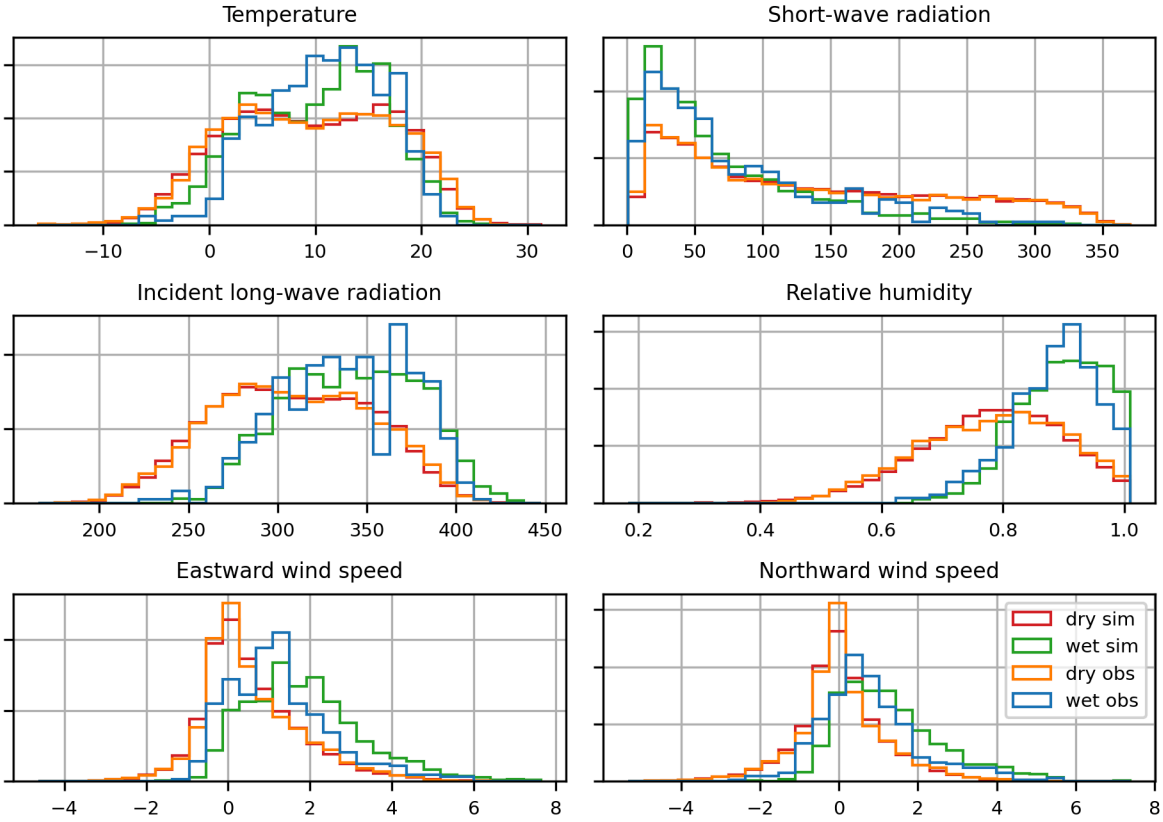
When interpreting the difference in quality of fit regarding dry and wet state, it is illuminating to pay regard to the population size of both parts of the data sets. In absolute numbers the deviation is similar, but normalizing to relative frequencies amplifies the differences in the smaller wet data set compared to the dry one. This result is not surprising, given that neither during the fitting of marginals nor during fitting the VAR-process, a higher weight is given to wet state. Thus the overall fit is optimized and the smaller subset of the data exhibits higher relative errors.

3.6.4 General Precipitation Fit and Dispersion Behaviour

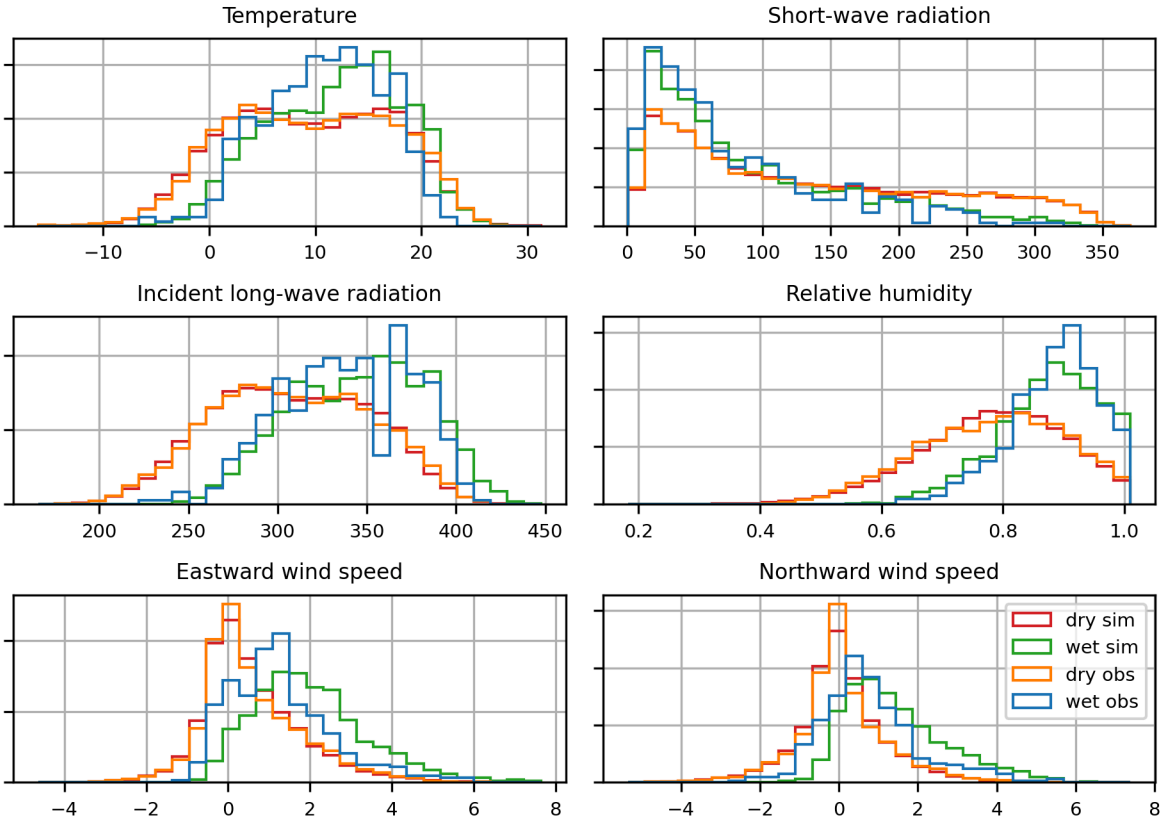
Figures 3.6 and 3.7 shows exceedance probabilities for intensity, dry- and wet spell lengths of the source data and an ensemble of 500 realizations. The median of the exceedance probability of the realizations for intensity regarding the distance as well as the regression method follows the observed intensities very precisely. The very similar shape of medians and ranges indicates that the marginal of precipitation does not depend on the dryness probability estimation method. The frequency of longer dry spells is underestimated more for the regression method, while the distance method's median follows the observed values more closely. In the frequency of dry spells, the two methods produce very similar, good results. Even the VAR process using i.i.d. Gaussian vectors show satisfactory results. Differences appear most visibly in the intensity using the regression method. The VAR process even has a better representation of medium-length dry spells compared to VG for the regression- and to a lesser extent the distance method.

Inter-annual variability in terms of monthly precipitation sums is shown in figure 3.8 based on the same ensemble with both dryness probability estimation methods. The regression method gives a better representations of large monthly sums in the months July to September than the distance method. Other than that, the observed monthly sums give the impression of being well abstracted by VG. The use of log-Normal kernels also show that values are allowed to be much higher than the observations.

For evaluating the scaling behaviour, it is important to make a distinction between the measurement and transformed domain as shown in figure 3.9 where the standard deviation is plotted over the aggregation length of precipitation sums. In the standard-normal domain, the inability of the VAR-process that uses i.i.d. Gaussian vectors to produce variability on longer time series is shown by having a markedly lower standard deviation than the observations. The variability on longer time scales is, however, maintained by using phase-randomized residuals as shown by blue and green line (for regression- and distance-method, respectively). The back-transformation has an influence on the scaling behaviour as well, as all methods now suffer from having a too high variability.

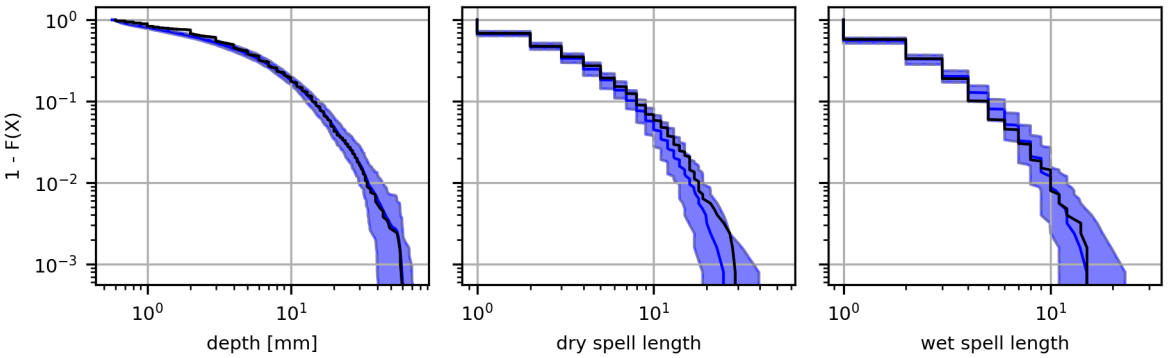


(a) Regression method

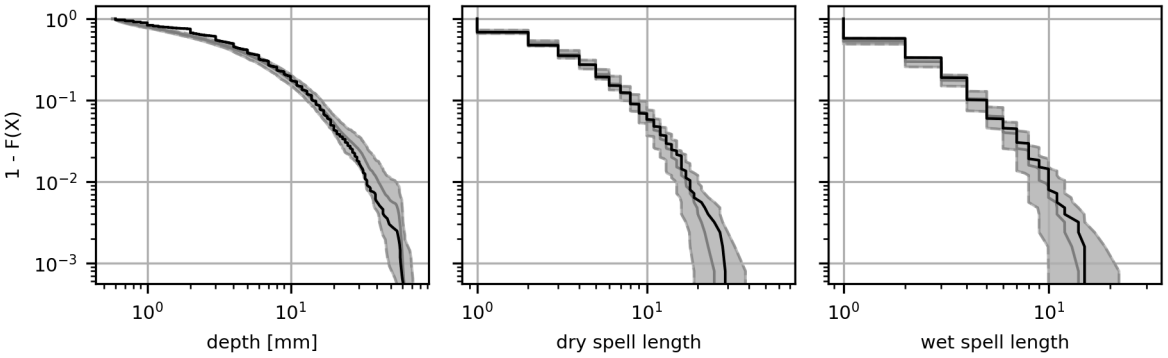


(b) Distance method

Figure 3.5: Histograms with normalized frequencies of non-precipitation variables conditional on precipitation occurrence.

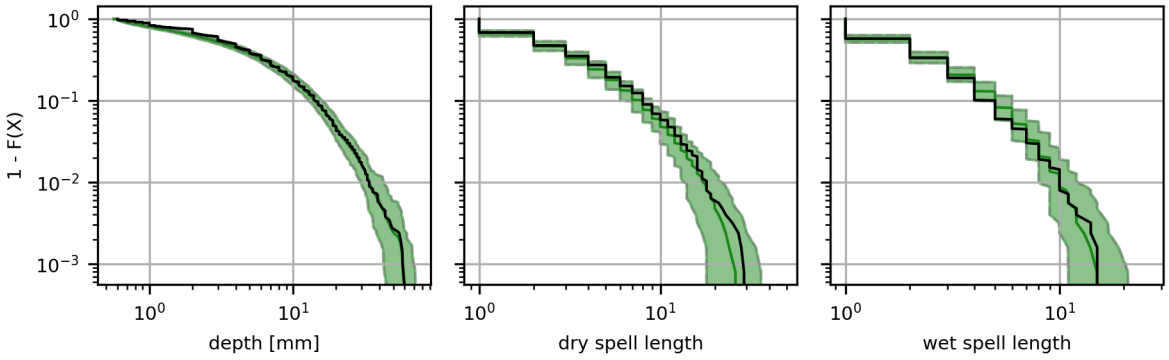


(a) Regression method with VAR generation using phase randomized residuals.

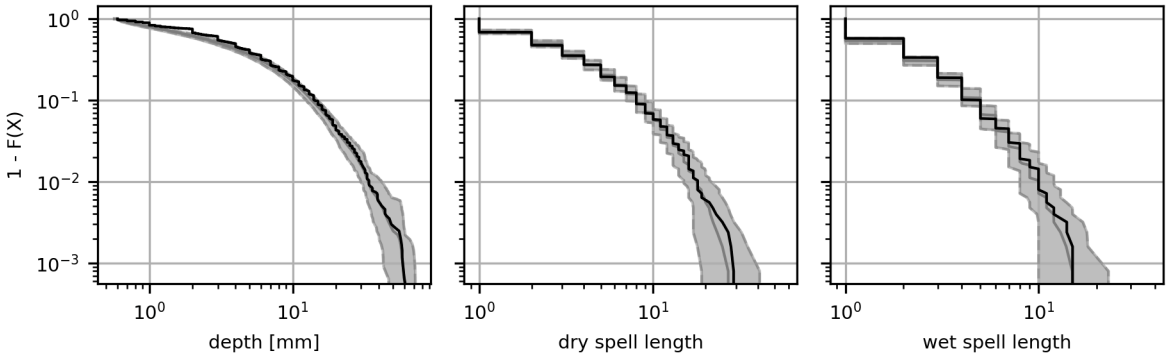


(b) Regression method with VAR generation using i.i.d. Gaussian vectors.

Figure 3.6: Precipitation exceedance probabilities for intensity, dry- and wet spell length using the regression method for dryness probability estimation. Thick lines shows medians. The shaded areas show the range of the ensemble. The black line shows observed values.



(a) Distance method with VAR generation using phase randomized residuals.



(b) Distance method with VAR generation using i.i.d. Gaussian vectors.

Figure 3.7: Precipitation exceedance probabilities for intensity, dry- and wet spell length using the distance method for dryness probability estimation. Thick lines shows medians. The shaded areas show the range of the ensemble. The black line shows observed values.

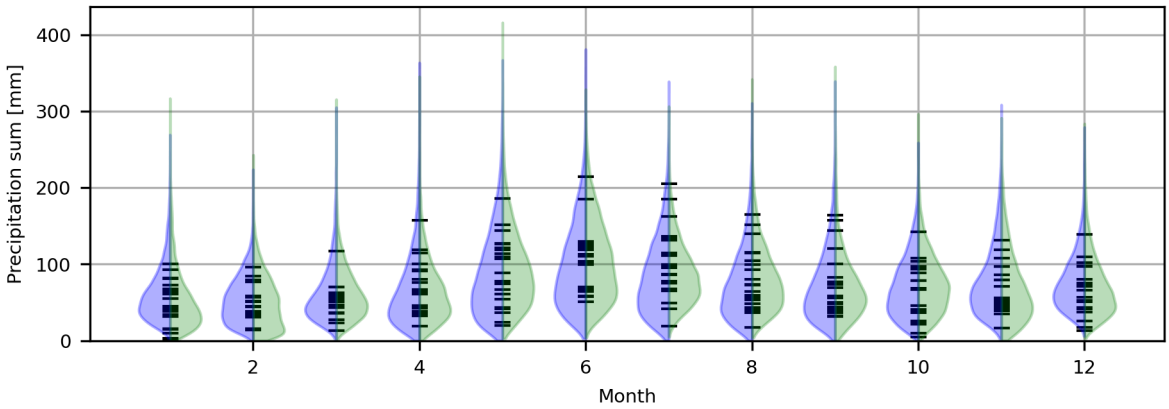


Figure 3.8: Violin plots of monthly precipitation sums. The black horizontal bars are observed values. Blue corresponds the regression, green to the distance simulation method.

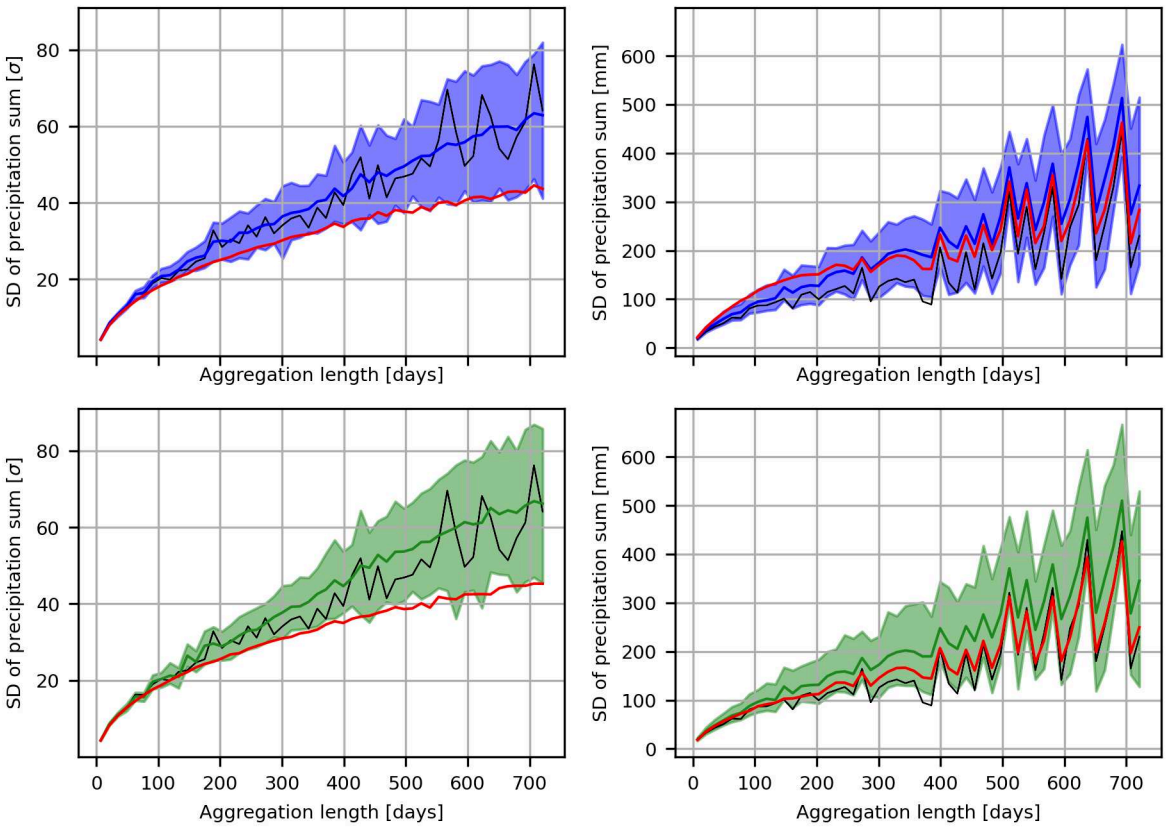


Figure 3.9: Standard-deviation of precipitation sums over aggregation lengths. X-axis starts at 7 days. Left panel: transformed. Right panel: measurement domain. Black line represents observations, blue regression and green distance method respectively. Red line is the median of the ensemble run with a VAR-process that uses i.i.d. Gaussian vectors.

3.6.5 General fit

Figure 3.10 shows the distributional fit of the ensemble for all variables. With the median of the ensemble being mostly on the x=y-line and all observed values within the range of the ensemble, the fit is generally good, but some tails show discrepancies. Upper quantiles of long-wave radiation are simulated lower than observed and lower quantile relative humidities are higher in the simulations. Other differences in the marginals between simulated and observed data based on the qq-plots seem minor.

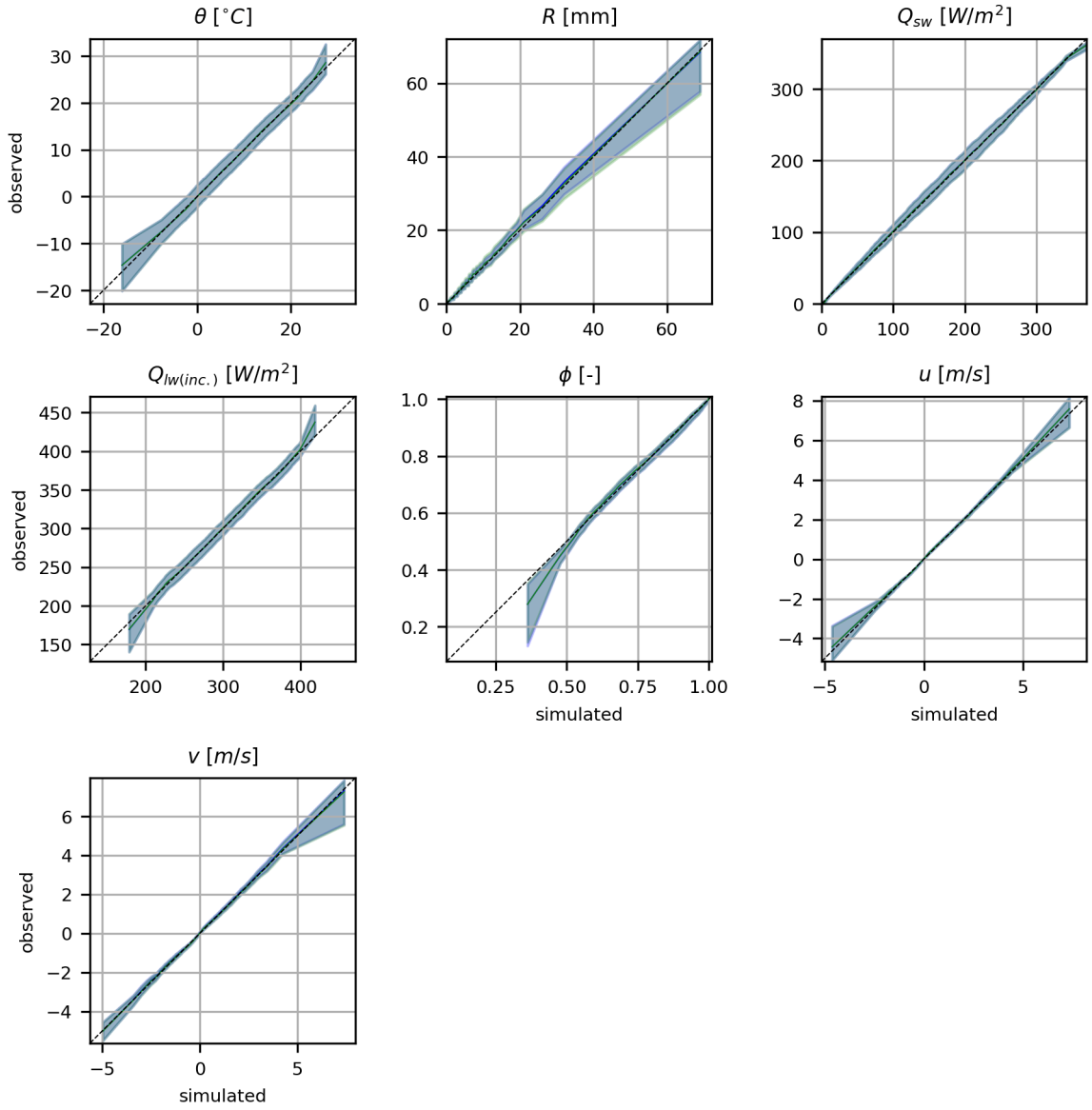


Figure 3.10: QQ-Plot of ensemble simulation versus observations. Blue region represents the range of the ensemble simulated with the regression method and blue for distance method. Thick lines are the median of the ensemble.

The representation of means is shown in figure 3.11. There exists a small overestimation in

precipitation as well as the wind speed components, however the differences to the observed means are small. The choice of dryness probability method has a negligible effect on most variables, with an exception of precipitation itself. With the distance method, the spread in means is higher. The spread is influenced by the dependence between temperature and the infilled precipitation time series, as the ensemble mean perturbations are only added to temperature before simulating a realization (see section 3.5).

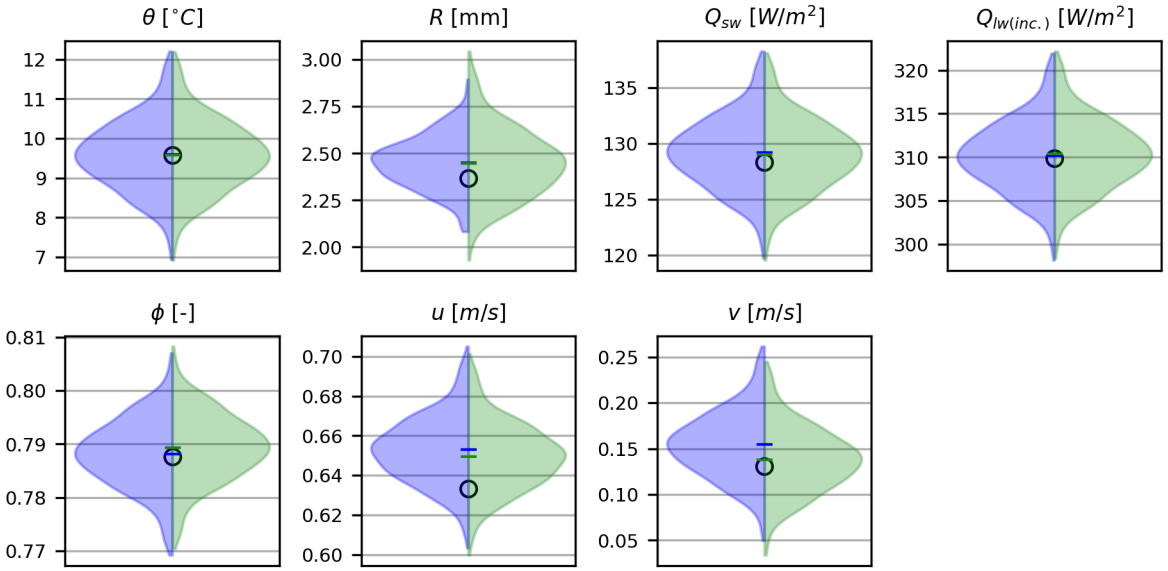


Figure 3.11: Violin plots of daily means of the ensemble simulation. Black circles show observed means over the whole time span, whereas horizontal bars show the median value of the ensemble. Colors represent the dryness probability method used: blue for regression and green for distance.

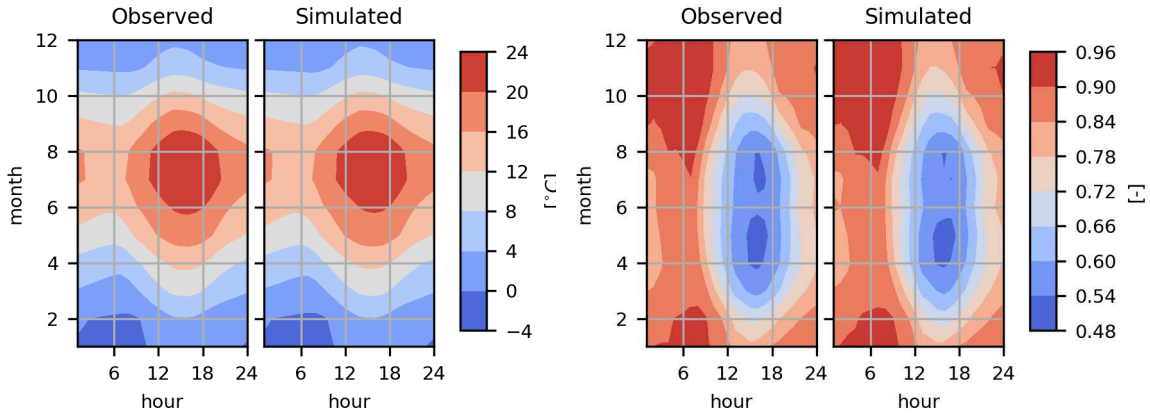
3.7 Adding Hourly Information

In order to arrive at hourly values, I implemented the following resampling scheme:

1. The hourly input data is aggregated to daily data.
2. Using linear interpolation, the aggregated daily input data is filled up to hourly time steps.
3. A pool of hourly 2-day chunks is formed by taking the difference between the hourly and the hourly interpolated data.
4. The simulated daily time series is interpolated to hourly values. This is the same method as in step 2.
5. Randomly selected chunks from the pool are added to the hourly interpolated simulated time series. The candidates from the pool are restricted by *doy*-distance in order to not deteriorate the annually changing (cross-) correlations.

For variables that follow a bounded distribution, chunks are not added, but applied as percentage differences, so that physically meaningless values such as negative radiation is avoided.

Figure 3.12 shows annually changing diurnal patterns of observed and generated data with the example of air temperature and relative humidity. As expected from a data-based method, the patterns are closely reproduced.



(a) Air temperature (b) Relative humidity

Figure 3.12: Annual cycles of diurnal patterns. Shown are hourly means.

This step is not to be confused with disaggregation which would generate higher-resolution data that aggregates to a given lower-resolution time series. Here, hourly information is added to daily simulated values, but it is not guaranteed that averaging over the hourly values yields the daily values.

3.8 Generation of Balanced Scenarios

With "balanced scenarios", I mean such simulations where a change in a guiding variable (typically temperature) is accompanied by changes in the other simulated variables according to the statistical dependencies between the guiding and the other simulated variables. As will be seen in the context of VG, these relationships are modelled as linear ones in the standard-normal transformed space.

The VAR-process is arguably simple, which allows for modifications that enable the generation of scenarios with changed statistics. The aim in the following subsections is not to show a path to downscale general circulation model (GCM) or regional climate model (RCM) output, but to describe how to generate time series that share some statistics with measured data while altering statistics that are defined by the impact modeller.

While the quantile-quantile transformation described in section 3.3 helps to deal with the problem of differing marginal distributions and the dependencies separately, it makes the generation of scenarios slightly less trivial. To change the mean of a variable within the

generated time series by a given value, it has to be known what the respective change in the standard-normal domain is. Further, in order to generate balanced scenarios, a perturbation is added to one variable *during generation*, so that this perturbation spreads to dependent variables via the parameters A_i to subsequent time steps in equation (2.2). Because of this spreading effect, achieving a predefined change in mean necessitates that the parameter matrices A_i are taken into account. For changing the means of the dependent variables, an explicit perturbation is also added to them individually. Together, these perturbations are meant to achieve statistically sound scenarios in which variables behave according to their linear dependence structure.

3.8.1 Simulation with Changed Mean

There are three necessary steps for generating balanced scenarios with a specified change in one guiding variable x_{prim}:

1. Transformation of the change in the guiding variable to standard normal: $\Delta x_{\text{prim}} \rightarrow \Delta y_{\text{prim}}$
2. Determining the appropriate change in dependent variables Δy_i
3. Obtaining the perturbation vector: $\Delta \mathbf{y} \rightarrow \mathbf{m}$

Transformation of the Change in the Guiding Variable

If the guiding variable is fitted with a normal distribution (such as temperature in section 3.9) the first step simplifies to dividing the change in the measurement domain by the variables' fitted standard-deviation. As the standard-deviation changes throughout the year, this transformed change also changes with the *doy*:

$$\Delta y_{\text{prim,doy}} = \frac{\Delta x_{\text{prim}}}{\sigma_{\text{prim,doy}}} \tag{3.14}$$

where
- prim – index of the primary/guiding variable
- $\Delta y_{\text{prim,doy}}$ – *doy*-specific change of the primary variable in the standard normal domain
- Δx_{prim} – selected change of the primary variable in the measurement domain
- $\sigma_{\text{prim,doy}}$ – fitted standard-deviation of the primary variable at *doy*

For the less common case in this work that the guiding variable is fitted with a non-normal distribution (see section 6.3 for an example), I implemented an approximating method. Δy_{prim} is then calculated as the matching normal quantile of the change in the measurement domain relative to the fitted median:

$$\Delta y_{\text{prim,doy}} = \Phi^{-1}\left(F_{\text{prim, doy}}(F^{-1}_{\text{prim,doy}}(0.5) + \Delta x_{\text{prim}})\right) \tag{3.15}$$

where $F_{\text{prim,doy}}$ – *doy*-specific fitted *cdf* of the primary variable.
$F^{-1}_{\text{prim,doy}}(0.5)$ – *doy*-specific fitted median of the distribution of the primary variable.

Determining the Change in Dependent Variables

Next, in order to propagate the change in the primary variable to all other variables, I opted for a simple linear approach:

$$\Delta y_i = \Delta y_{\text{prim}} \frac{\sigma_{y_{\text{prim}} y_i}}{\sigma^2_{y_{\text{prim}}}} \tag{3.16}$$

where $i \in \{\mathbb{N} | i \leq k \wedge i \neq \text{prim}\}$ – index of non-precipitation variable
k – number of variables
$\sigma_{y_{\text{prim}} y_i}$ – covariance between transformed primary variable and variable i
$\sigma^2_{y_{prim}}$ – variance of the transformed primary variable

This term is the same as the slope of a univariate linear regression.

Obtaining the Perturbation Vector

The vector $\Delta\mathbf{y}$ cannot be simply added to the recursive formulation of the VAR-process, because the parameters A_i amplify its influence to the next p time steps (see equation (2.2)). There is however, an identity connecting the term m to the mean of a stable VAR-process (Lütkepohl, 2006, p. 16):

$$\mu = \left(I_k - \sum_{i=1}^{p} A_i\right)^{-1} \mathbf{m} \tag{3.17}$$

where μ – $(k \times 1)$ vector of process means
I_k – $(k \times k)$ identity matrix
p – autoregressive order of the VAR-process
A_i – parameters of the VAR-process

This equation can be intuitively derived using the stability constraint (equation (2.5) on page 8), which states that the expected value μ is constant for every time step t ($\mu = E(\mathbf{y}_t)$). More specifically:

$$
\begin{aligned}
E(\mathbf{y}_{t-1}) &= E(\mathbf{y}_t) \\
\mu &= E\left(\sum_{i=1}^{p} A_i \mathbf{y}_{t-i} + \mathbf{m} + \mathbf{u}_t\right) \\
&= E\left(\sum_{i=1}^{p} A_i \mathbf{y}_{t-i}\right) + E(\mathbf{m}) + E(\mathbf{u}_t)
\end{aligned}
\tag{3.18}
$$

Because $E(\mathbf{m}) = \mathbf{m}$, $E(\mathbf{u}_t) = \mathbf{0}_k$ (see equation (2.3) on page 7) and $E(AX) = AE(X)$:

$$
\begin{aligned}
\mu &= \sum_{i=1}^{p} A_i E(\mathbf{y}_{t-i}) + \mathbf{m} \\
\mu &= \sum_{i=1}^{p} A_i \mu + \mathbf{m} \\
\left(I_k - \sum_{i=1}^{p} A_i\right) \mu &= \mathbf{m} \\
\mu &= \left(I_k - \sum_{i=1}^{p} A_i\right)^{-1} \mathbf{m}
\end{aligned}
\tag{3.19}
$$

This derivation is obviously only valid if the inverse $\left(I_k - \sum_{i=1}^{p} A_i\right)^{-1}$ exists, which it does if the VAR processes is stable (compare equation (2.4) on page 7).

Re-arranging equation (3.17) suggests the following estimator $\hat{\mathbf{m}}$ for $\mathbf{m}$:

$$
\hat{\mathbf{m}} = \left(I_k - \sum_{i=1}^{p} A_i\right) \Delta\mathbf{y} \tag{3.20}
$$

where $\Delta\mathbf{y}$ – $(k \times 1)$ vector of sample means

This estimator is useful for finding an $\hat{\mathbf{m}}$ that produces a time series with the desired means $\Delta\mathbf{y}$.

Equations (3.14), (3.16) and (3.20), together connect a supplied change in a primary variable in the measurement domain to the whole set of simulated variables in the transformed domain.

3.8.2 Simulation with Time-Varying Mean

$\mathbf{m}$ was previously assumed to be a constant $(k \times 1)$-vector. However it can also be applied as time-varying $\mathbf{m}_t$ with $M := (\mathbf{m}_0, \ldots, \mathbf{m}_T)$ ($(k \times T)$ matrix). Applying a time-dependent mean enables the following scenarios:

- applying a deterministic trend
- inducing variability on a larger-than-daily scale by stochastic variation of $\mathbf{m}_t$
- following a temperature signal from a RCM

3.9 Ability to Extrapolate

While it is convenient and necessary for freely defined scenarios to not rely on external variables, it is advisable to evaluate the methodology on how well it is able to project changes. In other words, can the change in one variable really be projected to the rest of the simulated variables? In order to do such a test, I fitted VG on the years 1984–1987 of the data set mentioned in section 3.2.1 and validated against the period 1992–1995 using the mean change in air temperature (1.304 °C) between these periods as the sole predictor for the rest of the simulated variables. For the simulated variables see table 3.1.

Figure 3.13 conveys the changes in means and the variability within a 500-member ensemble. Table A.2 in the appendix also lists changes in means. There are only small differences between simulated and observed means in the calibration period for most variables. The mean precipitation is, however, severely overestimated. This should be seen in context to the better fit when using the whole 21 years for fitting in figure 3.11. It seems to be an implementation issue regarding smaller numbers of years in the code that fits seasonally changing distribution parameters and performs KDE for values above a threshold that is found by maximum likelihood estimation (see section 3.3.3). Looking at the changes in observed means, the increase of 14.8% in temperature is accompanied by slight increases in precipitation (2.5%), incoming long-wave radiation (3.4%), relative humidity (2.0%) and short-wave radiation (1.5%), but the wind-speed components change drastically (northward: 11.6%, eastward -66.2%). Regarding the simulated means, most notable is the difference in projected precipitation depending on the dryness probability estimation method. With the regression method, the temperature increase is accompanied with a decrease in mean precipitation, while with the distance method precipitation increases. For the other variables the two dryness estimation methods do not make a difference in mean projection. Relative humidity (ϕ) and northward wind component (v) stand out insofar as the direction of their change in mean is misprojected. The change in mean short-wave radiation ($Q_{sw.}$) is overestimated with the true mean in the validation period being enclosed by both ensembles. The change in mean incident long-wave radiation ($Q_{lw(in.)}$) is underestimated, with the true mean still being inside of the range of the ensemble.

Despite the discrepancies in projected non-guidance-variable changes, the change in temperature is very precisely the one requested (observed change: 1.304 °C, simulated change: 1.306 °C for the regression and 1.306 °C for the distance method). This validates equations (3.14) and (3.20).

Arguably more important than changes in mean are changes in distributions as they convey more information than aggregate statistics. Before presenting the results for the distributional shifts, it is enlightening to regard how the distributions of the generated data agree

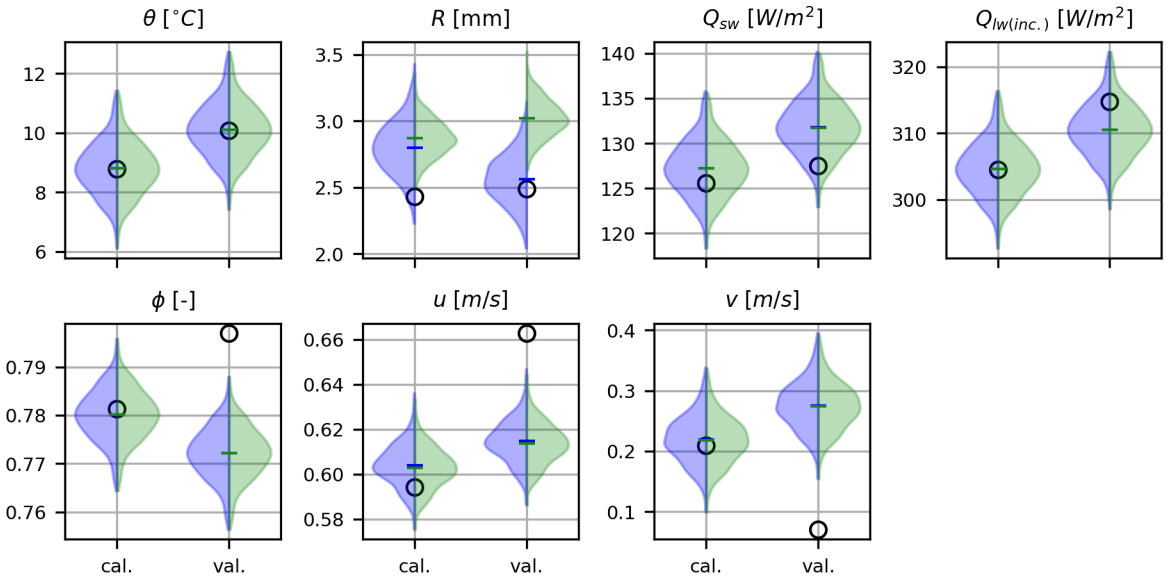


Figure 3.13: Violin plots of daily means in the extrapolation test. Densities shown are means of 500 realizations. Black circles correspond to observed values. Blue corresponds the regression, green to the distance simulation method.

with the observed data in the calibration period (see figure 3.14). Despite the differences in mean, the generated distributions agree well with those from the observed data set, with an exception of precipitation. The qq-plot for precipitation implies that the positive bias in the calibration period is due to an overestimation in the middle of the distribution, while the largest quantiles match those of the observations. It has to be mentioned again, that this unsatisfactory fit occurs only in this short calibration period. When using the whole 21-year data set, the quantiles show a much closer fit (see figure 3.10). In general, misrepresentations in the distributions occur more often in the extreme values. Affected extreme value misfit exist in: lower temperatures, lower relative humidity, high values in the wind speed components. Despite these deficiencies, the 1–99 % range envelopes all observed quantiles.

As figure 3.15 highlights, the shifts in distributions between calibration and validation periods are relatively weak. The black line matches quantiles between observed calibration and validation period and is shifted slightly to the right for temperature and longwave radiation. This implies a shift of all quantiles towards higher values from calibration to validation period in these variables. Relative humidity shows a change in the upper quantiles that is not projected by the validation ensemble and falls outside of its bounds[3]. Precipitation shows a complex change regarding its quantiles. Such a quantile-dependent change is not projectable with the methods described in section 3.8, which change all quantiles equally. The distributions of the wind speed components exhibit quantile-dependent changes as well, which are equally misrepresented in the validation ensemble.

[3]The observed time series suggests even an inhomogeneity like a change in instrumentation (not shown).

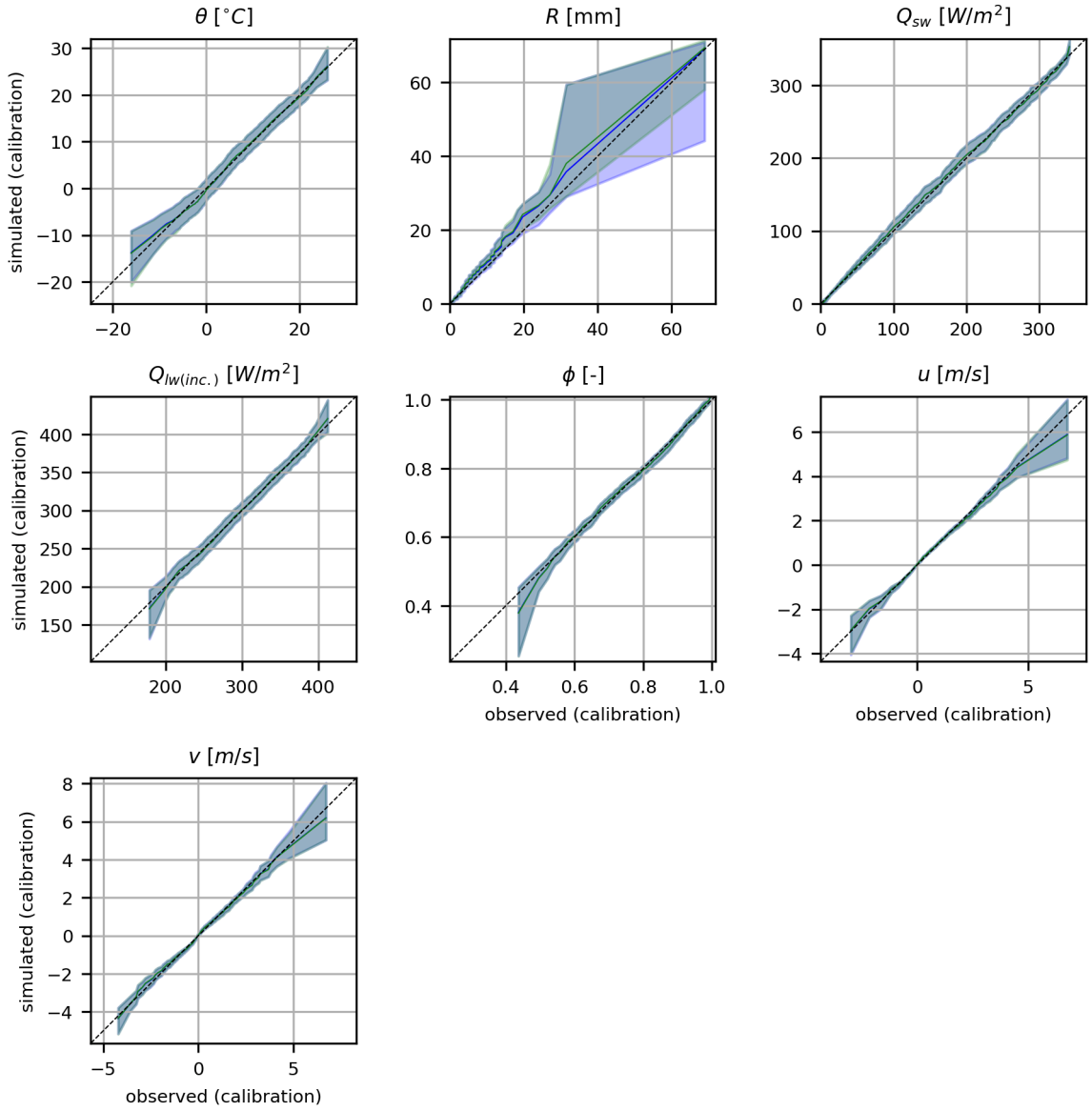


Figure 3.14: QQ-Plots for calibration period within extrapolation test. Blue and green lines show simulated against observed values in the validation period. Solid coloured lines show the median and dashed coloured lines the 1–99 % range of 500 realizations.

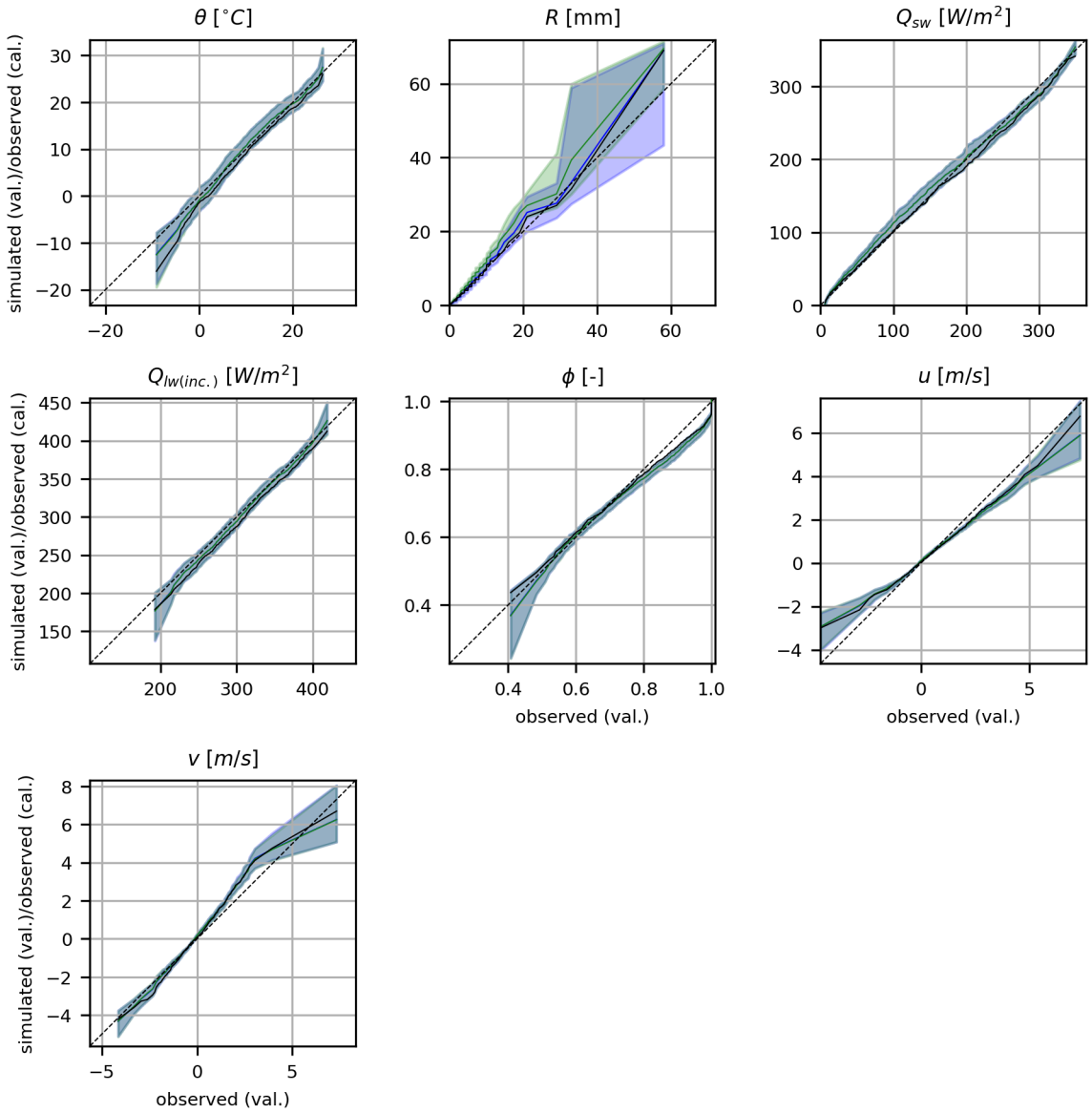


Figure 3.15: QQ-Plots for extrapolation test. The black line shows observed quantiles of the calibration against validation period. If this line falls on the x=y diagonal, the distributions of calibration and validation period are the same. If it is situated to the right of the x=y diagonal, it implies an increase from calibration to validation (e.g. for temperature θ). Blue and green lines show simulated against observed values in the validation period. Solid coloured lines show the median and dashed coloured lines the 1-99 % range of 500 realizations.

3.10 Summary and Discussion

This chapter presented the WG VG, starting with a general overview and then highlighting important detailed aspects. VG is a trans-Gaussian WG employing quantile-quantile transformations to convert the marginals of its input to standard-normal. While the use of an autoregressive model such as VAR is far from novel, the design choice underlying the generation of precipitation and the use of phase randomizing residuals to achieve better dispersion behaviour is.

VG can be seen as a framework for weather generation, as variables can be included as long as suitable distribution families for transformation to standard-normal can be found. For the underlying VAR-process, none of the variables are given a special treatment, which is achieved by the infilling of precipitation in a pre-processing step.

To achieve this, gaps in precipitation were filled with latent, sub-p_0 episodes, derived from the values of the other generated variables. The infilled data can be interpreted as dryness probabilities, transformed to the part of the standard-normal distribution below the p_0 value. With this information, the VAR-process recreates precipitation intensities, wet- and dry spell lengths reasonably well. Distributions of non-precipitation variables conditional on precipitation state show satisfactory agreement to the conditional distributions of observed data, despite the fact that VG does not have different parameters depending on rain state. With the observed data from the Konstanz measurement station, the method that used euclidean distance to "wet" means of non-precipitation variables outperformed the method based on multiple linear regression slightly in the aspect of dry spell length. Using the distance approach in the extrapolation test resulted in the right direction of change in precipitation. However, with the small differences between the results of the two dryness probability estimation methods, it seems hard to propose one definitely, especially, as the present chapter only used data from one station. In addition, it is questionable whether the dryness probability estimation method *should* influence the propagation of temperature change to precipitation change, as the dependence regarding the dry part of the transformed distribution is of a different nature than the wet part.

I have shown how to generate "balanced scenarios" in which a change in one guiding variable is propagated to the rest of the simulated variables in the context of VAR processes with qq-transformed marginals. The method assumes a linear relationship between the guiding variable and the other variables. I showed that the parameters of the VAR process must be taken into account to achieve the desired change in means. Filling up the precipitation gaps with dryness probability in the standard-normal domain was a preparatory step for it to join the proposed scenario generation method.

By using phase randomization as a source of randomness for a VAR processes, I was able to combine advantages of both methods. This made the VAR process honour longer-term linear correlations and offers a possibility to generate scenarios using phase randomization.

Apart from phase randomization inducing longer-term correlations into the VAR process, VG does not include explicit measures to avoid overdispersion. There exist different proposals on addressing overdispersion in the literature such as a link to a monthly WG (Khazaei et

al., 2013) or the use of wavelets (Steinschneider and C. Brown, 2013). Overdispersion in VAR is observable in the transformed domain, but turns to underdispersion in the measurement domain, suggesting that the marginal distribution plays a bigger role than the underlying process. Despite the non-perfect fit of the precipitation marginal, I could show that phase randomizing VAR residuals leads to a much improved dispersion behaviour in the transformed domain. The use of phase randomization to address dispersion behaviour in WGs is not new, but I am unaware of a work using it on residuals. Chen, Brissette, and Leconte (2010) used phase randomization to adjust monthly and yearly precipitation in an iterative post-processing step.

Projecting changes from one variable to another by exploiting linear relationships is fragile as seen in section 3.9. Evaluating the projection is made harder as observation-based extrapolation testing is limited by data availability. Further discussion about extrapolation using statistical means is given in section 4.6.2.

4 WeatherCop: A Combined Phase-Randomization Vine-Copula Weather Generator

The motivation for developing a copula-based Weather Generator (WG) came from the need to have a WG capable of what Vector-Autoregressive Weather Generator (VG) offers, but replace the Vector-Autoregressive (VAR)-model at its basis with a model that handles dependencies better. VAR-models can only represent linear dependencies whereas copulas are models of dependence, meaning that regarding the kind of dependence present in observed data, a suitable copula family can be chosen and fitted accordingly. Additionally, **Weather** Generator based on Phase Randomization and Vine **Cop**ulas (WeatherCop) has a broader area of application because it is a multi-site WG.

In order to keep the number of free parameters low, a vine copula is only used for time-invariant inter-variate dependence. Using a copula for inter-variate relationships enables a statistically better propagation of a change in a guiding variable to the other variables; hence more realistic scenarios. Temporal and inter-site linear dependence is maintained here by the use of phase randomization.

4.1 Overview of WeatherCop

Both VG and WeatherCop use the same methods and code to transform (section 3.3), infill rain (section 3.6), retransform variables and add hourly information (section 3.7). Figure 4.1 gives an overview over the parts of WeatherCop that are different from VG.

The steps WeatherCop executes are as follows:

1. The de-seasonalized and individually standard-normally distributed observations are transformed to relative ranks. This results in a data structure that has the dimensions: number of measurement stations, number of variables, number of time steps.
2. A canonical vine (C-Vine) copula is fitted to the observed ranks. This involves automatic selection and fitting of one-parametric bivariate copula families based on likelihood[1]. Appendix C lists the copulas available for this step.

[1] As only one-parametric copula families are compared, using an information criterion would not change the selection result.

3. The obtained vine is used to decorrelate the observed ranks using the inverted C-Vine sampling algorithm (see section 4.4).
4. The decorrelate is transformed via the inverse cumulative distribution function (*cdf*) of the standard-normal distribution.
5. The phases of the transformed decorrelate are randomized.

   The same random phases are used for all variables at all locations, thus not only maintaining temporal, but also inter-site dependencies.

   Optionally, a constant is drawn for each station from a normal distribution with zero mean and 0.25 standard-deviation and added to the central variable.
6. The phase-randomized variates are transformed to ranks using the univariate standard-normal *cdf*.
7. Finally the so obtained phase-randomized ranks are re-correlated using the C-Vine sampling algorithm 2.

In **step 2** above, only inter-variate dependencies are considered, resulting in a drastic decrease in dimensionality. This means that it is assumed that the dependencies between variables are independent from location. Concretely, the ranks for copula fitting are reorganized by concatenating the whole data matrix along the time dimension:

$$\begin{matrix}\text{Station 1:}\\ \\ \\ \\ \text{Station i:}\\ \\ \\ \\ \text{Station L:}\end{matrix}\begin{pmatrix}\begin{bmatrix} x_{1,1,1} & \cdots & x_{1,1,T} \\ \vdots & \ddots & \vdots \\ x_{1,D,1} & \cdots & x_{1,D,T}\end{bmatrix}\\ \vdots \\ \begin{bmatrix} x_{i,1,1} & \cdots & x_{i,1,T} \\ \vdots & \ddots & \vdots \\ x_{i,D,1} & \cdots & x_{i,D,T}\end{bmatrix}\\ \vdots \\ \begin{bmatrix} x_{L,1,1} & \cdots & x_{L,1,T} \\ \vdots & \ddots & \vdots \\ x_{L,D,1} & \cdots & x_{L,D,T}\end{bmatrix}\end{pmatrix} = \begin{pmatrix}\begin{bmatrix} X_{1,1} \\ \vdots \\ X_{1,D}\end{bmatrix}\\ \vdots \\ \begin{bmatrix} X_{i,1} \\ \vdots \\ X_{i,D}\end{bmatrix}\\ \vdots \\ \begin{bmatrix} X_{L,1} \\ \vdots \\ X_{L,D}\end{bmatrix}\end{pmatrix} \tag{4.1}$$

where D – number of variables
L – number of stations
T – number of time steps
$X_{i,j}$ – $(x_{i,j,1}, \ldots, x_{i,j,T})$, vector of data ranks for station i and variable j

Becomes:

$$U = \begin{pmatrix} X_{1,1} & \cdots & X_{i,1} & \cdots & X_{L,1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{1,D} & \cdots & X_{i,D} & \cdots & X_{L,D}\end{pmatrix} = (u_1, \ldots, u_D)^T \tag{4.2}$$

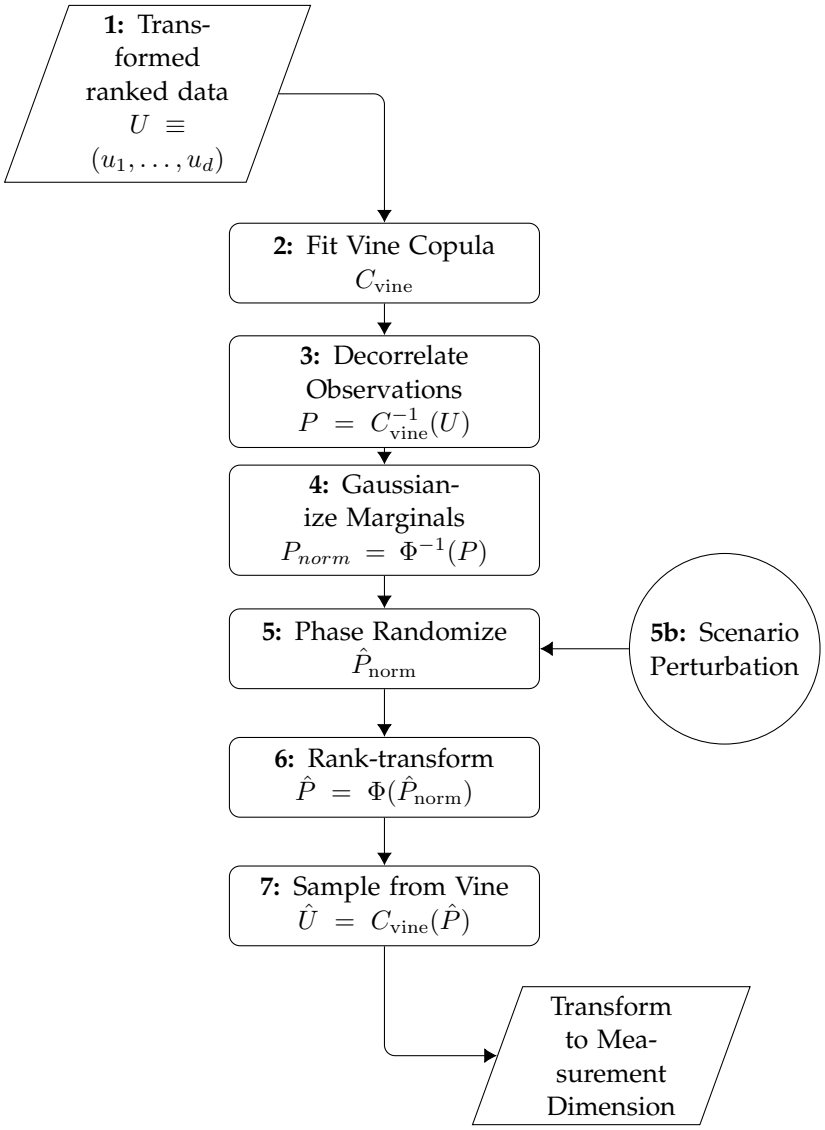


Figure 4.1: Structure of the WG WeatherCop. Not shown are the steps for transforming the marginals, which are the same as in the WG VG. The steps here replace the boxes "Fitting the VAR-Process" and "Simulate Time Series" in figure 3.1.

Thus the number of dimensions for the vine copula to consider is reduced from DL to D.

The C-Vine used in WeatherCop is of the simplified kind (see section 2.2), so the bivariate copulas are only dependent on the (virtual) ranks from the preceding tree and not on the values of the conditioning variables.

Another detail in **step 2** concerns dealing with the inhomogeneous dependency structure between precipitation and other variables, induced by dryness probability estimation (see section 3.6). Because of the estimation, ranks concerning dry conditions arise from a different process than those regarding wet conditions, which results in distinctly dissimilar patterns in copula density (see figure 4.2). With the aim of not influencing the fit of naturally occurring wet rank pairs, I amended the Log-Likelihood function to not consider the concrete position of rank pairs that relate to dry conditions. This change to the Log-Likelihood function is only made for bivariate relations that involve precipitation as one of the variables. The change is achieved by using dryness frequency instead of the probability density for such pairs (see equation (4.3)):

$$l(U, V, \theta) = \sum^{t \in t^+} \log \left\{ c_\theta \left(U_t, V_t \right) \right\} + \sum^{t \in t^-} \log \left\{ C_{u|v,\theta}(p_0, 1) \right\} \tag{4.3}$$

where t^+ – time steps with precipitation
t^- – time steps without precipitation

The bivariate copula fit compromises between the density pattern for wet conditions and the proportion of precipitation occurrence.

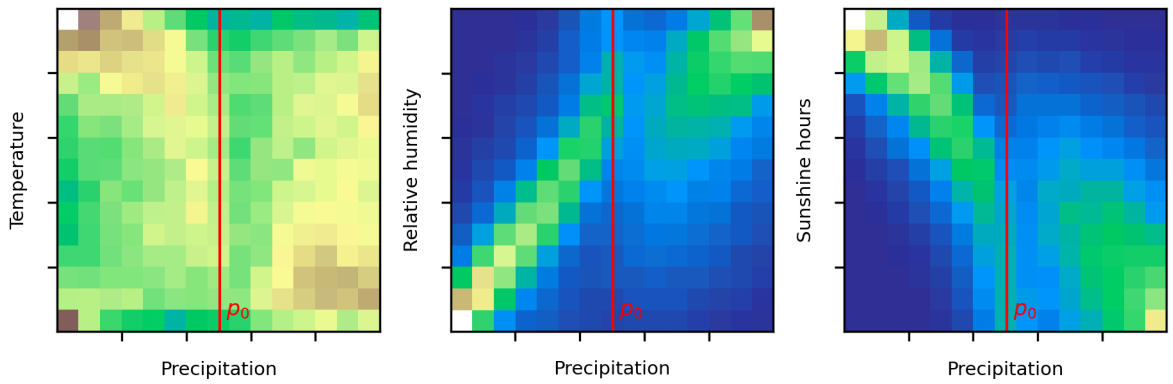


Figure 4.2: Empirical copula densities arising from estimation of dryness probability. Only the density to the right side of the vertical p_0 line, which are unrelated to the infilling method, are used during likelihood estimation.

The transformation to standard-normal prior to phase randomization in **step 5** provides two advantages over a raw phase randomization of the decorrelate: (1) phase randomization

could generate values outside of the range of relative ranks ($[0, 1]$) and (2) phase randomization produces time series with marginals that are closer to the Gaussian than the original signal[2]. The approach I used here differs from the one described in J. Theiler et al. (1992) in that the *cdf* (and its inverse) is used for transformation. J. Theiler et al. (1992) uses a gaussian sample for quantile-matching to obtain a normally distributed transformed time series and the empirical *cdf* of the original time series for back-transformation. The advantage of my approach over J. Theiler et al. (1992)'s is that different values than the ones from the source data set can appear in the generated time series.

Within **step 5**, there exists the option to add a constant to the central variable. The reasoning is that phase randomization does not change the mean of the source data but some variation might be desired. While having a constant mean is an advantage for some situations, it offers no abstraction of a naturally variable statistic. Adding it only to the central variable and before step 7, in which re-correlation via Vine Copula is done, ensures that the correlation structure fitted by the Vine is kept, resulting in dependent changes to the mean of the non-central variables.

After phase randomization, a scenario perturbation can be added to the variable that represents the central vine node (**step 5b**). This variable is usually the temperature. Every change done to the variable that represents the central node propagates through the conditional copula *cdf* of the first tree to all other variables.

4.2 Scenario Generation and the Choice of Vine Family

As formulated out in section 2.2, vine structures can be categorized as canonical (C), drawable (D) or regular (R). R-vines that are neither C- or drawable vines (D-Vines) have no central node as in C-Vines nor an "origin" node like D-Vines. These nodes represent starting points on which all other conditional distributions in the sampling process are dependent on. Those R-vines that lack a C- or D-Vine structure are more decentralized. The sampling in such R-vines is initialized independently along sub-branches around their local central nodes.

Because of the requirement to generate scenarios that are guided through one variable, only C- and D-Vines are applicable here[3]. Because of their simplicity, I chose C- over D-Vines. Despite their suitability in modelling time series, D-Vines do not add an additional value in the context of WeatherCop, because here all temporal dependencies are dealt with by the use of phase randomization. This is a compromise between model sophistication and simplicity as, by this choice, temporal dependencies are assumed to be linear.

By employing a C-Vine structure with the central node representing the variable that guides the other variables in scenario generation, only direct dependencies between this and the other variables are considered. Second-order effects are thus out of the scope of this kind of scenario generation.

[2]This is also the reason why phase randomization is problematic when working with untransformed data. Especially if their marginals are bounded (such as precipitation, relative humidity or sunshine duration).

[3]R-vines that are neither C- or D-Vines could be used to generate scenarios guided by several variables at once.

4.3 Annually Changing Dependence

Inter-variate dependence of meteorological variables is not generally stationary, but exhibits cyclical patterns in strength and even sign over the year. Dependence in vine copulas is influenced not only by the bi-variate copulas on each edge of the vine, but also by the graph. In WeatherCop, the configuration of the first vine tree is fixed in order to have one defined central variable. A change in graph structure to follow the seasons would impose abrupt changes in dependence. In order to avoid these sudden changes, I opted for a static vine graph, but changing bi-variate copula parameters throughout the year, should those yield a higher likelihood than stationary ones. Changing copula parameters are obtained by fitting a bi-variate copula for each day of the year, using also data from neighboring *doy*s with the help of a moving data window. This series of 365 copula parameter values (per vine edge) is smoothed by using its 4 longest harmonics obtained by Fast Fourier Transform (FFT).

4.4 Decorrelation by C-vine Copula

Algorithm 2 shows how to generate samples from a d-dimensional C-Vine given a vector $P = (p_1, \ldots, p_d)$ containing random numbers drawn from $U(0, 1)$. Algorithm 3 is the inverse formulation, returning the vector P calculated from a given observed vector $U = (u_1, \ldots, u_d)$. In other words, using P obtained from algorithm 3 as input to algorithm 2, one recovers U.

Algorithm 3: Inverse C-Vine sampling algorithm.

Data: $u_1, \ldots, u_d$ observed ranks
Let $p_1 \leftarrow u_1$
Let $p_2 \leftarrow C_{2|1}(u_2|u_1)$
for $j = 3, \ldots, d$: **do**
    let $q \leftarrow u_j$
    for $l = 1, \ldots, j-1$ **do**
        let $q \leftarrow C_{j|l;1:(l-1)}(q|p_l)$
    let $p_j \leftarrow q$
return $(p_1, \ldots, p_d)$

Under the assumption that the correlated observations U follow the copula represented by the fitted vine, by definition, the univariate time series P are independent.

4.5 A Few Implementation Specifics

Some choices made during coding warrant a description here, because they helped in achieving a reasonably fast execution speed and reduced error-prone typing of lengthy equations despite a high level of abstraction. I coded WeatherCop using the interpreted

programming language Python. In order to optimize run-time, I implemented parts of the code in Cython, a superset of Python which compiles to C (Behnel et al., 2011). While this approach is common in the scientific Python software community, I adopted a less common method for bivariate copulas. I employed sympy, a computer algebra system with code-generation capabilities (Meurer et al., 2017). The Archimedean type of copulas is mathematically fully determined by its generator function (see line 8 on page 18). From the generator function, the copula *cdf* can be constructed, from which in turn expressions for the copula density and the conditional *cdf*s can be derived. With a few exceptions, implementation of a copula family of the Archimedean type in WeatherCop consists only in explicit coding of the generator function. In the case of non-Archimedean copulas with closed-form copula *cdf*s, the *cdf* is supplied before continuing with deriving the necessary expressions. All other needed mathematical expressions are constructed symbolically during a one-time setup step with sympy[4]. What follows is automatic Cython code generation with tools from sympy for each mathematical expression.

Not every algebraic task can be done automatically or even manually. Problematic steps during equation manipulation are inversions, which are required operations for obtaining the copula *cdf* from the generator function and for sampling from conditional *cdf*s. For generator functions the inverses can be found in the literature (see Joe (2015, Chapter 4)), but closed equations for inverse conditional *cdf*s are often non-existent. However, for the latter, expressions for first and second derivatives are constructed programmatically with sympy in WeatherCop which help in inverting the conditional *cdf*s numerically.

In order to have a higher number of available bivariate copulas, I employed sympy again for rotating copulas in 90 degree steps. These rotations can be achieved by the following substitutions:

$$\begin{aligned} C_{90}(u,v) &= v - C(1-u,v) \\ C_{180}(u,v) &= u + v - 1 + C(1-u,1-v) \\ C_{270}(u,v) &= u - C(u,1-v) \end{aligned} \tag{4.4}$$

These substitutions are also constructed symbolically with sympy and compiled to C via Cython.

4.6 Simulation Example

The aim in this example is to evaluate the ability of the weather generator to generate data in the likeness of its input data and to see how extrapolation via a change in temperature performs. In contrast to the simulation done for evaluating possible climate impacts on Lake Constance in section 6.2, this example is not indirectly validated using an impact model.

[4]The copula expressions in appendix C were automatically generated this way.

I fitted WeatherCop on freely available data from 14 measurement stations in southern Germany operated by the German Meteorological Service (DWD) (see figure 4.3). The dataset spans from 1996 to 2017 and contains the variables air temperature, sunshine duration, precipitation and relative humidity. The period was chosen in order to maximize data availability and to have a reasonably large time frame for extrapolation testing. In the following, I present the performance of WeatherCop in two modes: (1) representation of stationary conditions, by fitting on the full data set (section 4.6.1) and (2) ability to extrapolate from colder to warmer conditions by split-sampling (section 4.6.2). I generated 500 realizations for each experiment. In order to have variation of means, despite using phase randomization, a disturbance constant is drawn for each realization from a normal distribution with zero mean and 0.25 standard deviation (see also section 4.1). With few exceptions, throughout the example, only the outcomes following the regression method for estimating dryness probability (described in section 3.6 on page 36) are shown, as the choice of this method has negligible influence on the outcome in WeatherCop.

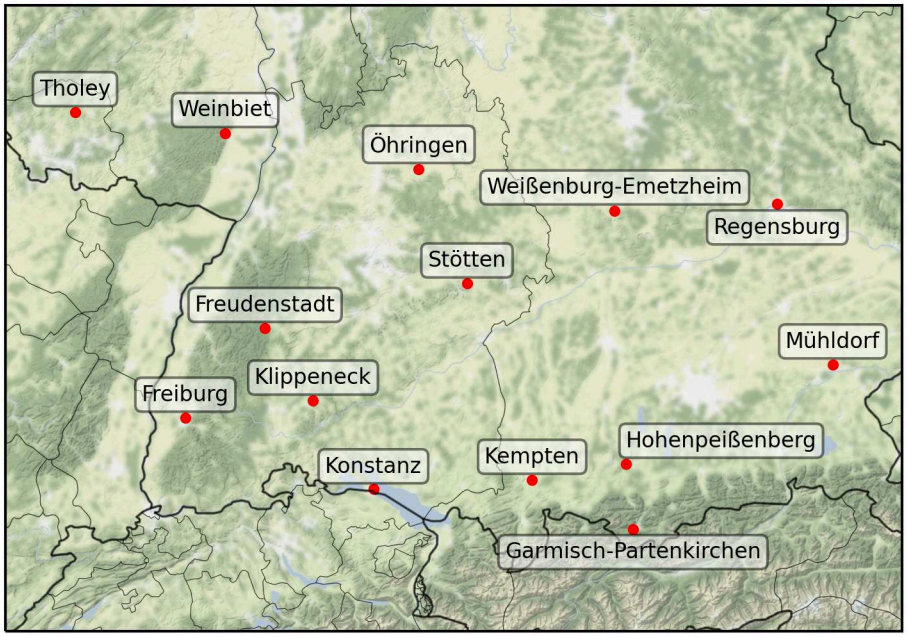


Figure 4.3: Location of measurement stations in southern Germany.

Table 4.1 lists the distribution families I used for transforming meteorological observations to ranks. Precipitation was modelled with a Kumaraswamy distribution to match values below a fitted quantile and kernel density estimation (KDE) with a log-Normal kernel above. The Kumaraswamy distribution (Kumaraswamy, 1980) is closely related to the beta distribution, but does not require the evaluation of a transcendental function (as the Gamma function, when using the beta distribution), resulting in faster code. The aim was to have a

parametric abstraction for the lower part and a flexible method for higher values. By using a log-Normal kernel, it is possible to generate values above the historical maximum, which might be especially appropriate in climate scenarios and large ensemble sizes. In Schlabing et al. (2014) a truncated Normal distribution was fitted to relative humidity from Constance, but unfortunately, this distribution is not a suitable model for other sites. Thus, I opted for KDE with Gaussian kernels as distribution approximation for relative humidity in this example.

Table 4.1: Variables and their distribution families used for transformation to ranks.

Variable	Symbol	Distribution
Air temperature	θ	Normal
Precipitation	R	Kumaraswamy with KDE for upper tail
Sunshine duration	sun	Kumaraswamy with KDE for upper tail
Relative humidity	ϕ	KDE

4.6.1 Performance Under Stationary Conditions

The vine structure with the bivariate copulas chosen automatically by maximum likelihood are shown in figure 4.4. Kendall τs for seasonally changing copulas are shown as averaged over the year. As these copulas each have a *doy*-varying parameter, correlations can be higher or lower during different times of the year. Opposed to the motivation behind C-Vine graph selection (see section 2.2), the highest correlations do not occur in the first tree. This can happen if the central node in the first tree ($\mathcal{T}_0$) is not chosen by maximizing the sum of absolute correlations involved, but pre-defined externally, as is the case here with temperature. The highest correlations actually appear in the second tree ($\mathcal{T}_1$) with precipitation in the central node. This tree handles the dependence between the pairs sunshine duration – precipitation and relative humidity – precipitation with the influence of temperature removed. Removing the influence of temperature with a statistical method from this triplet of variables can only go so far, as the underlying physical processes defining their relationships (e.g. cloud formation) depends on much more than the value of that variable, hence the still high correlations in the second tree.

Notable is also that there is no difference in selected copulas with regard to the dryness probability estimation method, due to the likelihood-estimation described in section 4.1.

Figure 4.5a shows the fit of marginal distributions in terms of qq-plots using the Konstanz station as an example. While temperature, sunshine duration and relative humidity are represented well, there is an overestimation of precipitation for middle to high quantiles. This overestimation does not appear systematically at all stations as exemplified by the Weinbiet station (see figure 4.5b). The distributions vary from realization to realization according to the perturbation induced to temperatures. Most variation exists in the tails of extremes of unbounded variables[5]. The methods for dryness probability estimation produce no different marginals.

[5]Relative humidity is in a strict sense bounded, but its theoretic lower bound is never reached in observations.

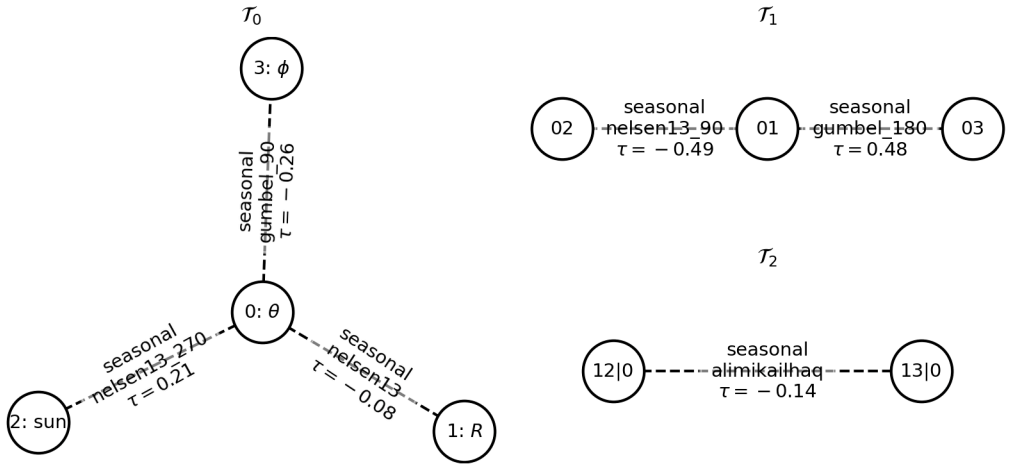
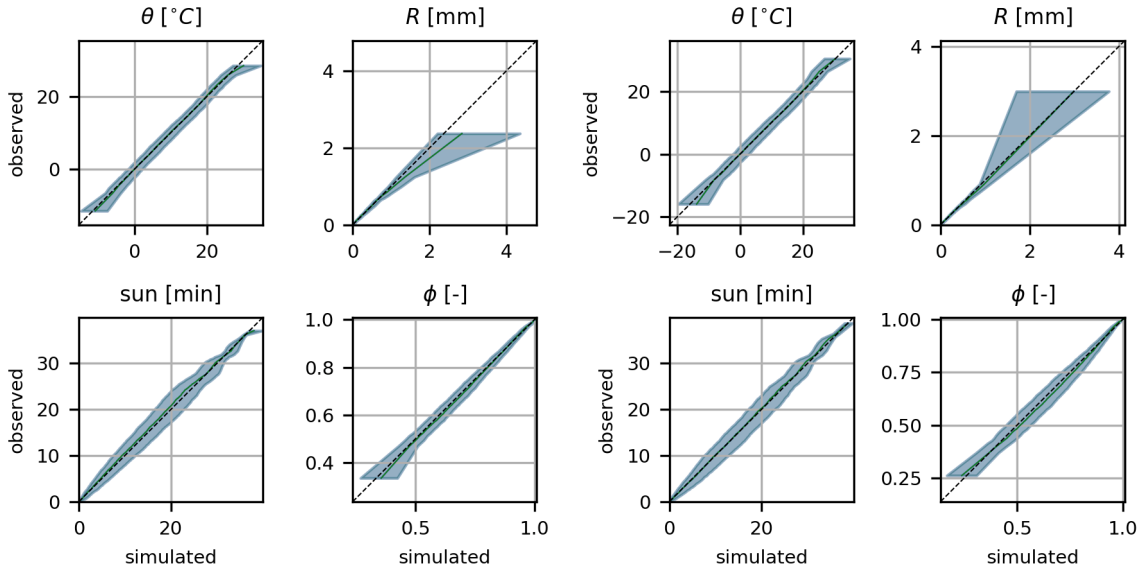


Figure 4.4: Vine trees for the full data period with the regression method for dryness probability estimation. The correlation coefficients are Kendall's τs. "[S]easonal" copulas have *doy*-specific parameters.

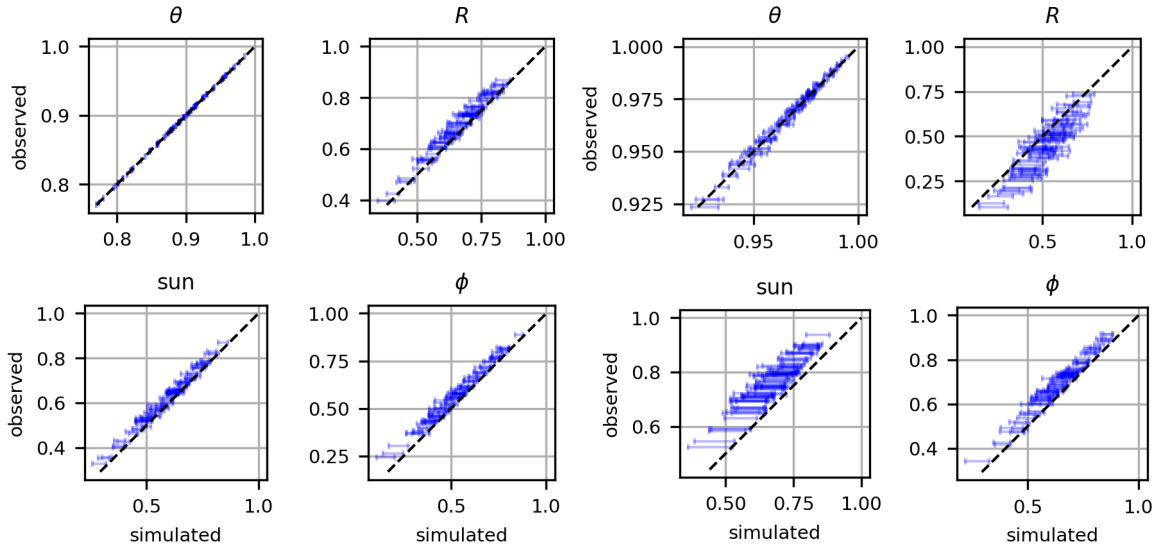


(a) Konstanz

(b) Weinbiet

Figure 4.5: QQ-Plots for the stationary example. Shaded area is the full range of all ensemble realizations. Solid black line the median of all realizations. Colors code two different dryness probability estimation approaches: blue is for regression and green for distance method.

Figure 4.6 shows the reproduction of inter-site (spatial) correlation coefficients for the same meteorological variable at each station. The near-perfect fit in the transformed domain for temperature (see figure 4.6a) is not surprising as this arises purely from phase randomization, which keeps correlations constant by design and the vine sampling algorithm does not change the first variable. Back-transforming the marginals, which includes also recovering the annual cycle of the data, results in a degradation of the correlation coefficients (see figure 4.6b). This is most notable in lower correlation for sunshine duration and relative humidity and higher correlation in precipitation. There are virtually no differences for different methods of dryness probability estimation (not shown).



(a) Transformed domain

(b) Measurement domain

Figure 4.6: Inter-site correlation coefficients. These are related to phase randomization. Lines in the right panel show ranges of simulated correlation coefficients.

The vine copula is able to reproduce the various observed inter-variate rank correlations to a large extent (see figure 4.7), but correlations in the measurement domain show large variations between realizations (see figure 4.8). It should be stressed however, that the absolute deviations from the observed correlation coefficients are small with a mean of 0.031.

In general, the correlation coefficients of variable pairs that are represented in earlier trees are better reproduced than ones that are dealt with late in the algorithm. This can be seen from the tight grouping of lines near the x=y-line for the pair air temperature – precipitation in figure 4.7. Due to the order of these variables, only one bivariate copula is involved in generating this pair. In contrast, the pair sunshine hours – relative humidity, appears in the last tree and shows worse fit in terms of correlation coefficients. The generation of these variables happens last, in dependence to all variables generated before. As more fitted bivariate copulas are involved and the vine is a simplified one (see section 2.2), inaccuracies can add up. Affected by the simplifying assumption in the present 4-dimensional case are 3 bivariate copulas – the ones in tree 1 ($\mathcal{T}_1$) and tree 2 ($\mathcal{T}_2$) handling the variable pairs sun-

shine hours – precipitation, precipitation – relative humidity and sunshine hours – relative humidity.

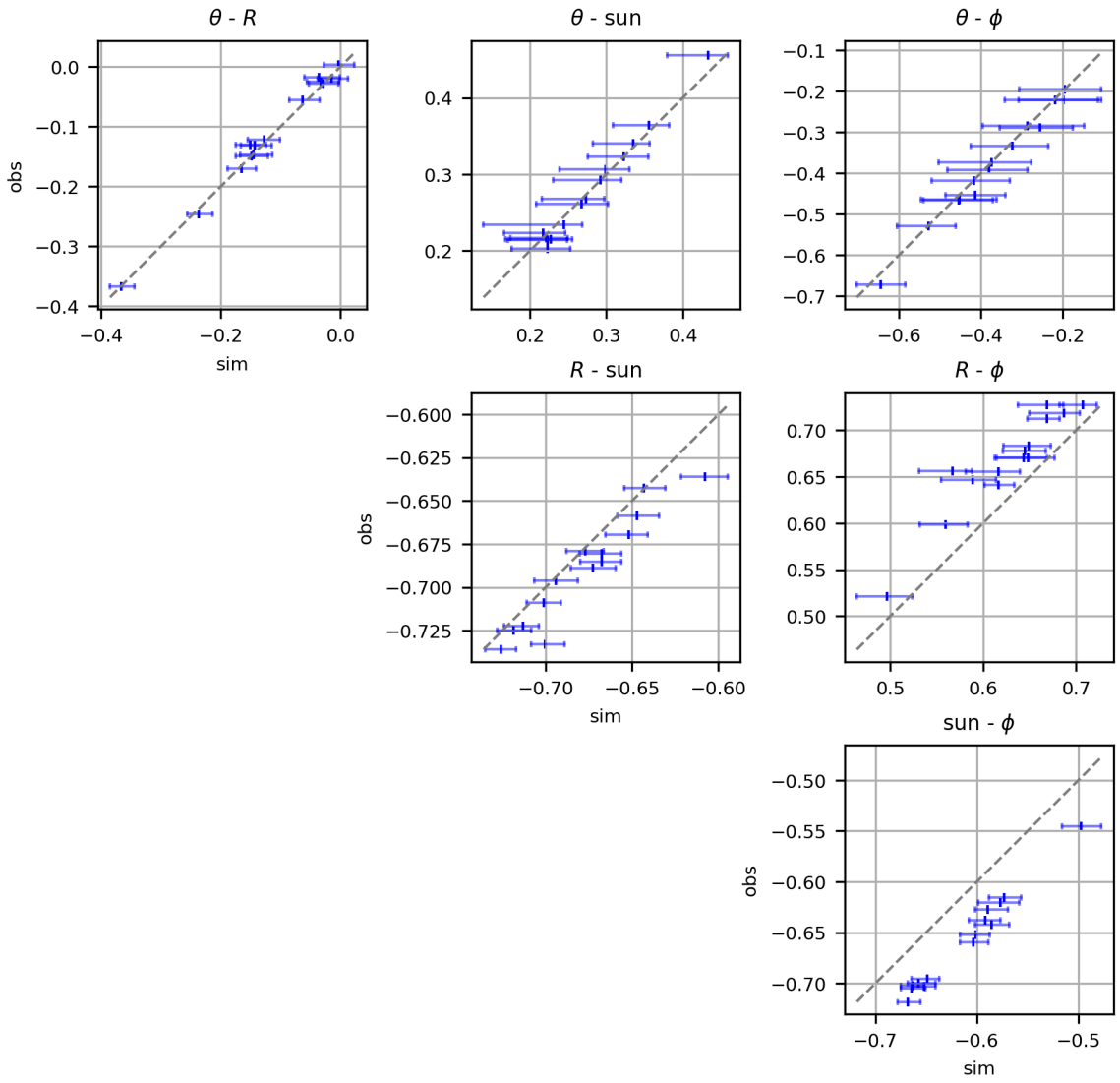


Figure 4.7: Intra-site correlation coefficients in rank-space. These are related to the vine copula. Lines show ranges of simulated correlation coefficients. Vertical bars near the middle of the lines show medians.

Figure 4.9 shows the fit of precipitation intensity and spell length distribution for two exemplary stations. While WeatherCop seems to provide a good abstraction of intensity, the quality of spell lengths fit differs from station to station. The lengths of dry spells for Konstanz are underestimated (see figure 4.9a). A different example poses Weißenburg-Emetzheim (see figure 4.9b) for which both dry- and wet-spell distributions show good agreement with observed values regarding the regression method. Both methods show a underestimation rather than overestimation of long dry spells while no such consistent mis-

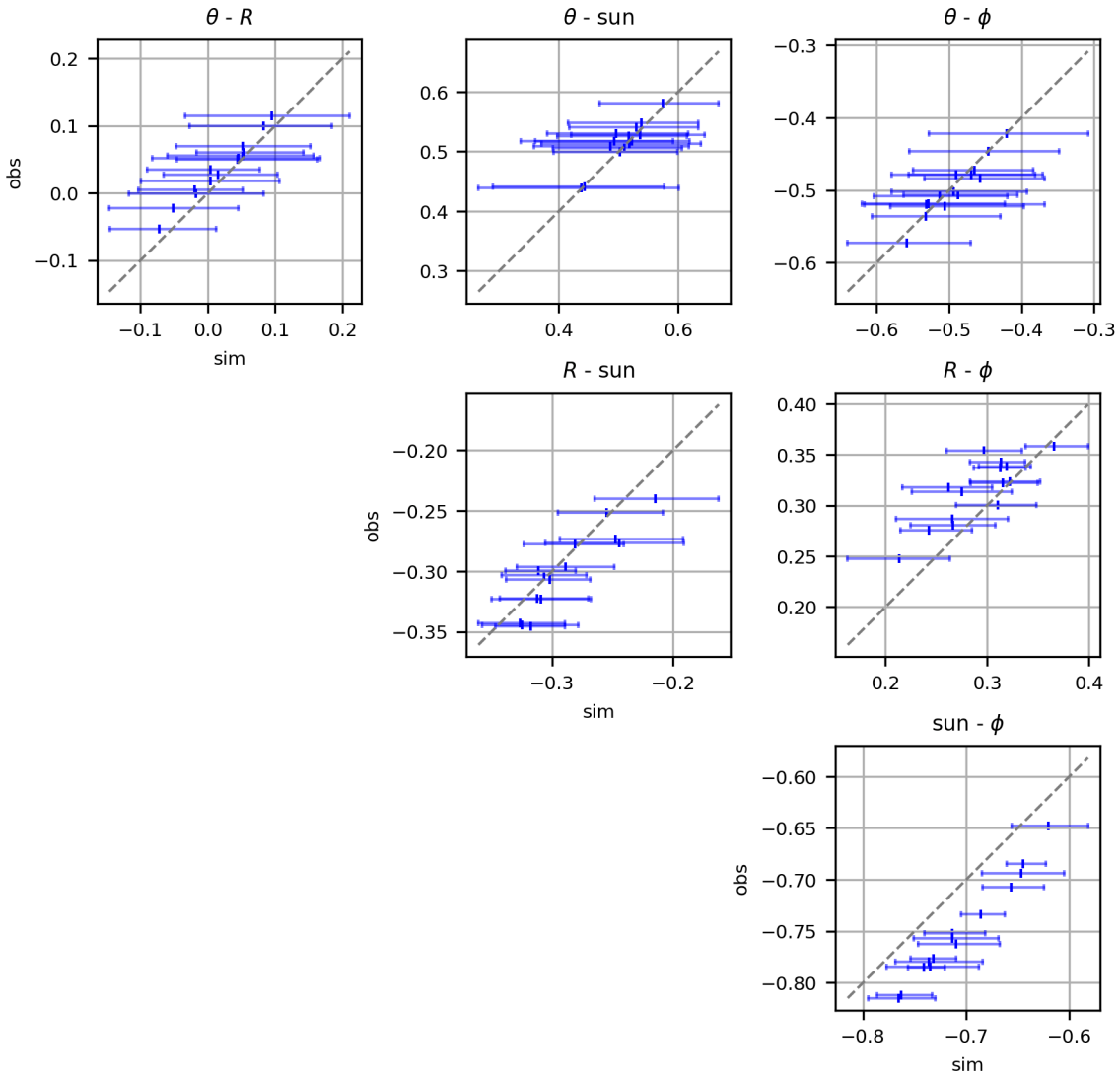
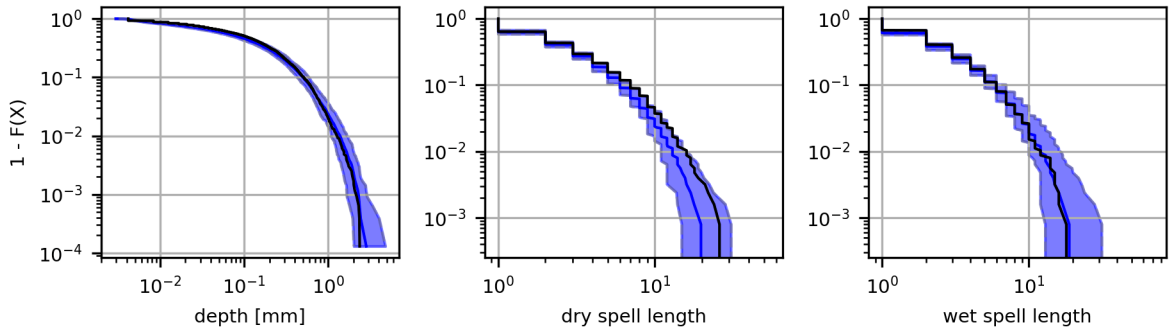
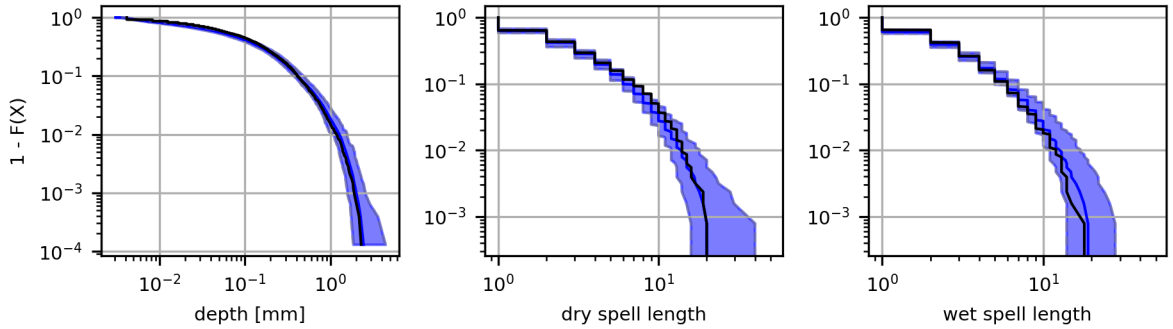


Figure 4.8: Intra-site correlation coefficients in the measurement domain. These are related to the vine copula and the marginal distributions. Lines show ranges of simulated correlation coefficients. Vertical bars near the middle of the lines show medians.

match exists for wet spells (not shown).



(a) Konstanz



(b) Weißenburg-Emetzheim

Figure 4.9: Precipitation exceedance probabilities for intensity, dry- and wet spell length. The thick blue lines shows the median of the ensemble, while the shaded area shows its bounds. Black line shows observed values.

4.6.2 Ability to Extrapolate

As WeatherCop is intended for climate scenario generation, it is pertinent to assess its ability extrapolate. Being fitted on a cold period and run with station-specific temperature difference to a warm period as mean change in temperature, it would be ideal if WeatherCop matches to not only the temperature increase, but also the change in non-temperature variables. This expectation relies on the copula to capture the underlying dependency structure of the data during the calibration period, and on the hope that this structure still exists in the time that connects calibration and validation period. This kind of split-sampling validation approach emulates the situation of projecting climate change in freely-defined scenarios, albeit with a smaller amount of data. I chose the first 5 years of the full data set (1996–2001) as calibration- and the last 5 years (2013–2017) as validation set.

Using only the calibration period for fitting, the chosen bivariate copula families change in 4 out of 6 cases for each of the dryness probability methods when compared to fitting on the

full-data period (compare figures 4.4 and 4.10). The stable pairs are sunshine duration – air temperature and precipitation – sunshine duration. The strengths of the dependence, measured in Kendall's τs change only little. This indicates that the relationships between the variables are not completely stable when comparing the 5-year calibration period to the full 22 years of available data.

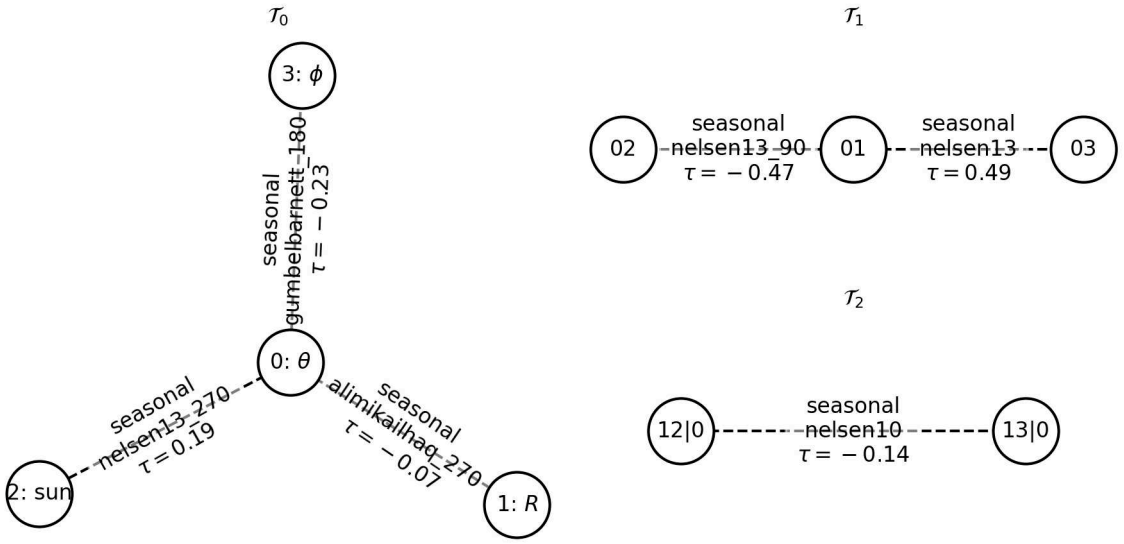
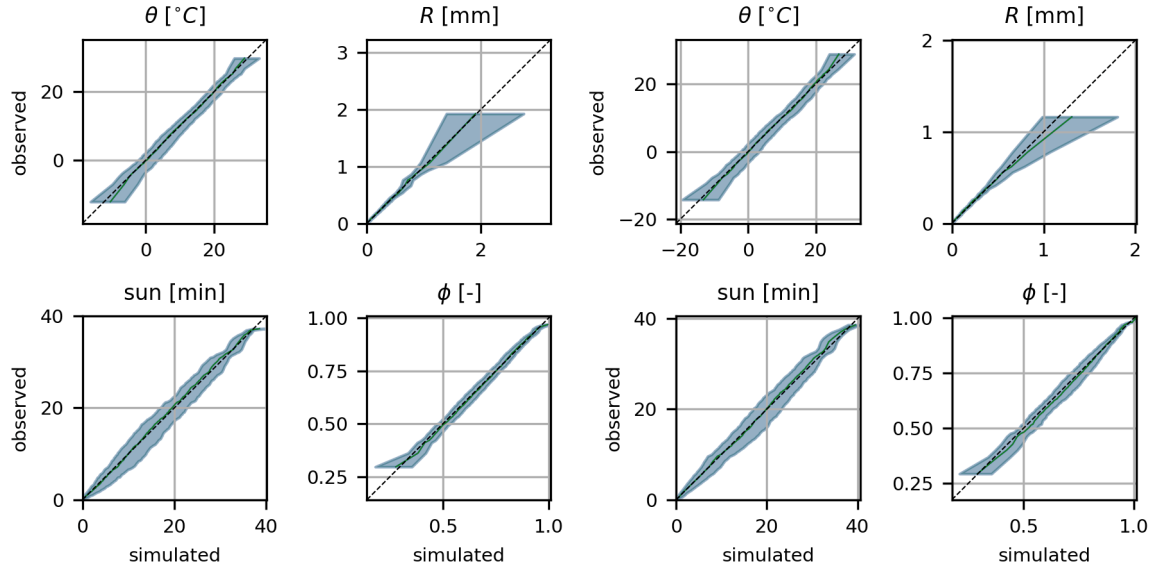


Figure 4.10: Vine trees fitted on the data from the calibration period. The correlation coefficients are Kendall's τs.

The calibration shows in general good agreement with the calibration-period source data with regard to the marginals. Still, upper-tail values of precipitation are overestimated for 5 out of the 14 stations, where rest of the station have a good representation in their upper-tail precipitation values (not shown). Figure 4.5 shows two examples of marginal fit in terms of qq-plot. Figure 4.11a (Freiburg) is an example for good and figure 4.11b (Weinbiet) an example for less good fit. There are no noteworthy differences between the results depending on dryness probability estimation method.

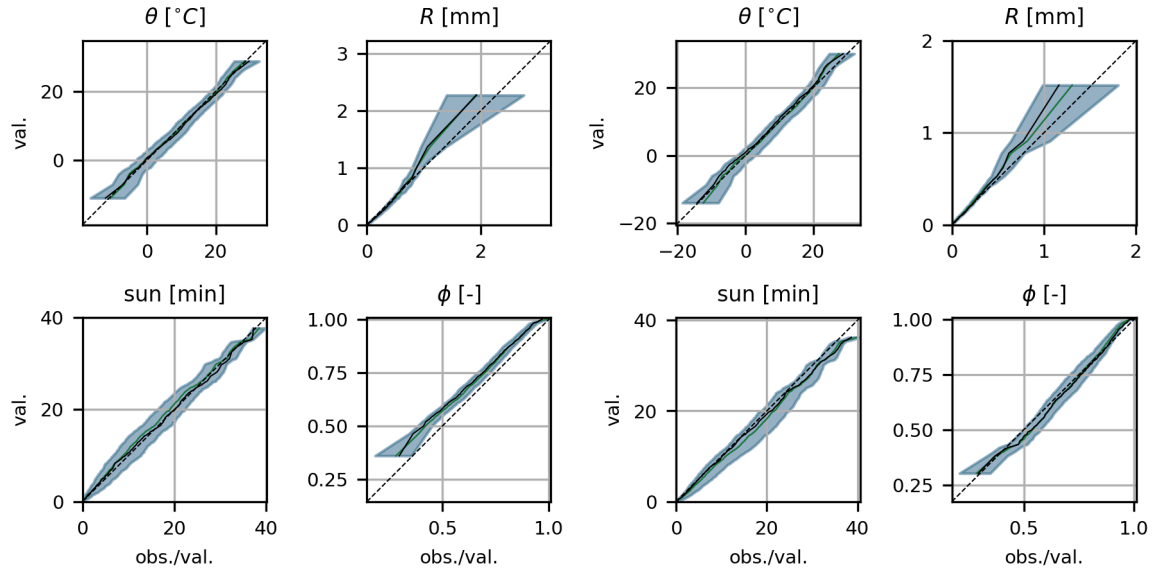
The changes from calibration to validation period are not drastic. Figure 4.12 shows the change in the form of qq-plots for the two exemplary stations Freiburg and Weinbiet. The only severe change in distribution is relative humidity for the Freiburg station. Less pronounced but still significant is the shift towards higher values in upper quantiles of precipitation for both stations. Upper tail precipitation values at Weinbiet were overestimated in the calibration, causing them now to be closer to the validation period values.

Figure 4.13 shows the changes in daily means and the variability of the ensemble for the stations Freiburg, Konstanz, Weinbiet and averaged over all 14 stations. Means are also listed in tables B.1 and B.2 in the appendix. Means for temperature very closely meet both the observed and the targeted means in the validation period. Observed means during the calibration period for the other variables are not as precisely met, but still lie within the range of the ensemble.



(a) Freiburg (b) Weinbiet

Figure 4.11: QQ-Plots for calibration period.



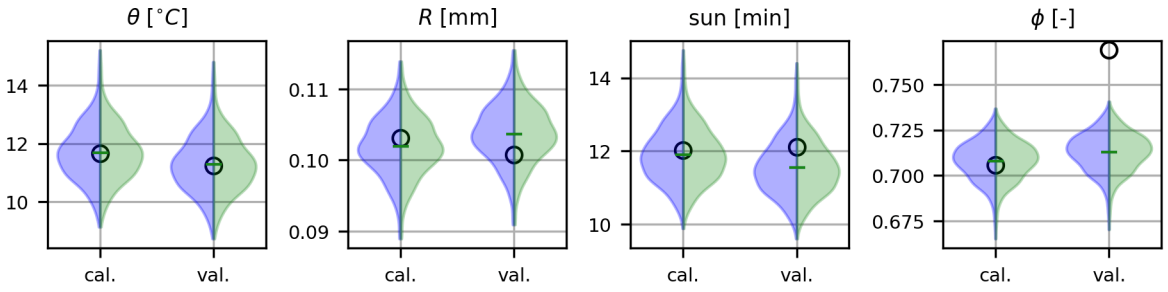
(a) Freiburg (b) Weinbiet

Figure 4.12: QQ-Plots for extrapolation test. Black lines show quantiles of the validation over the calibration period, indicating how the distribution changed from the former to the latter.

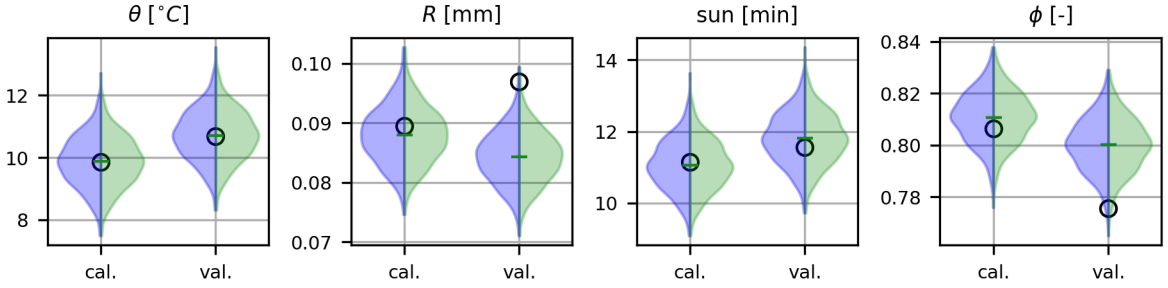
An outlier in terms of observed change is relative humidity at Freiburg. As figure 4.12a already showed a shift in distribution, the change in mean is thus as drastic as well and lies outside of the range of the ensemble in the validation period.

The change in observed sunshine duration highlights an interesting problem. The increase in temperature is accompanied by an increase in sunshine duration for Konstanz, but a decrease in Weinbiet. Generally, temperature and sunshine duration are positively correlated in the combined data set (see figure 4.7) which manifests as an increase in sunshine duration for the simulated values. A reason that site-specific changes are not projected efficiently might also be due to the fact that the vine copula is fitted globally and not site-specific (see section 4.1). Averaging the projected changes over all stations yields satisfactory results for precipitation and to a lesser extent, sunshine duration. The precipitation decrease is slightly underestimated, whereas the increase in sunshine duration is overestimated. The observed increase in relative humidity is misprojected as a decrease.

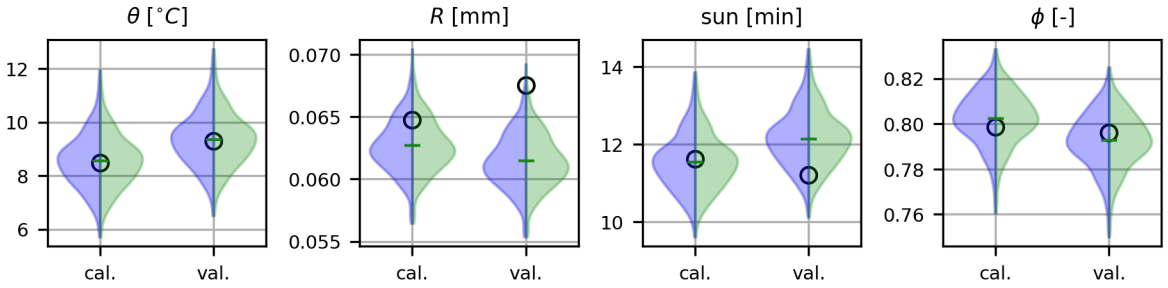
The example of Freiburg in figure 4.13a is interesting in that it shows that spatial variability plays a role in the temporal trends. In Freiburg, the air temperature decreased from the calibration to the validation period, contrary to the overall trend.



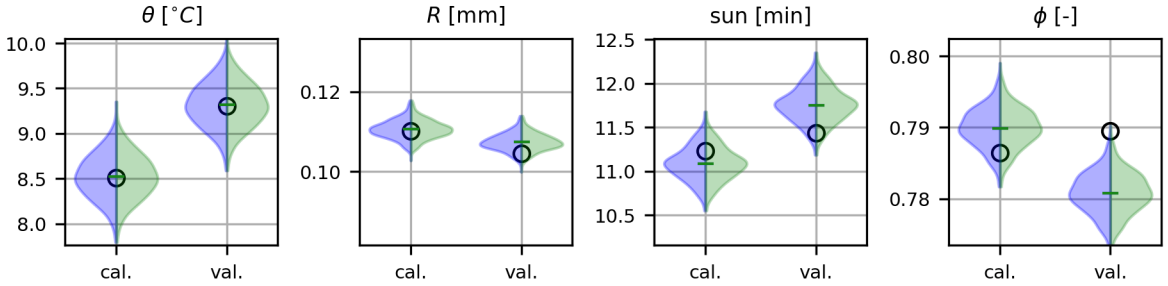
(a) Freiburg



(b) Konstanz



(c) Weinbiet



(d) Mean over all stations

Figure 4.13: Violin plots of changes in mean. Densities shown are means of 500 realizations. Black circles correspond to observed values. Blue corresponds the regression, green to the distance simulation method.

4.7 Summary and Discussion

This chapter introduced WeatherCop, a multi-site, daily weather generator based on vine copulas and phase randomization. Its design was guided by the same aims as VG, providing the variables and features needed for climate impact assessment on lakes. The most meaningful difference to VG is arguably not the fact that it generates weather for multiple stations at once, but that the dependencies between weather variables are modeled with copulas instead of linear models.

Phase randomization in a multi-site WG is able to reproduce correlations across space and time without the need for any additional parameters. Its limitation of only being able to produce "unchanged" scenarios were overcome with the help of vine copulas. Joining phase randomization and vine copula sampling was achieved by decorrelating observations through an inverse formulation of the C-Vine copula sampling algorithm, phase randomizing the decorrelated observations and using them as random numbers in the forward formulation of the C-Vine sampling algorithm.

Both VG and WeatherCop are WGs that use a transformation of marginals. This enables the use of methods that dictate specific input distributions, but has the disadvantage of an indirect and often less good fit in the measurement than in the transformed domain. In the case of WeatherCop, inter-variate dependence is modelled in rank-space and not in the measurement domain of the observed values. Hence the overall fit does not only depend on the vine, but also on the transformation. This dilemma of fitting on transformed data is very evident in the case of correlations being reproduced well in the transformed domain with the help of phase randomization and deteriorating through back-transformation to the measurement domain.

In VG, the choice of dryness probability estimation method plays an important role for projecting change. This is different and deliberately the opposite in WeatherCop, where the maximum likelihood estimation of bivariate copula parameters was adjusted to "ignore" the infilled, dry part of the data. In most plots in this chapter, I did not show graphs for results for the different dryness probability estimation methods as they are largely indistinguishable.

The result of dependence-structure conflicting change of some variables in the observed data set during the extrapolation test highlights a problem that can arise when projecting changes via the assumption of constant dependency structures: a statistical model cannot predict that variables might behave contrary to their statistical relationships. Or when the central assumption of stationarity of the statistical relationships is violated, as the fact that half of the bivariate copulas change when using a subset of the full data set, indicates. The fact that WeatherCop employs copulas, which are models of dependence, cannot help in this regard. A better fit might be achieved, however, by fitting copulas per site and not globally as done here. A more thorough test for WeatherCop would be to repeat the exploration test with a data set that does not violate said assumption, if such a data set could be found.

5 Comparison with a Resampling Method

"Anyone who studies present and ancient affairs will easily see how in all cities and all peoples there still exist, and have always existed, the same desires and passions. Thus, it is an easy matter for him who carefully examines past events to foresee future events in a republic and to apply the remedies employed by the ancients, or, if old remedies cannot be found, to devise new ones based upon the similarity of the events. But since these matters are neglected or not understood by those who read, or, if understood, remain unknown to those who govern, the result is that the same problems always exist in every era."

Niccolò Machiavelli, Discourses on Livy, 1517

"The past does not repeat itself, but it rhymes."

Attributed to Mark Twain

This chapter describes some specific features of the K-Nearest Neighbors (KNN) resampling method and compares its output to the parametric methods from chapters 3 and 4. It is a is a direct validation that compares observed and generated data of the same variables. Some examples for indirect validation using various lake models are discussed in chapter 6.

The structure of this chapter is as follows. First, a motivation for the comparison of resampling and the parametric Weather Generators (WGs) is given in section 5.1. Then, changes to the resampling method introduced in section 2.3 are described in section 5.2. These changes are necessary for generating time series that match the annual cycle of observations and in order to generate scenario output. Output characteristics of the resampler, Vector-Autoregressive (VAR), Vector-Autoregressive Weather Generator (VG) and **Weather** Generator based on Phase Randomization and Vine **Cop**ulas (WeatherCop) are presented in section 5.4.

5.1 Motivation

Resampling approaches offer a seemingly convincing set of advantages. They produce "synthetic" data with statistical properties (auto-, spatial and cross-correlations, higher-order statistics) close to the observed data, can generate climate scenario output and are easy to implement. However, a rather obvious restriction is often overlooked. Namely, they do not contain any abstraction from the data – inherently, they assume that the observed data (or key aspects of it) *is the abstraction*. Hence, the past is all that could have happened and

might happen. The remainder of this chapter describes consequences of the KNN selection process.

The comparison is, however, not meant to underline possible advantages of parametric over non-parametric methods. The measures used in this chapter reveal deficiencies of the parametric methods as well. In the case of the VAR-process the deficiencies can be attributed in part to the fact that it generates multivariate Gaussian distributions, which might not be a good approximation to the underlying data.

5.2 Bringing the Resampler Closer to the Weather Generator VG

In order to achieve a comparison that relates to the nature of the different methods and not to aspects of implementation, I modified the classical KNN-resampler slightly. The KNN resampling method is described in section 2.3. I opted for using the classical euclidean distance of the transformed variables for forming the candidate sets. The transformation of the marginals to standard-normal brings every variable to the same scale and also eases the comparison to VG and WeatherCop, because the same transformation method is used there (see section 3.3).

5.2.1 Annually Changing Dependencies

In order to recreate the annual cycle in inter-variable dependencies, I constrained the set of candidate time-steps. This is achieved by only selecting time steps a certain maximum *doy*-distance apart from the resampled time step. Note that annually changing marginals are captured by the transformations described in section 3.3.

5.2.2 Generation of Scenarios

In order to generate scenarios, the resampling process can be biased towards the desired statistical properties. To be able to compare KNN with the WGs presented in this work I generated time series with a higher mean by biasing the selection method.

Simulation with Changed Mean

Consider equation (2.37) which defined the squared euclidean distance used as a dissimilarity measure to limit each resample step in order to maintain the autocorrelation of the source data set. This distance measure is modified here so that the selection of time series results in a changed mean temperature in the resampled time series:

$$d_s = \sum_{i=1}^{p} \sum_{j=1}^{k} (y_{t-i,j} - x_{s-i,j} - d_j)^2 \quad \forall s \in \{p+1, \ldots, T\} \tag{5.1}$$

where
- d_s – squared euclidean distance between current chunk and chunk at time step s – biased by d_j
- p – number of previous time-steps considered
- k – number of variables
- t – time index in the resampled data
- $y_{t-i,j}$ – previously resampled values
- $x_{s-i,j}$ – values from the source data
- d_j – constant to increase the mean of variable with the index j of the resampled time series
- T – number of time steps in the source data set

The only change is the addition of the vector $\mathbf{d}$. In principle, biases to all variables can be included, but I restricted myself to just adding a change in temperature and leaving all other elements as zero. Relating d_{j_θ} (with j_θ being the column index of temperature in $\mathbf{x}$) to the change in mean temperature $\Delta\overline{\theta}$ is not trivial. To find a useful approximation of d_{j_θ} as a function of $\Delta\overline{\theta}$ for a fixed p and number of considered candidates, I generated time series with varying d_{j_θ}. I fitted a function of the following form to the empirical changes in mean temperature obtained that way:

$$\Delta\overline{\theta}(d_{j_\theta}) = a \left(1 - \exp\left(-\frac{d_{j_\theta}^b}{c}\right)\right) + m \tag{5.2}$$

where a, b, c, m– constants

After finding suitable values for the constants a, b, c by numerical optimization, I used the inverse of equation equation (5.2) to obtain d_{j_θ}:

$$d_{j_\theta} = \left(-c \ln \left|1 - \frac{\Delta\overline{\theta} - m}{a}\right|\right)^{-b} \tag{5.3}$$

Figure 5.1 shows the fit of equation equation (5.2) to empirical data generated with varying d_{j_θ} (the fitted values are: $a = 0.98$, $b = 1.94$, $c = 0.48$ and $m = 0.11$). The generated data makes it clear that the achievable increase in mean temperature is limited. Equation (5.2) reflects that as its limit evaluates to

$$\lim_{d_{j_\theta} \to \infty} a \left(1 - \exp\left(-\frac{d_{j_\theta}^b}{c}\right)\right) + m = a + m \tag{5.4}$$

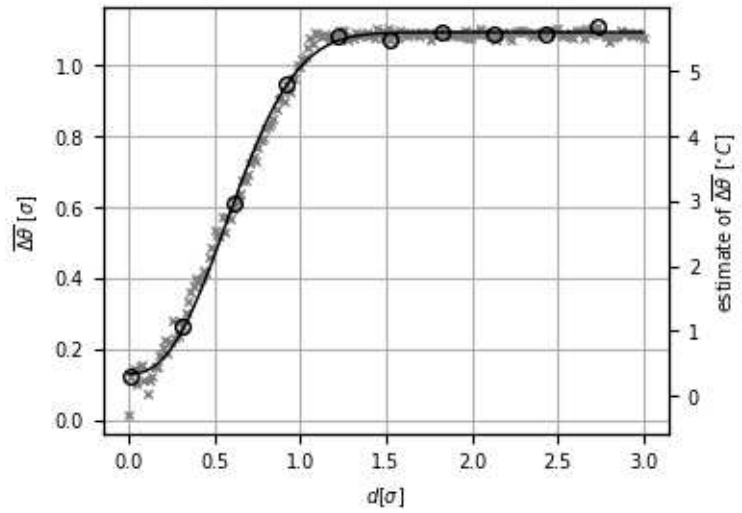


Figure 5.1: Calibration of the bias d for attaining a specific temperature increase. Line shows the function $\Delta\overline{\theta}(d)$. Circled crosses are values used for calibration of a, b and c. The scale in the measurement dimension on the right side of the plot is a rough estimation. The imprecision on the right scale is due to the fact that σ is changing with the *doy*.

5.3 Centrality Bias Within KNN Resampling

Section 5.4 shows that it is not strictly true that KNN resampling maintains the distribution of the source data in that the resulting distributions are narrower in their marginals. Before presenting these empirical results, the current section outlines the mechanism in non-mathematical terms that leads to narrower distributions.

Figure 5.2 shows a simplified schematic of why the resampling scheme results in narrower distributions. The simplifications amount to the following: (1) only one variable is considered, (2) only one previous time step is considered for identifying nearest neighbors and (3) the candidates are not weighted. The underlying assumption is that the source data has a significant autocorrelation[1]. Thus, the candidate distributions are biased towards the values of previously resampled time steps. Continually sampling in this way results in overlapping candidate distributions for central values, leading towards a resampled distribution with a lower variance. The extremes are limited to those in the data set, so that for extreme previous values, the candidate distribution is biased towards more central values.

This happens because the KNN resampling method approximates the conditional distribution $F(x_t|x_{t-1}, \ldots, x_{t-p})$ by identifying the nearest neighbors in the source data set and without building an abstraction of the source data values. Hence there is no extrapolation beyond the range of the observed values and a stronger effect of regression towards the mean.

[1] If there is no significant autocorrelation, the argument does not apply. But in this case the resampling scheme has no advantage over resampling serially independent values.

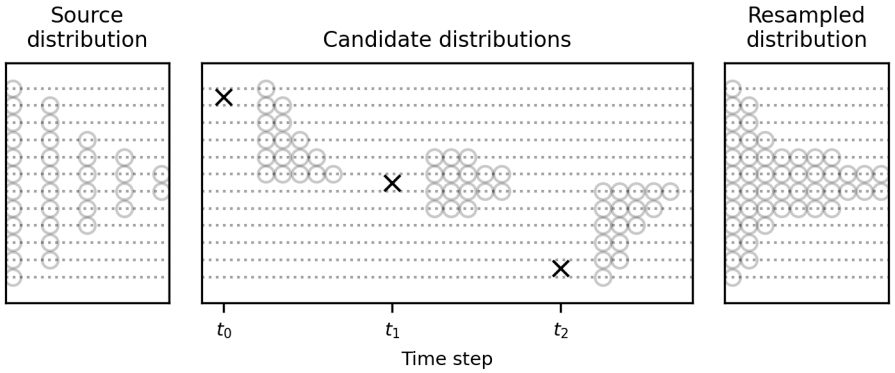


Figure 5.2: Schematic of a centrality bias in KNN resampling. Left panel shows the source distribution (frequencies correspond to a normal distribution). Middle panel shows the distribution of resampling candidates according to three different starting values (black crosses). Right panel shows the resampled distribution, arising from adding the circles from the middle panel.

In the following, I present an experiment of the effect of increasing autocorrelation in the source data set with the help of synthetic time series. For this, I generated synthetic source data sets with a first order univariate autoregressive model:

$$x_t = \rho x_{t-1} + \epsilon_t \tag{5.5}$$

where x_t– value at time step t
ρ – lag-1 autocorrelation
ϵ_t– noise term at time step t. Drawn from a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = \sqrt{1-\rho^2}$ in order to set $\sigma_x = 1$.

The experiment is setup in such a way that it is comparable to the analysis with measured data in section 5.4. The autocorrelations of the synthetic source data sets vary from 0.50 to 0.99. 40 source data sets of a length of 7671 (corresponding to a 21-year daily time series as in section 3.2.1) where generated. Asymptotically these time series have a mean of 0 and standard deviation of 1, empirically, however, not. In order to ensure standard normal marginals, I transformed the time series x by subtracting their sample mean and divided by their sample standard deviation. For each of these source data sets, I generated a resampled time series with a length of 365608 days (corresponding to a 1000-year daily time series) using q=87 neighbors (the square root of the number of time steps in the source data set, as advised in Rajagopalan and Lall (1999)).

The autocorrelation of the source data has a slight effect on the standard deviation of the resampled data (see figure 5.3). Considering more previous time steps in the resampling method (p) further reduces the standard deviation. In order to show that the effect is relevant measured time series, lag-1 autocorrelations of the variables in the observed time series are also shown. Temperature (θ) has the highest lag-1 autocorrelation in the observed time series

and is expected to show the largest decrease in standard deviation when resampled.

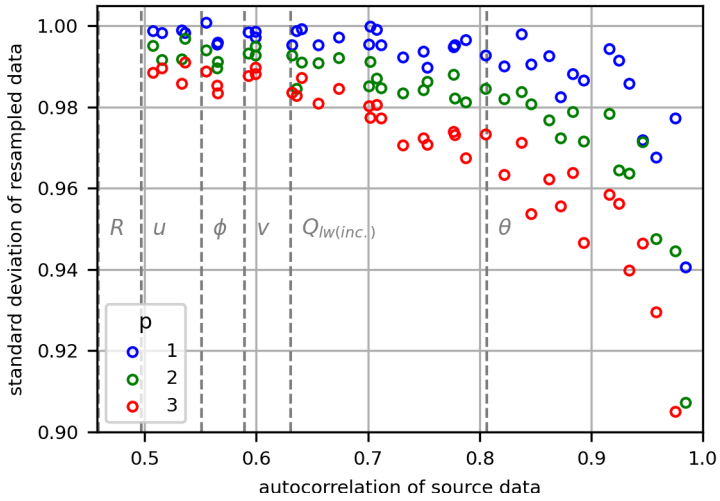


Figure 5.3: Standard deviation of resampled data declines with increasing autocorrelation of the source data. Vertical dashed lines show the lag-1 autocorrelation of the variables within observed data set.

The changing standard deviation alludes to having a more narrow distribution, which is shown in figure 5.4. Increasing source autocorrelation does lead to tighter distributions, with the most visible change in the tails. The middle and right panel in figure 5.4 underline again that the narrowing is systematically increasing with increasing source autocorrelation. Univariate statistics of the resampler will be revisited in section 5.4.2, where a measured time series is used instead of synthetic ones here.

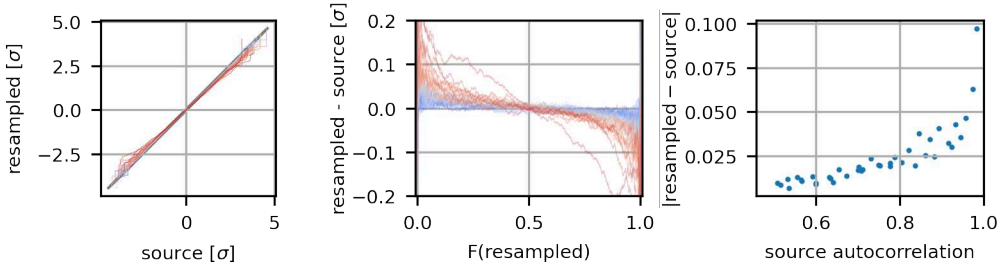


Figure 5.4: Changed distributions of resampled data due to higher autocorrelation in the source data. This uses only the data from the resampled time series with considering p=3 previous time steps. Colors range from blue to red with blue mapping to low and red to high autocorrelation. Left panel: ordered resampled data against ordered source data (qq-plot). Middle panel: deviations of $y-x$ from first panel over resampled quantiles. Right panel: mean absolute deviations from second panel over source autocorrelation.

A reduction of standard deviation is in line with results from T. A. Buishand and

Brandsma (2001), where precipitation and temperature at multiple sites were resampled. The fact that they resampled in the measurement domain hints that the bias towards smaller standard deviation is independent from the method used to deal with different dimensions of the variables, as they used weights for distance normalization whereas I used transformation to standard-normal marginals.

5.4 Output Characteristics of KNN-Resampling and Parametric Methods

The following comparison encompasses three types of analysis: resampled indices, univariate statistics and multivariate aspects. The output characteristics are presented in that order, as the selection of specific time steps is the reason for different univariate statistics and the multivariate analysis builds upon the results from the univariate case. The index-based analysis is done only for the resampling method, as it is not possible to be conducted for the parametric methods. Following the univariate statistics, a few multivariate extensions are discussed. The last subsection recapitulates some of the previous analyses for time series generated with a positive temperature bias.

The data used throughout this chapter is the single-station Konstanz data set described in section 3.2.1. It contains 7 variables (precipitation, temperature, incident long-wave radiation, short-wave radiation, relative humidity and east- and northward wind speed component). The comparisons are based on generated time series of lengths of 1000 years or the original 21 years of the source data, depending on whether the comparison is feasible with a longer time series or not. The number of considered neighbors was set to the square root of time steps in the source data (q=87), as suggested by Rajagopalan and Lall (1999).

Other than the setup described in the previous paragraph, configuration of VG is the same as for the example in chapter 3, as the same data set is used here. WeatherCop is fitted on the same data from just the Konstanz weather station. Both WGs use the regression method for estimating dryness probability in the present chapter.

5.4.1 Index-Based Analysis

> *"Those who fail to learn from history are bound to repeat it."*
>
> Attributed to George Santayana and Winston Churchill

The resampling method reuses concrete time steps from the source data set, thus it is possible to analyze the frequency of these time indices occurring in the resampled data set without strictly taking the values at these time steps into account.

Forgotten and Over- and Undersampled Indices

Equation (2.37) formulates a distance between resampled and source data chunks. The distances form the basis of the algorithm in section 2.3, in which only a proximity-based subset

of chunks is considered for selection. Because at each iteration step, only q nearest neighbors are considered, the question arises whether a subset of time steps is misrepresented in the resampled data set. Additionally, there might be time steps in the source data set that are unreachable and can *only* be sampled in the first p randomly selected time steps and are neglected in all later time steps.

Figure 5.5 shows the relative frequency of time steps in a large resampled time series (1000 years). Time steps from the source data do not have an equal probability of appearing in the resampled time series. A small proportion of the source time steps appear exceedingly often while other time steps have a low probability of being resampled. In the following, time steps above the horizontal, dashed line in figure 5.5, which represents the expected relative frequency of each time step under the assumption that every time step has the same probability of appearance, will be referred to as "oversampled" and time steps under the line as "undersampled". The resampled time series consists to 68.72 % of oversampled time steps.

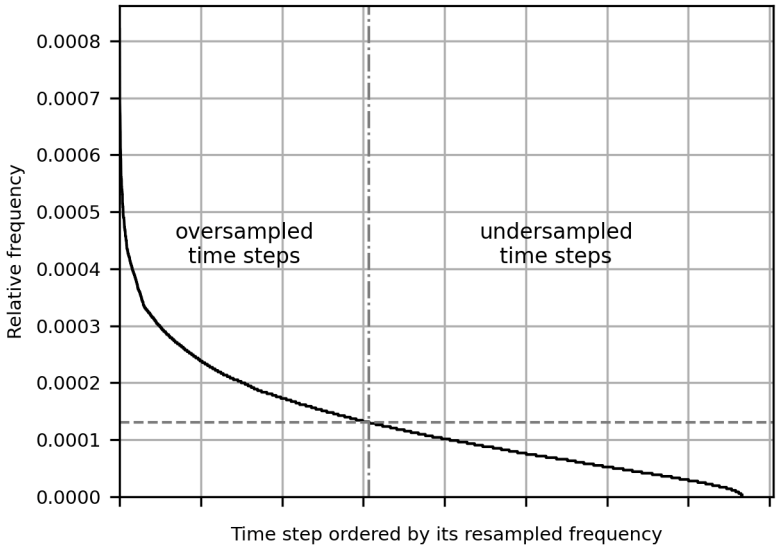


Figure 5.5: Over- and undersampled time steps. Relative resampling frequencies based on a 1000 year resampled time series ordered by number of repetitions. The gray, dashed line shows the expected relative frequency of repetitions based on the assumption that every value has an equal chance of occurring.

Out of the 7671 time steps from the source data, 6 (i.e. 0.08 %) do not appear in the resampled time series. It is unlikely that such a low number of "forgotten" indices has an effect on the overall resampled distribution.

Resampled Sequences and Repeated Traces

The distance measure as defined in equation (2.37) together with the additional selection bias towards the time step with the lowest distance (see equation (2.38)) favors selection of time series in the order they appear in the source data. A histogram of these sequences is

shown in table 5.1. 88.8% of the time steps appear in a different order than in the source data, followed by 9.1% of two- and 1.7% three-step sequences.

Table 5.1: Lengths of sequences in the resampled data that appear identically in the source data set.

Sequence length	absolute frequency	relative frequency
1	284900	0.887509
2	29335	0.0913832
3	5419	0.016881
4	1066	0.00332076
5	239	0.000744523
6	43	0.000133952
7	9	$2.80364e{-}05$

Resampling historical traces has also been studied by T. A. Buishand and Brandsma (2001) and found to be dependent on the number of considered neighbors q during the selection process. They derive a simple equation for the expected length of historical traces and an approximation to the expected longest resampled historical trace given the number of considered neighbors and the length of the resampled time series. Their derivation is based on the probability given to selecting the historical successor of a time step, which is the one given the highest probability in equation (2.38).

Beyond repeating sequences of the source data, there are also traces in a different order which are repeated throughout the resampled data. These are more common than repeated historical sequences. The resampled time series consists to 59.8 % of 2-tuples. This proportion declines for longer traces (3-tuples: 8.9 %, 4-tuples: 0.9 %, 5-tuples: 0.1 %). The longest reappearing episode has a length of 6 days.

Other Distributional Aspects

From the issue about unequally likely selected time steps raised in section 5.4.1, the question arises whether this selection bias targets *special* values. In other words, are certain kinds of values more likely to be selected and which kind of values are oversampled?

Looking at the empirical marginal distribution, it becomes obvious that the resampler overly selects data from the middle of the source distribution. This can be seen in figure 5.6, where oversampled time steps exhibit a more narrow and slightly left-shifted temperature distribution.

The effect is not as pronounced but still visible when comparing the empirical *cdf*s of the complete resampled and source data set without paying regard to over- or undersampled data. Figure 5.7 shows the expected narrowing of the resampled *cdf* relative to the source data. The narrowing holds for all other variables as well, but there is no consistent bias for low or high values.

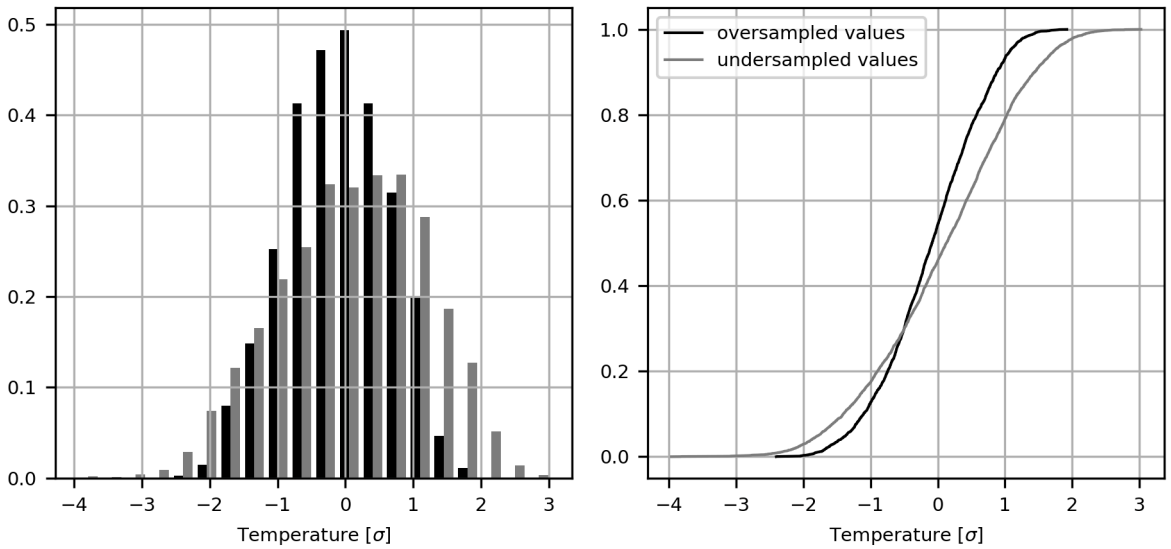


Figure 5.6: Distributions of over- and undersampled values. Histogram and empirical cumulative distribution function (*cdf*) of over- and undersampled values.

5.4.2 Univariate Statistics

"Only the fool believes that the highest mountain in the world will be the size of the tallest one he has ever seen."

Attributed to the Roman poet and philosopher Titus Lucretius Carus (ca. 99 BC – 55 BC)

An obvious comparison between the source and generated data is via simple univariate statistics. A graphical comparison of various univariate statistics is shown in figure 5.8 while the concrete numbers are presented in table D.1 in the appendix.

The mean of the resampled data set shows differences to the source data. However, regarding all variables, there is no consistent sign of the bias. The mean of the VAR-generated time series is close to zero[2]. VG and WeatherCop exaggerate the means slightly.

The standard deviation is smaller in the resampled data for all variables and almost equal in the time series generated with the parametric methods. The interquartile range (i.e. the distance between the lower and upper quartile) is, as the standard deviation, another measure for the spread of the data. In contrast to the range, it is smaller in the resampled time series, compared to the source data. The interquartile ranges of the time series generated with parametric methods are closer to the source data, yet still lower for all variables other than the wind speed components.

As the resampling method directly accesses the source data, minimum, maximum and range are maintained exactly. These statistics are, however, exceeded in the time series generated with parametric methods. Despite maintaining approximately the same marginal

[2]This appears to be an implementation issue, not a limitation of the method.

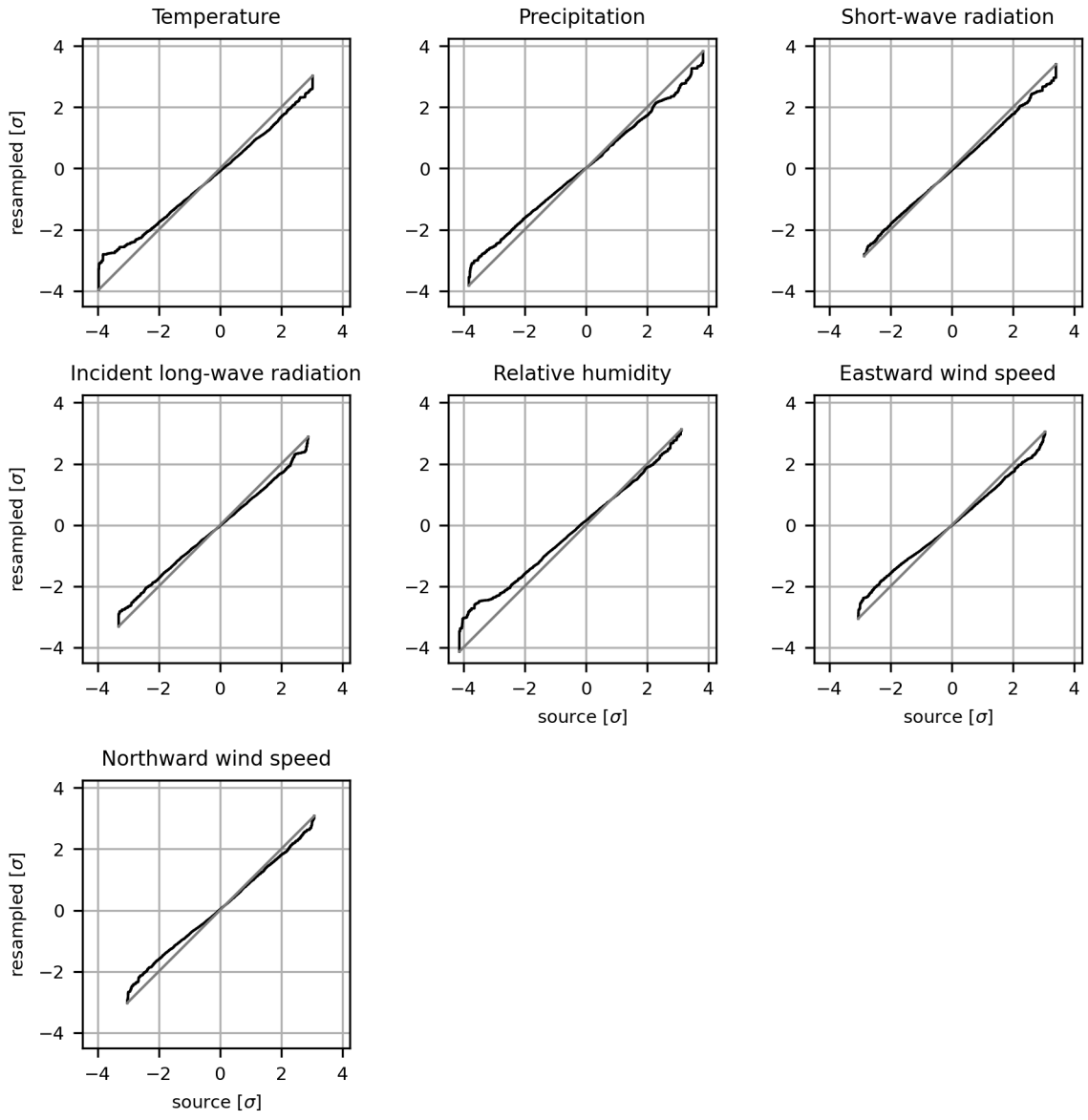


Figure 5.7: Narrowing effect of over- and undersampling on output marginal distributions. The sorted values of the source are plotted against the sorted resampled data.

distributions as the source data set, a larger sample would naturally contain values outside of the range of the source data. In other words, a 1000-year time series can be seen as having a lower variability if it has the same range as a 21-year time series.

The source data shows non-zero values of skewness, indicating an imperfect transformation to standard-normal marginals. The skewness of the resampled time series are also non-zero and exceed the values of the source data set. The VAR- and VG-generated time series show low absolute values of skewness, which is expected as the asymptotic value for skewness of these processes is zero. The WeatherCop-generated time series has non-zero values of skewness, however they are differing in sign and value to the observed time series.

The kurtosis values show a few interesting patterns. First, the source data is not exactly normally distributed in terms of this statistic, as it is significantly non-zero[3] (based on a two-sided test on 5% significance level) for all variables except the wind speed components. The negative kurtosis values (with the exception of the wind speed components) indicate less mass in the tails compared to a normal distribution. Kurtosis values in the resampled time series are significantly non-zero and a bit higher than the observed values. The kurtosis of the VAR- and VG-generated time series is very close to zero, hence more values fall in the tails of the distribution compared to the observed time series. Only the wind speed components have more values in the tails of the distribution compared to the VAR- and VG-generated series. The WeatherCop-generated time series contain variables with non-zero kurtosis values, but like with the skewness they do not match with those of the observed time series.

The difference in performance of the VAR-process for the first statistical moments to that regarding higher-order moments is in line with the fact that it asymptotically produces multivariate normal distributions, which have normal marginals. The mean and standard deviations can be fitted to data, whereas skewness and kurtosis are fixed (to 0 and 1 respectively). Of the shown methods, the resampler shows the best skill in reproducing the two higher-order moments.

5.4.3 Multivariate Analysis

Comparing multivariate aspects between the 5 data sets is a little less intuitive than with the univariate statistics. The analysis in the current subsection is based on convex hulls, data depth and entropy, all of which warrant a small explanation.

Multivariate Output Ranges in Terms of Hypervolumes

While narrowing of the univariate distributions of the resampled variables is easily shown, the detection and analysis of a changed multivariate distribution, however, requires more

[3] I computed kurtosis according to Fisher's definition which is the fourth normed statistical empirical moment minus three, so that the kurtosis of the normal distribution is 0.

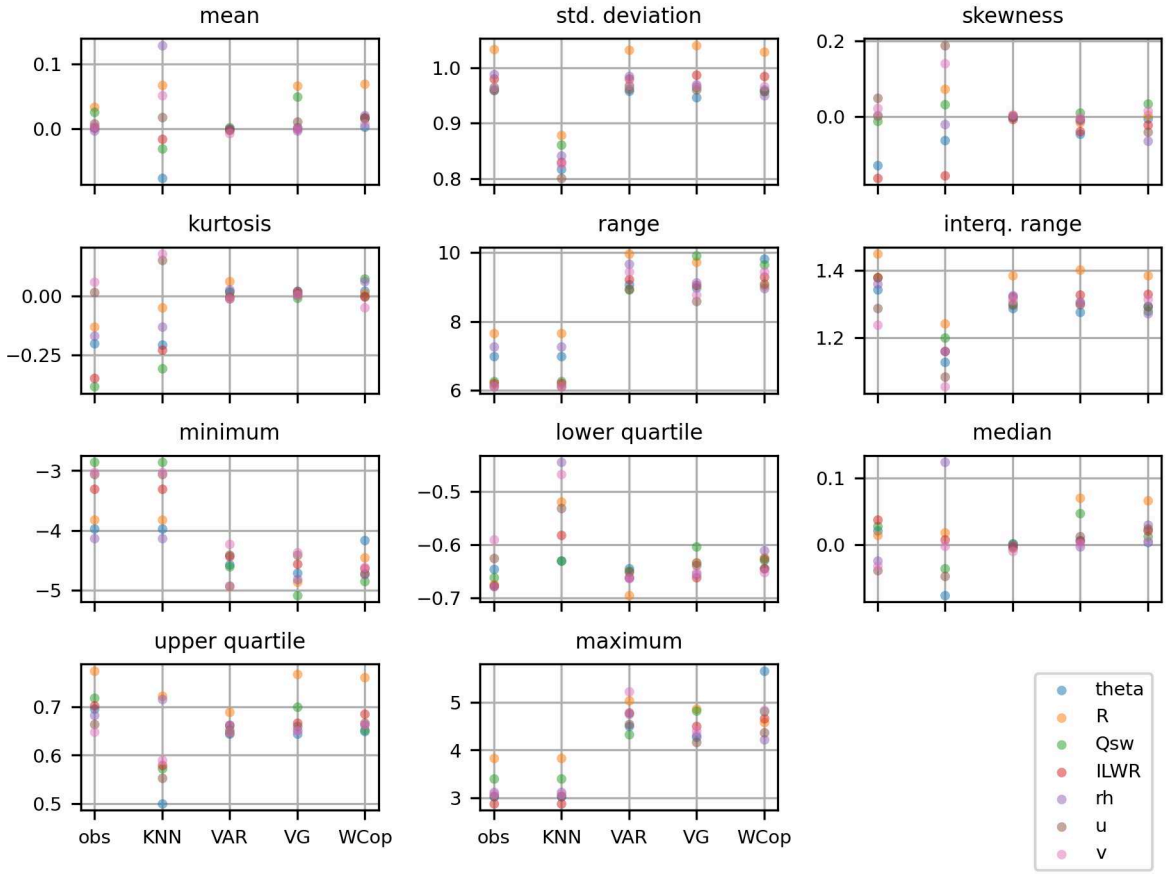


Figure 5.8: Univariate statistics of observational (obs) and generated data. KNN refers to the resampling method, VAR to VAR-modelling with serially independent Gaussian residuals and VG to VAR with phase-randomized residuals. WCop refers to the copula-based method. All generated time series have a length of 1000 years.

imagination. Even recognizing whether a point is outside of the range of the original data in the multivariate sense is not trivial, but possible by leveraging the convex hull of the data set. The convex hull can be seen as an extension of the univariate range in a multivariate sense. One useful definition of it is the following: given the dataset $X : x_1, \ldots, x_n$ (where each x_i is a point in k-dimensional space), the convex hull is the set of all points that are linear combinations of the form $\alpha_1 x_1 + \cdots + \alpha_n x_n$ where all $\alpha_i \geq 0$ and $\alpha_1 + \cdots + \alpha_n = 1$. From this definition it follows that all data points that include an extremum in one of their dimensions are part of the boundary of the convex hull. Setting $\alpha_j = 1$ for $j = \operatorname{argmax}(x_{1,k}, \ldots, x_{n,k})$ (and all other α_i's to zero), one obtains the point x_j which contains the univariate maximum value in the dimension k. The minimum values can be obtained with the argmin function respectively. Additionally, the set of points on the convex hull contains a lot more than the $2k$ univariate extremum points (in the data set used here, the convex hull is spun by 919 points).

Table 5.2 gives three points on the boundary as examples. The last column is an example that underlines that boundary points can appear to be located within the data "cloud", judging from its univariate ranks.

Table 5.2: Some examples for points on the convex hull. Shown here are some data points that form the convex hull of the source data set. Daily mean values in the measurement dimension and relative ranks with respect to their standard-normal values.

	Unit	21.03.1980	12.01.1987	11.12.2000
Precipitation	m	0.000	0.000	0.000
relative rank	-	0.907	0.202	0.241
Temperature	$°C$	2.667	-16.062	10.550
relative rank	-	0.202	0.000	0.995
Short-wave radiation	W/m^2	102.742	70.750	62.896
relative rank	-	0.402	0.981	0.982
Incident long-wave radiation	W/m^2	266.531	178.508	312.398
relative rank	-	0.271	0.000	0.951
Relative humidity	-	0.926	0.810	0.663
relative rank	-	0.964	0.228	0.006
Eastward wind speed	m/s	-0.402	0.756	1.305
relative rank	-	0.152	0.566	0.678
Northward wind speed	m/s	-0.317	-0.524	0.718
relative rank	-	0.271	0.200	0.703

If a point lies outside of the convex hull of a data set, it is more extreme as the points in the given data set in the sense that the combination of the values involved is more extreme than observed ($\alpha_1 + \cdots + \alpha_n \geq 1$).

Where points on the convex hull can be seen as multivariate extreme values, the multidimensional volume ("hypervolume") of the convex hull is in part akin to the univariate range. It is not a trivial extension from the range, as dependence is a factor when interpret-

ing the hypervolume of the convex hull. Data sets with a high linear dependence will have narrower shapes than ones with low linear dependence.

The volume of the resampled data set is $2234.0\sigma^7$, which is a little less than for the observed data ($2234.3\sigma^7$), indicating that not all boundary points from the source data set are resampled. The parametric methods generate data that spans a much higher volumes, with $18106.6\sigma^7$ for VAR-based generation, $14458.0\sigma^7$ for VAR with phase randomized residuals and $31396.5\sigma^7$ for WeatherCop.

78.9% of the VAR-generated points fall within the hull of the observed data set and 78.8% and 75.5% of the points generated by VG and WeatherCop, respectively. Naturally, all of the points acquired by resampling fall within the hull of the source data set.

As the convex hull of a data set can be seen as a description of its most extreme values in a multivariate sense, the comparison between long time series generated with parametric methods to a short observed time series, mostly shows that more extreme values can be produced in contrast to the nonparametric method. Additional insight can be gained by comparing volumes of inner layers, obtained by repeatedly identifying the points of the convex hull, removing them and finding the next hull. This process is called convex hull peeling (Eddy, 1982). I generated time series for each model with the original length of 21 years, as the difference in length complicate the interpretation of the peeled volumes. For each data set I peeled until all points were gone or the number of points are not sufficient to form a hull in the 7 dimensions. Figure 5.9 shows the layer volumes over the proportion of remaining points in the data set. The parametric models lie a little above the line of the observed data set, while the resampling approach is consistently below it. The logarithmic y-scale on the right panel reveals that the inner layers of the parametric models slightly overestimate inner layer hull volumes.

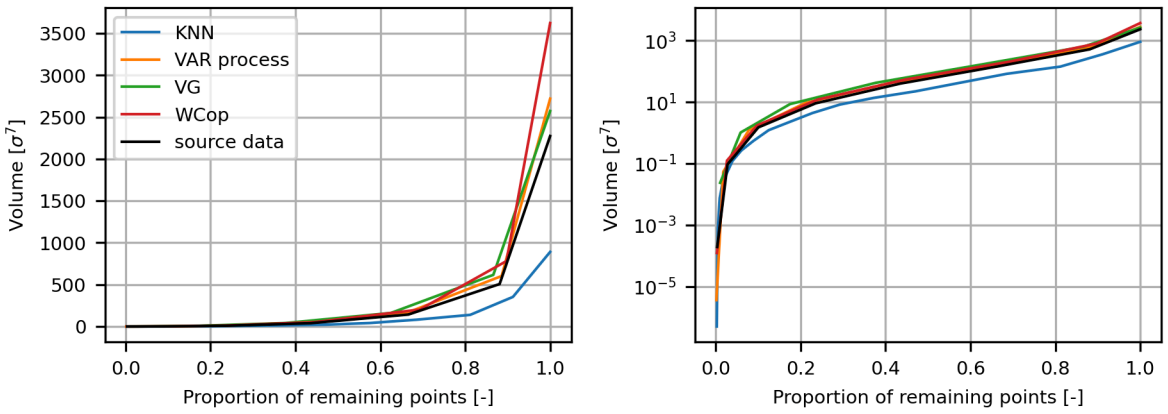


Figure 5.9: Convex hull peeling. Inner layers would be on the left of and outer layer (the convex hull) on the right the x-axis. To not complicate the comparison between observed and simulated values, the length of the simulated series is as long as the observed data set (21 years).

Using the remaining proportion of the peeled data, the different models can be compared in a similar manner as with the interquartile range in the univariate case. An equivalent

measure is to remove 50% of the data by peeling and compute the volume of the remaining points. The concrete values for this approach are: $68.4\sigma^7$ for the observations, $27.4\sigma^7$ for KNN, $82.8\sigma^7$ for VAR, $101.1\sigma^7$ for VG and $85.2\sigma^7$ for WeatherCop, continuing the pattern, that the parametric methods overestimate spread, while the non-parametric method underestimates it.[4]

It is to be expected that the multivariate range of a data set changes as time progresses. In order to investigate this change and to contrast it with the resampling method and the VAR, I repeatedly split the 21-year data set into calibration and validation set. I used 4-year continuous time periods as validation sets and the rest of the data set (17 years) for calibration and evaluated all such possible full-year periods. The generated time series are as long as the validation period. As comparison, the overlap between the convex hull of the validation set with respect to resampled and VAR-generated time series, as well as the calibration set is shown in figure 5.10. In all considered cases, the convex hull of the calibration set does not fully include the validation set, despite the fact that the validation period is shorter than the calibration period (4 versus 17 years). Following the previous discussion on the limitations of the resampling method, it is not surprising to see the resampled time series showing the lowest amount of points within the convex hull of the validation period. This happens because it only uses values from the calibration set with a bias towards those from the middle of the calibration set distribution. The VAR-generated time series shows a higher percentage of overlap than the resampling method for all of the 17 run configurations.

Figure 5.11 shows the hypervolumes of convex hulls for the observed, VAR-generated and resampled data set in the split-sampling experiment. The VAR-process generates data that is consistently more spread out than the validation sets. The volume of the resampled time series is larger than the validation set in only 3 out of the 17 cases, with an underestimation being more likely than an overestimation. The multivariate range of the VAR-generated data is consistently larger than that of the resampled time series and the validation period. This is an indication that the multivariate Gaussian distribution represents an imperfect approximation to the observed data, as it is spread out too far. The fact that the VAR method showed a higher percentage of overlap in figure 5.10 is most likely due to its highly spread out output.

Data depth

I showed in section 5.4.1 that the KNN resampling method overly selects values from the middle of the marginal distributions. It requires different tools to determine whether this tendency towards the distribution's center, that the resampling method shows in the univariate case, extends to the multivariate case. The approach I take here is using data depth, which is a measure of how central a value or vector is with respect to a data set (Tukey, 1975). A conceptually easy way to express data depth is the so called halfspace depth. Through a point for which the halfspace depth is to be estimated, one finds the hyperplane that separates the dataset in the most unequal way possible. One then counts the proportion of points

[4] As peeling is not guaranteed to provide exact numbers for a given reduction in data size, these values were obtained by linear interpolation.

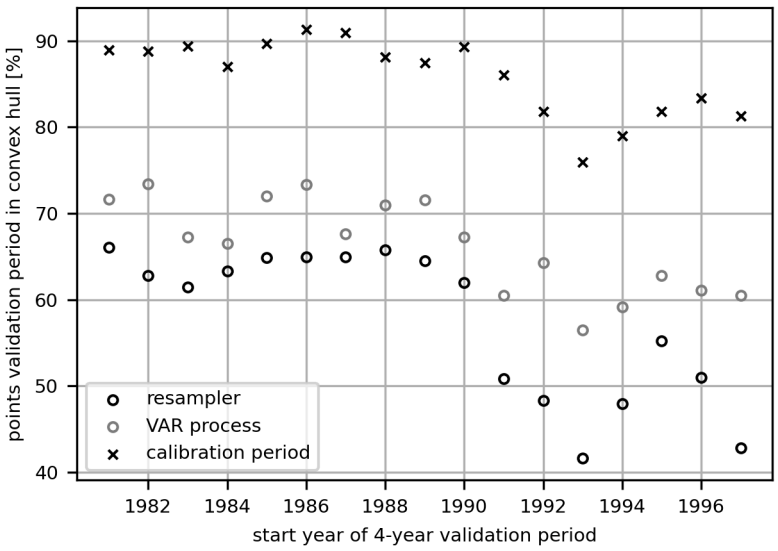


Figure 5.10: Overlap of validation and convex hull of generated data in a split-sampling experiment. The position on the y-axis gives the percentage of points that fall in the convex hull of the validation set (e.g. $\approx$ 72 % of the time steps from the VAR-generated time series fall within the convex hull of the validation period from 1981–1984). The crosses show the overlap between validation and calibration data of the source and are thus an indicator of how much the convex hull changes between those periods.

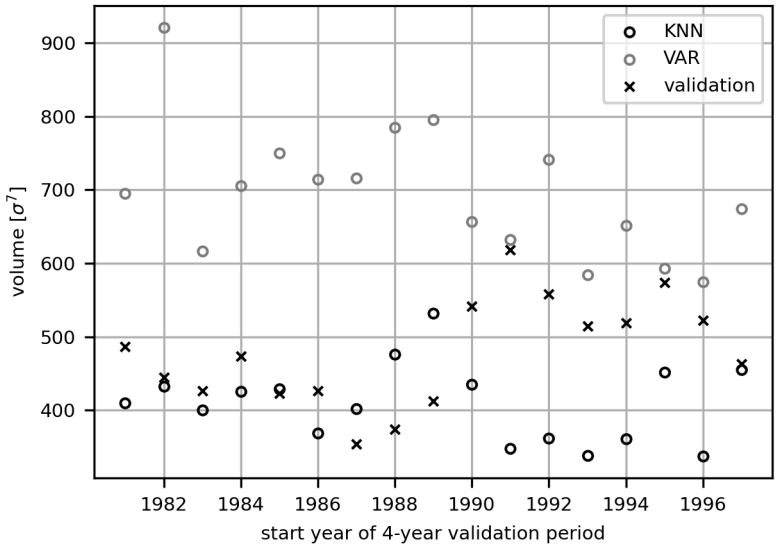


Figure 5.11: Hypervolume of convex hulls in a split-sampling experiment. Position on the x-axis marks the start of the 4-year validation period, while the rest of the years were used as calibration.

on each side of the plane and assigns the smaller count to the data depth of the given point. For example, through a point on the convex boundary of the dataset, a plane can be found that has all data points on one side. The depth of such a boundary point is 0. For a very central point (and a rotationally symmetric data set), only planes that half the data set might be found. The depth of that point is then $n/2$ (with n being the number of points in the data set).

The data depth can also be used to infer how deep one data set is within a reference data set. For this endeavour, planes at the points of the to-be-checked data set are sought that separate the reference data set in the most unequal way.

The data depth is an empirical measure and thus devoid of any assumption of an underlying theoretical distribution. The use of planes in estimating halfspace depth, however, makes it unsuitable to identify boundaries of concave shapes.

The mean depth of the source data is 174.1, while the mean depth of a 1000-year resampled data set (with respect to the source data) is 266.3, indicating that the resampling method favours data that is more central in a multivariate sense.

The mean depth of a 1000-year VAR-generated data set with respect to the source data is 147.8, which is lower than the mean data depth of the source data itself (174.1). VG generates data with a similar but slightly lower depth than pure VAR-processes with a mean of 145.4. WeatherCop is very similar in terms of data depth as well (mean depth with respect to the source data: 147.4). The comparison using simple statistics in section 5.4.2 illustrated that the parametric models generate values outside of the univariate ranges of the source data. Hence, in the multivariate perspective, part of the generated points lie outside of the convex hull of the source data and are assigned a depth value of 0, causing the mean depth to decline.

Plotting the depth values as *cdf*s shows that the parametric models are very similar to each other and all generate slightly less deep values than the source data. The *cdf* of the KNN resampling output shows again that the selected points are deep within the source data set.

The higher centrality of the resampled data follows from the very different distribution of data depths of the over- and undersampled values (see figure 5.13). The mean depth of the oversampled time steps is 337.3 and of the undersampled time series 110.3. 740 points from the set of oversampled time series are points that form the convex hull of the source data, as opposed to 6378 of the undersampled time steps.

Entropy

The fact that resampling reuses the values from the source data – and that possibly more than once – imposes a peculiar orderliness on the generated time series, which potentially affects its information entropy. The entropy of a multivariate data set is (Shannon, 1948):

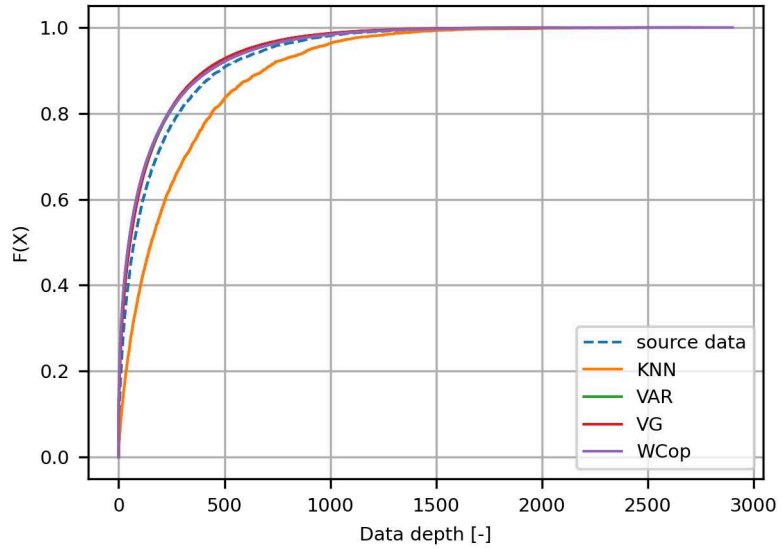


Figure 5.12: *Cdf*s of data depths of observed and simulated data. The depth values of the simulated series are all with respect to the observed data. Dashed line shows the non-exceedance probabilities of the observed data depths. The parametric model's lines (VAR, VG and WeatherCop) overlap on the left side of the dashed line.

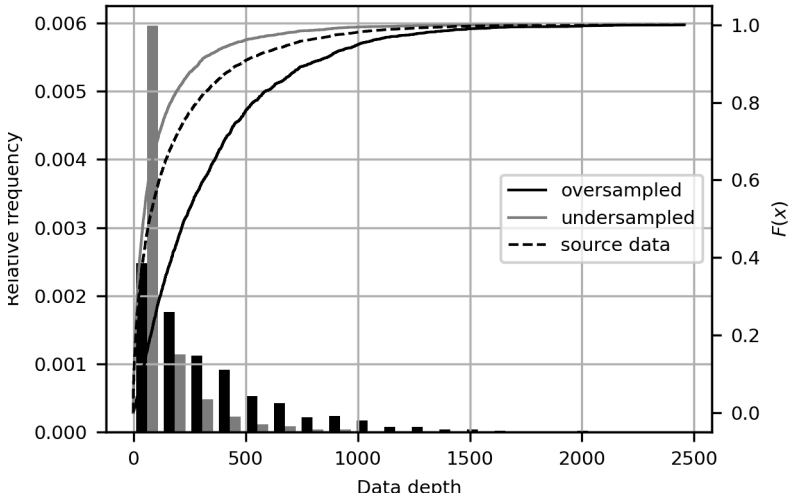


Figure 5.13: Histogram and empirical *cdf*s of data depths of over- and undersampled time steps. Dashed line shows the *cdf* of the source data.

$$H = -\sum_{i=1}^{L} \dots \sum_{l=1}^{L} p_{i\dots l} \log_2 \left(p_{i\dots l}\right) \tag{5.6}$$

where L – Number of bins per dimension
$i \dots l$ – index for bins in the respective variable dimension
$p_{i\dots l}$ – probability (here: relative frequency) of a point falling in the bin with index $i \dots l$

In order to calculate the entropy of a data set, one can use relative frequencies as estimates for $p_{i\dots l}$. The entropy is at minimum if all data is located within one bin and at maximum when it is spread out evenly over all bins. Hence, the entropy can be interpreted as a multivariate measure of uncertainty inherent in the distribution (Singh, 1997). The entropy in the above formulation is a measure of information content in a multivariate sense, but this interpretation is misleading in the present context, as the multidimensionality of the resampled time series is only superficial (the values of the different variables are directly determined by the resampled index). Nonetheless, determining how different the level of order appears in the output of the models at hand is the aim of this section.

Figure 5.14 shows how the entropy of the source data set and the generated time series is changing with the length of the time series. The bin edges are held constant for the three different data sets in order to compare them on the same scale. The resampling method shows the smallest entropy, followed by the observed data and then the parametric models. All entropy lines show large changes at the beginning and smaller changes the longer the data set is. However, the change of entropy of the source data does not approach zero towards the end, hinting that entropy is not a stationary aspect of the data. The entropy of the resampling method approaches an upper limit that is lower than the entropy of the full source data. By comparison with the entropy of a random sample of the source data (without maintaining the autocorrelation), the reduction of entropy within the KNN framework becomes even more apparent. The time series generated with parametric models approach the entropy of the multivariate Gaussian distribution[5]. WeatherCop maintains order the longest of the parametric models before reaching the Gaussian entropy. The fact that the KNN-resampler does not select values from the source distribution at random results in an order that is too high. Asymptotically, the VAR process generates time series with a multivariate Gaussian distribution, which is underlined by the proximity of the entropy to the Gaussian sample. In comparison with the data it was fitted on, this contains too much entropy.

5.4.4 Consequences of Increased Bias

A bias in the form described in section 5.2.2 constrains the selection process to data close to the target mean. This necessarily reduces the variability of the output. Figure 5.15 shows the decline of unique time steps with increasing bias. Based on a 1000-year resampled time series, with an increase of 4 $^\circ C$, 64.0 % of the source data is ignored. The inverted "S"-shape of the curve indicates that the change from a relatively high variability to a low variability

[5] $H = \frac{K}{2}\left(1 + \ln(2\pi)\right) + \frac{1}{2}\ln(|\Sigma|)$, where K is the number of variables and Σ the covariance matrix.

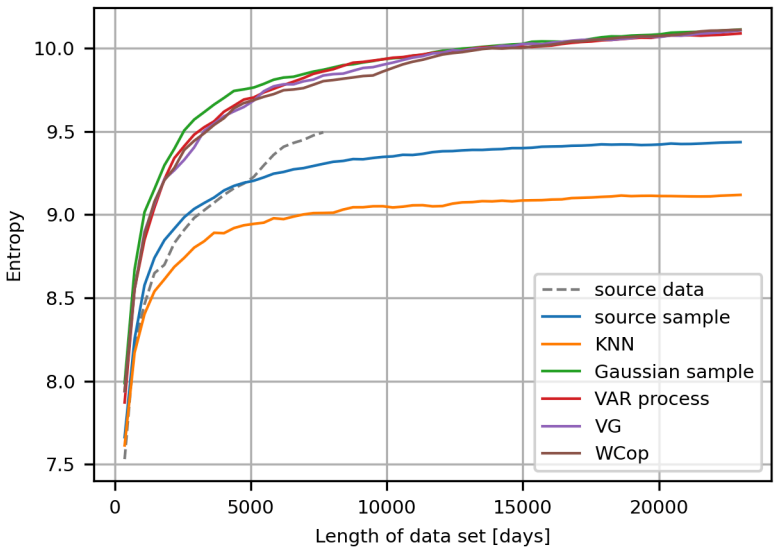


Figure 5.14: Entropy with increasing length of generated time series. The Gaussian and source sample where each generated by serially independent drawings. The Gaussian sample was generated with the mean vector and covariance matrix of the VAR-process.

regime happens rapidly as the temperature change exceeds $2\ ^{\circ}C$. As the number of unique time steps tends towards one[6], the method ceases to be stochastic. As the bias increases, the proportion of over- to undersampled time series increases as well, as more of the same time indices are selected repeatedly (see figure 5.16). Figure 5.17 shows *cdf*s of data depths for different changes in mean temperature. The data depth decreases with increased bias as less values from the middle of the multivariate distribution are made available during neighbor selection. As a last example, the decrease in variability under increased temperature bias can be seen in figure 5.18, where the volume for the resampled data sets vanish while the VG maintains a stable high spread.

[6]The minimum number of time steps in this experiment is actually 1381, because different time steps with maximum temperature values are available throughout the year and the composition of the neighborhood is *doy*-dependent (see section 5.2.1).

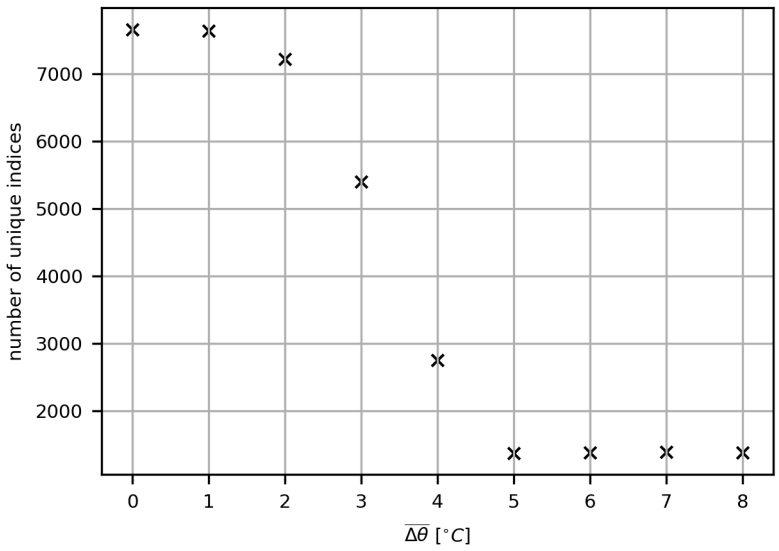


Figure 5.15: Declining variability with increasing temperature bias. Number of unique indices over change in mean in the measurement domain. The full Constance data set was used, which contains 7671 days (21 years). The lengths of the resampled time series are 1000 years.

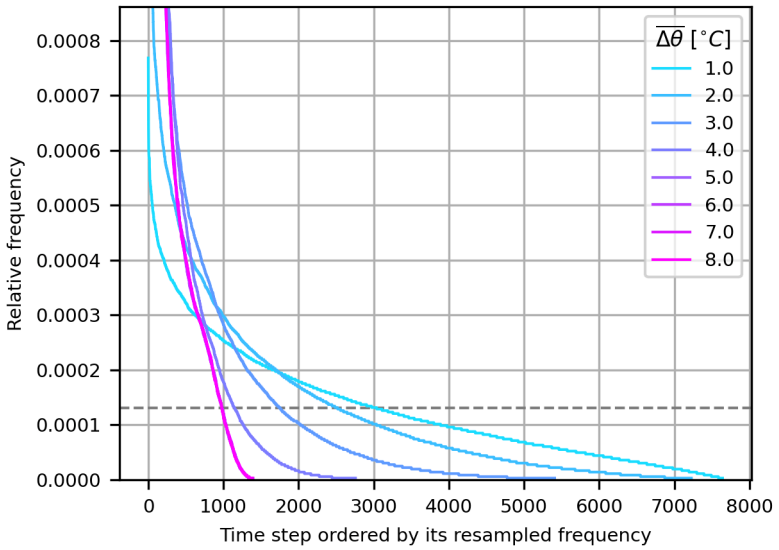


Figure 5.16: Over- and undersampling under increasing temperature bias. Relative resampling frequencies of resampled time series under bias, ordered by number of repetitions. The gray, dashed line shows the expected relative frequency of repetitions based on the assumption that every value has an equal chance of occurring.

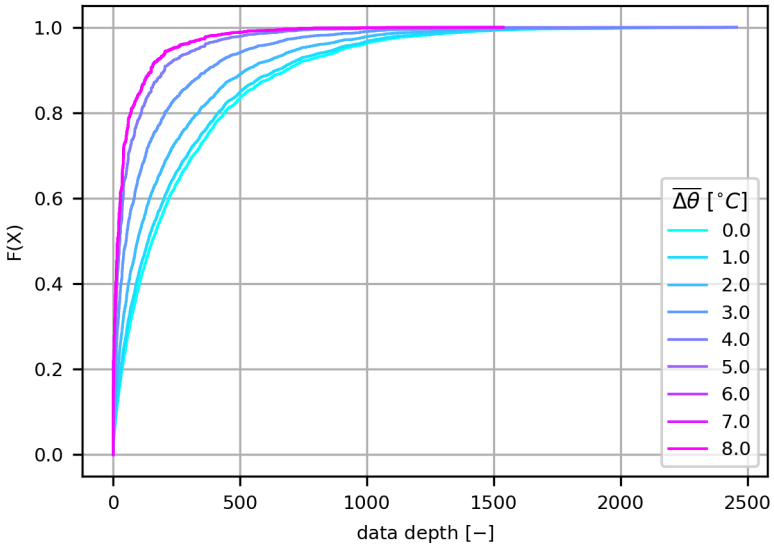


Figure 5.17: Data depth distributions in the form of *cdf*s for different changes in means. The increasing bias favours selection of low-depth values.

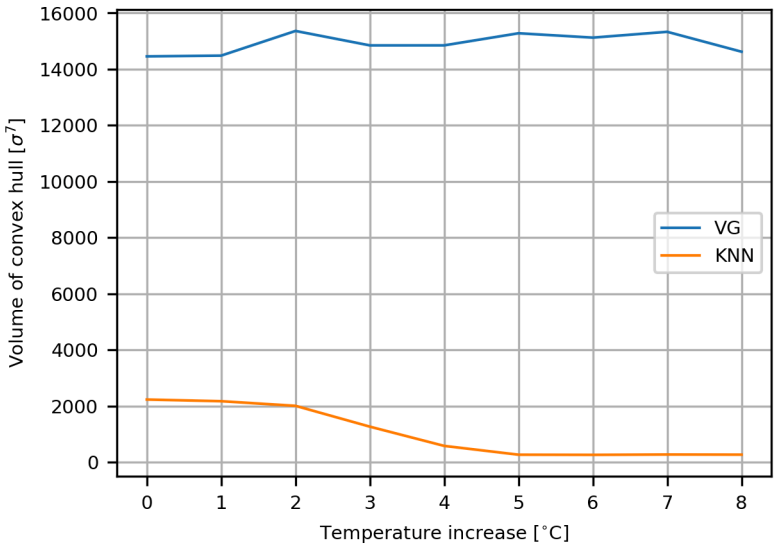


Figure 5.18: Volume of convex hull under increased temperature bias.

5.5 Generating Values Outside the Historical Range

The foundational paper on nearest-neighbor bootstrapping in hydrology by Lall and Sharma (1996) already suggested a scheme based on resampling residuals in an additive fashion, which is able to generate values outside the historical range. Sharif and Burn (2006) add Gaussian noise terms with variance dependent on the distribution of the candidate set. This method requires additional precautions to not generate values outside of physical ranges (e.g. negative precipitation). The latter issue is handled in the work of **Leander2009** by using a multiplicative instead of an additive scheme or by King, McLeod, and Simonovic (2015) who opt for using log-normally distributed perturbations which are strictly positive. Lee (2017) adds random perturbations to "intentionally biased bootstrap" precipitation data to match sample variances while downscaling general circulation model (GCM) generated climate warming scenarios. Another adaptation of KNN resampling for generating time series following a climate change scenario is described in Lee and Singh (2019). The authors propose a combination of KNN resampling and genetic algorithms. In the context of multi-site resampling, selected feature vectors are subjected to a "crossover", whereby precipitation occurrences at some stations are exchanged with those of a similar feature vector from the source data producing a new spatial pattern. An additional "mutation" is allowed, where elements of the feature vector are replaced by an arbitrary feature vector from the source data set. In the context of a precipitation generator for climate scenarios, Agilan and Umamahesh (2019) proposes a perturbation of extreme precipitation in the form of a qq-transformation to a generalized pareto distribution after each KNN-based selection step.

Without having tested these approaches, they are likely to increase the entropy towards that of measured data, increase the volume of the convex hull and might help in breaking out of repeated traces. On the other hand the dependency structure (in terms of inter-variable correlation, auto- and cross-correlation) is weakened, because the perturbations are added to each variable or location (in the case of Lee and Singh (2019)) independently. The "crossover" and "mutation" steps of Lee and Singh (2019) likely break spatial dependence and while they generate different configurations than the past, their values do not leave the historical range.

5.6 Summary and Discussion

Drawing a random sample from source data with the additional aim to maintain its autocorrelation comes at a price. The reduction of output possibilities within the KNN resampling framework necessarily restricts the variability of the "generated" time series. A trivial sign of this restricted variability is the fact, that the values remain within the univariate range of the source data set. The slightly reduced standard deviation is a first hint that the variability is somewhat reduced. The reduction of the standard deviation is very small, but shows that the resampled marginal distributions do not remain as they are in the source data. I showed that the reduction in standard deviation is dependent on the autocorrelation of the source data.

The KNN-resampling method tends to overly select values from the middle of the univariate distributions as well as the multivariate distribution.

Implementing resampling of multiple variables at once is easy but it introduces subtle deficiencies. This is because despite having multiple output variables, the resampled resource (the time-step) is one-dimensional, but the available hypervolume for observational data to be filled grows exponentially with the number of dimensions. The basic assumption that the source data is the population thus becomes quickly unfeasible with the inclusion of more resampled variables. A similar argument can be made when applying a resampling scheme for multiple sites at once.

In terms of the multivariate range and entropy of output, observed statistics tend to lie between the resampling method and the parametric models. The resampling method's output is too ordered. The VAR-process generates multivariate Gaussian distributions, which is a crude approximation to the given data even after transforming the marginals back to the measurement domain. These multivariate characteristics can be easily overlooked when the focus is set on statistics of individual variables or their marginal distribution.

Biasing the selection process in KNN resampling exacerbates issues with variability in terms of staying within a subset of historical values. This results in reduced variability in terms of oversampled indices and also in a multivariate sense measured in terms of smaller convex hull.

Staying within a subset of historical data is problematic especially when conducting Monte-Carlo-style ensemble modelling, where the aim is to generate a multitude of plausible outcomes. The presented parametric approaches overshoot in terms of variability but are capable of generating values outside of historic ranges and value combinations while maintaining plausibility by relying on dependencies present in the source data set.

6 Application in Modelling the Climate Impact on Lakes

"The purpose of computing is insight, not numbers."

Richard Hamming, 1962

Lakes are affected by climate change as their processes are driven in part by meteorological conditions. Upper-layer water temperatures of lakes are increasing in Europe (Woolway, Weyhenmeyer, et al., 2019) and world-wide (O'Reilly et al., 2015). Many are projected to mix less frequently (Woolway and Merchant, 2019). At the same time, they are important fresh-water resources, which makes it prudent to estimate possible impacts of the ongoing climate change on them.

This chapter presents applications of the Vector-Autoregressive Weather Generator (VG). While the **Weather** Generator based on Phase Randomization and Vine **Cop**ulas (WeatherCop) offers the same interface to generate scenarios, it has not been used in lake-oriented studies yet.

6.1 Direct versus Indirect Validation

The applications described in the following subsections serve an important purpose: indirect validation. In contrast to direct validation which compares Weather Generator (WG) input to WG output, indirect validation tests whether the aspects of the observations that are relevant for the application are present in the synthetic data even if deficiencies were identified during direct validation (Dubrovský, Buchtele, and Žalud, 2004). Compared are not observed and generated meteorological data, but impact model output, driven by observed meteorological and synthetic data.

This approach is more cumbersome than direct validation, as the run of another model is needed – furthermore the indirect validation has to be repeated as often as a different impact model is used. This added cost provides the benefit of a more rigorous test of applicability. There exists no complete description of an observed data set in terms of aggregate statistics, hence the limitation of direct validation.

6.2 Lake Constance

Lake Constance is located just north of the central European Alps and borders Germany, Switzerland and Austria. Its size (48 km^3), depth ($>$ 250 m) and good trophic state (olig-

otrophic with phosphoric concentration around 10 μg L^{-1}) make it a valuable drinking water resource for 5 million people in Germany and Switzerland.

6.2.1 One-dimensional Lake Modelling

This application was published in Schlabing et al. (2014) and is described here only briefly. The collaboration with lake modellers that lead to this publication determined major aspects of VG's design, such as the choice of variables (with an initial exclusion of precipitation), the method for scenario generation and the overall avoidance of abrupt changes in parameter values (achieved by wide use of smoothing annual parameters with Fast Fourier Transform (FFT) tools). I deemed the latter feature necessary for evaluating the influence of scenario changes to dates of stratification and plankton blooms.

One-dimensional lake models simulate the water body in terms of homogeneous, vertically stacked layers. In the case of the DYnamic REservoir Simulation Model (DYRESM), these layers have a Lagrangian structure. The run time of one-dimensional lake models is short enough to enable the calculation of ensembles, which brings with it the benefit of having a range of plausible outcomes that can be interpreted as an estimate of uncertainty.

Model Setup

VG was fitted on air temperature, short-wave radiation, incident long-wave radiation, relative humidity, eastward and northward wind speed component from the years 1980 to 2000 measured by the German Meteorological Service (DWD) at the Konstance measurement station. The 500 realizations per scenario were used as boundary conditions for the lake model system DYRESM-Computational Aquatic Ecosystem DYnamics Model (CAEDYM). Tributary discharge and precipitation time series were kept constant throughout the realizations and scenarios.

Scenarios

Table 6.1 gives an overview of the four "what-if"-scenarios we defined.

The *stale* scenario serves two purposes: (1) it is an indirect validation which aims to determine whether the properties of the observed meteorological time series that are relevant for the studied impacted variables are reproduced well and (2) it provides context for the other three scenarios.

The three scenarios that contain a change to statistics (*hot*, *spicy* and *hot & spicy*) are balanced scenarios, which means that the temperature change is accompanied by changes in the other variables according to the linear dependence structure in the source data set (see section 3.8). This stands in contrast to other studies in lake research where a change is applied only to temperature (Trolle et al., 2011; Kupisch et al., 2012; Schwefel et al., 2016).

Table 6.1: Climate scenarios generated by VG for Lake Constance study. Adapted from Schlabing et al. (2014).

		increase in mean air temperature	
		– (C-scenarios)	4 °C (F-scenarios)
climate variability	unchanged (T-scenarios)	T_C – *stale* "current temperatures"	T_F – *hot* "future climate"
	increased (V-scenarios)	V_C – *spicy* "current temperatures with higher variability"	V_F – *hot & spicy* "future climate with higher variability"

Scenarios *spicy* and *hot & spicy* contain a kind of variability enhancement that was designed to have an influence on the lake. Instead of a time-invariant multiplicative change, which would succeed in increasing the variance of the data but not overcome the water's thermal buffering effect, episodes of alternating cold and warm temperature were added. These are applied in a saw-tooth-like fashion, with episode lengths drawn from an exponential distribution with a mean of 7 days and magnitudes drawn from a normal distribution with zero mean and standard deviation of 5 °C.

These scenarios were not meant to be predictive, but rather serve as a "stress-test" to see how a large system like Lake Constance would react to big changes in meteorological forcing. The *hot* scenario is however still within the range of IPCC (2007) projections (RCP8.5 has a 5 to 95% range of 2.6 °C to 4.8 °C for the period 2081 – 2100 relative to 1986 – 2005).

Results

Figure 6.1 shows results of the *stale* scenario in terms of input and lake model output while also showing a reference lake model run driven by the observed meteorological time series. The reference run has a similar appearance as the *stale* realizations in all panels, giving confidence in the idea that VG forms a reasonable stochastic abstraction of the observed time series.

In order to study an important ecological aspect, we decided to evaluate the impact of these climatological changes on the timings of spring phytoplankton blooms. Concretely, three output parameters were defined and evaluated:

doy_{spring}, *i.e. start of spring:* first *doy* since 1st of January at which the sum of daily mean air temperature surpassed 300 °C.

doy_{start}, *i.e. beginning of stratification:* first *doy* with a water temperature difference between 0 and 20m depth of 1 °C.

doy_{chla}, *i.e. start of phytoplankton bloom:* first *doy* at which mean total chlorophyll concentration in the upper 20 m is larger than 3 $\mu g L^{-1}$.

Figure 6.2 shows how these cardinal dates are related in the scenario output. The F-scenarios shift all cardinal dates to earlier times in the year. While the median doy_{spring} is April 29 and 27 for T_C and V_C respectively, it is at April 13 and 9 for T_F and V_F. The standard deviation of the cardinal dates is increased in the V-scenarios but the strength of that change decreases from doy_{spring} to doy_{start} and again from doy_{start} to doy_{chla}. Nonetheless, adding variability motivates more earlier than late stratification, clearly visible in figure 6.2. The timing of phytoplankton blooms is less sensitive to the variability change in the V-scenarios.

6.2.2 Three-dimensional Lake Modelling

This application is a short summary of some aspect of the thesis of Eder (2013), which are relevant in the current context. Her simulations complement the ones done in Schlabing et al. (2014) in that three-dimensional modelling offers a more detailed look, e.g. into mixing regimes. Furthermore, the simulations here consisted of 19 years, allowing the author to make statements about possible medium-term developments and mixing frequencies. Despite being supplied by single-site weather data, the 3D lake model was able to reproduce the variables deemed important for the study.

Model Setup

Eder (2013) used VG output as input for the three-dimensional lake model Estuary, Lake and Coastal Ocean Model (ELCOM) which was coupled to the ecological model CAEDYM in order to simulate phytoplankton populations. The data to fit VG on was the same as for the 1D-study in section 6.2.1.

Scenarios

Eder (2013) leveraged multiple features of VG in order to gain insight into Lake Constance's sensitivity towards climatic changes. Several sets of scenarios were thus defined:

stale: undisturbed statistics. Serves as a reference to the other scenarios.

hot4: temperature mean is increased by 4°C.

hot5: temperature mean is increased by 5°C.

spicy: includes enhanced hot and cold episodes.

hot & spicy: Combines *hot* and *spicy*.

seasonality: hot and warm winters and summers respectively.

Especially the changes in seasonality consist of very specific and intentional changes made by the modeler, which I made possible in VG by implementing the possibility to supply a "climate signal". This time series given in the measurement domain of the primary variable

(here: air temperature) is converted to changes in the theoretic mean vector of the Vector-Autoregressive (VAR) process as described in section 3.8.1. It is especially apt for supplying a smooth change in seasonality, because the VAR process adds its variability to it instead of reproducing the signal exactly[1]. It is also different from changing the FFT parameters of the marginal distribution, because that would be a univariate change in contrast to the method proposed, which propagates changes in the primary variable to the other variables.

Results

Eder (2013) found that the air temperature increase warmed the epilimnion faster than the hypolimnion, resulting in stronger stratification stability and because of that less mixing. Less mixing meant that the half-life of residence time of water in the deepest parts of the lake increased by more than a year. Nonetheless, even with 5°C of warming and subsequently less mixing, the simulated lake underwent full circulation every 3-4 years. Warmer water is able to hold less oxygen, which was reflected in the simulations in all depths.

An increase in variability had more subtle effects. Eder (2013) emphasizes the importance of cold episodes in the *spicy* scenarios as they tend to destabilize the stratification and thus allow mixing. A particularly cold winter with complete mixing enhances the stability of the stratification in the coming year, as temperature differences between deep and upper layers are then higher. Warm episodes are only effective at increasing the temperature in the uppermost layers, stabilize stratification and increase surface heat emissions.

Changing the seasonal cycle of air temperature revealed that the lake system is most sensitive to increases in December and January, leading to higher changes in water temperature than in the months of May – July when the water column is stratified.

[1] Taking a mean along the realization-dimension of an ensemble generated with a "climate signal" would approach said signal.

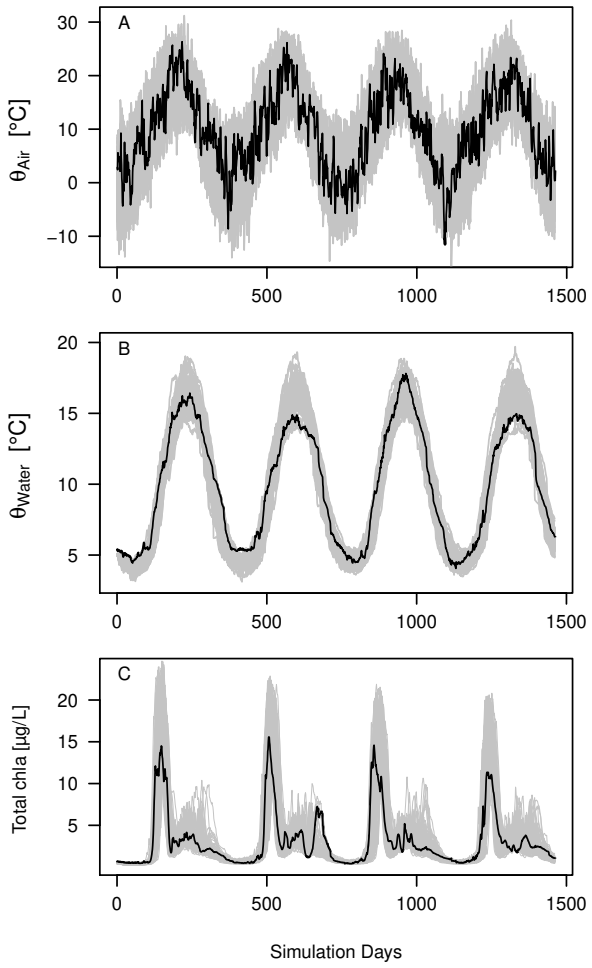


Figure 6.1: Indirect validation. Upper panel: daily mean air temperature. Middle panel: mean water temperature of the upper 20m. Lower panel: mean Chlorophyll concentration of the upper 20m. Black line represents the measured air temperature in the uppermost panel and DYRESM-CAEDYM model results based on observed meteorological input. Gray lines correspond to stale scenario realizations. As published in Schlabing et al. (2014).

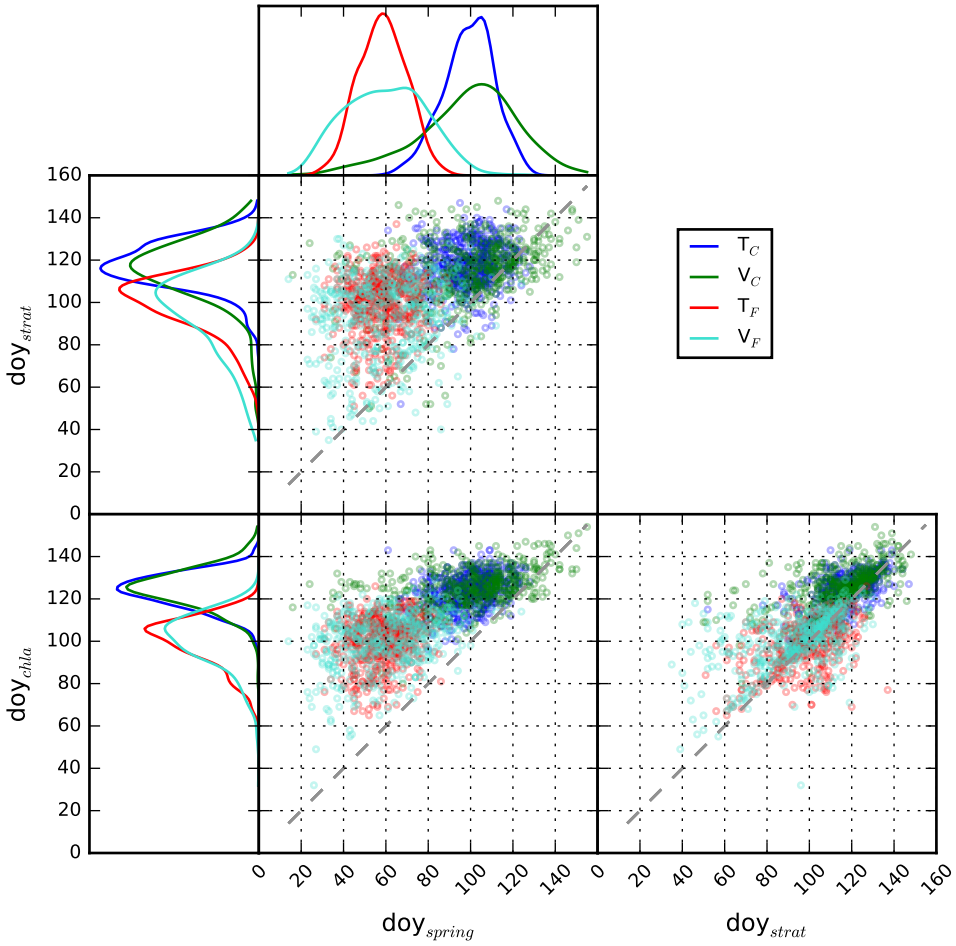


Figure 6.2: Relationship between measures of cardinal dates for chlorophyll concentrations and air and water temperature. As published in Schlabing et al. (2014).

6.3 Lake Kinneret

This section summarizes the publication Gal, Gilboa, et al. (2020). Lake Kinneret and its unique setting has been studied and modelled extensively. Jason P. Antenucci and Jörg Imberger (2003) analyzed the seasonal appearance of internal waves in reaction to wind forcing. Gal, Jorg Imberger, et al. (2003) first modelled the lake with DYRESM. Using the 3D model ELCOM, Gómez-Giraldo, Jörg Imberger, and Jason P Antenucci (2006) studied the spatial patterns of internal waves.

Located in the Jordan Rift valley in northern Israel at approximately 209m below sea level, Lake Kinneret is subject to very different meteorological conditions than Lake Constance. The lake is less deep with a maximum depth of 43m and smaller with a volume of about 4 km^3. Not only is it situated in a different climate zone than Lake Constance, its proximity to the Mediterranean Sea (≈ 45 km) with the Galilee Mountains in between, provide a distinctive summer diurnal pattern that is a driver for hydrodynamic processes and evaporation in the lake (Shilo et al., 2015). At early afternoon in the summer months, the Mediterranean Sea Breeze (MSB) reaches the lake and brings moist and slightly colder air. Wind speeds increase noticeably from approximately 3 to 8 m s^{-1}.

During very hot days, the marine inversion layer decreases under the height of the Galilee Mountains, which blocks the MSB from reaching the lake. This is not only accompanied by high local air temperatures, but also by low humidity and wind speeds. In other words, the blocking of the MSB, causes a very different diurnal pattern.

Analysis of observed temperature shows that Israel is subject to climate change regarding mean annual and summer extreme temperatures (Ziv, Saaroni, and Alpert, 2011). The summer heat waves have an impact on Lake Kinneret's ecology. An exceptional heat wave in the summer 2010 was accompanied by a clear water-phase that is unusual for that time of the year (Gal, personal communication).

6.3.1 Adaptations to VG

It was necessary to adapt VG to the meteorological conditions at Lake Kinneret. Air temperatures there exhibit a skew towards higher values, rendering the otherwise used symmetric Gaussian distribution unsuitable. Instead, I used the Gumbel distribution to better represent high values, which are directly related to the heat waves.

The heat waves caused by the blocking of the MSB by a low inversion layer represent a non-stationary behavior. I implemented these heat waves by introducing episodes of low relative humidity of 25% during the months of June to September. Per summer, 3 of those heat waves with a length of 7 to 14 days are generated by drawing the lengths from a uniform distribution. These uni-variate disturbances are propagated to the other variables in the same way as temperature changes are implemented in section 3.8 on page 47.

With air temperature and relative humidity both driving the other generated variables, an additional way of generating realizations with two "primary" variables had to be devised.

I implemented an additive approach in which the influence of each of the primary variables is calculated and added to the result. So each change can be interpreted as a distinct driver for the whole set of generated variables.

6.3.2 Model Setup

Data used for VG came from the Kinneret Limnological Lab (KLL) and was collected at an on-lake site. A period of 11 years from 1997 to 2007 was used for calibration. Simulated meteorological variables were air temperature, relative humidity, global short-wave radiation, wind speed and -direction.

In this example, multiple lake models were used in order to estimate model uncertainty. The lake models used were General Lake Model (GLM), General Ocean Turbulence Model (GOTM) and DYRESM-CAEDYM. GLM is similar in process description to DYRESM and uses Lagrangian layers that adapt to changes in vertical gradients. GOTM is an ocean model, but this application uses a version adapted to lakes.

6.3.3 Scenarios

In order to investigate the impact of summer heat waves in a warming environment in a fully crossed design, 4 scenarios were implemented:

Stale: undisturbed statistics. Serves as a reference to the other scenarios.

Gradual: includes a linear positive trend in air temperature of 0.65 °C per decade, resulting in a total increase of 2 °C.

Spicy: includes additional summer heat waves modelled after the ones described in Ziv, Saaroni, and Alpert (2011).

Spicy–Gradual: combines *Gradual* and *Spicy*.

Each scenario consists of 1000 realizations, which are 30 years long and have hourly output.

In this experiment, the same Gaussian vectors[2] and random numbers for generating the heat waves where used for the i^{th} realization of each scenario. This was implemented so that individual realizations can be consulted for discerning the effect of scenario-changes.

6.3.4 Results

The output analyzed in this study was mainly water temperature in the upper 10m and stratification lengths. The upper layer of the lake is the biologically active one and higher water temperatures would accelerate metabolic rates such as respiration and decomposition

[2]The term u_t in equation (2.2) on page 7.

Table 6.2: Climate scenarios generated by VG for Lake Kinneret study

		gradient in mean air temperature	
		– (C-scenarios)	0.65 °C / 10a (F-scenarios)
heat waves	unchanged (T-scenarios)	T_C – *Stale* "current temperatures"	T_F – *Gradual* "developing climate"
	enhanced (V-scenarios)	V_C – *Spicy* "current temperatures with enhanced heat waves"	V_F – *Spicy-Gradual* "developing climate with enhanced heat waves"

rates (J. H. Brown et al., 2004; Yvon-Durocher et al., 2012). While cardinal dates of plankton blooms could be affected by increasing air temperatures at Lake Kinneret as well, we were more interested in direct consequences of extremes.

The *Stale* scenario does not only provide a reference for the other scenarios, it also serves as an indirect validation of the model chain consisting of the weather generator and the model ensemble. Figure 6.3 reveals that the distribution of simulated summer water temperatures of individual models in the upper layers of the lake are in the range but more narrow than the distribution of historical values. Taken together, the ensemble does provide a wide enough distribution in comparison to historical values. The historical (1969-2008) length of stratification is approximately 286 days, which is comparable to the values in figure 6.4.

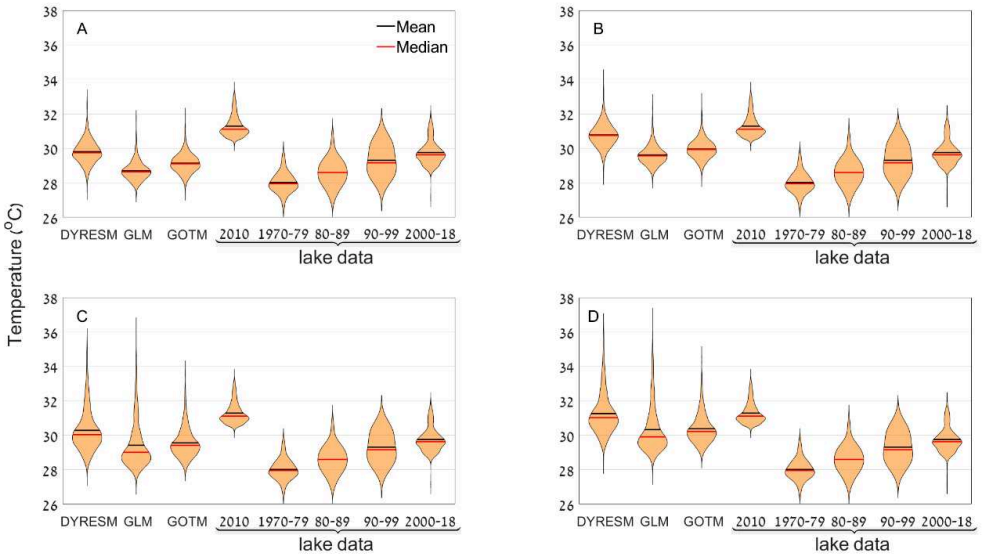


Figure 6.3: Violin plots of upper 10m water temperatures in August for the last 5 years of simulations for *Stale* (A), *Gradual* (B), *Spicy* (C) and *Spicy-Gradual* (C). As published in (Gal, Gilboa, et al., 2020).

Each dimension of the scenario table 6.2 results in distinct consequences. The temperature gradient has a long-lasting effect on the lake that can be seen in warmer upper-layer water temperatures (see panel B in figure 6.3) and longer periods of stratification (see panel B in figure 6.4). Stratification length is, however, unaffected by scenario-inputs for the GOTM model. The distributions of upper water summer temperatures are shifted upwards without changing their shape (panel B in figure 6.3).

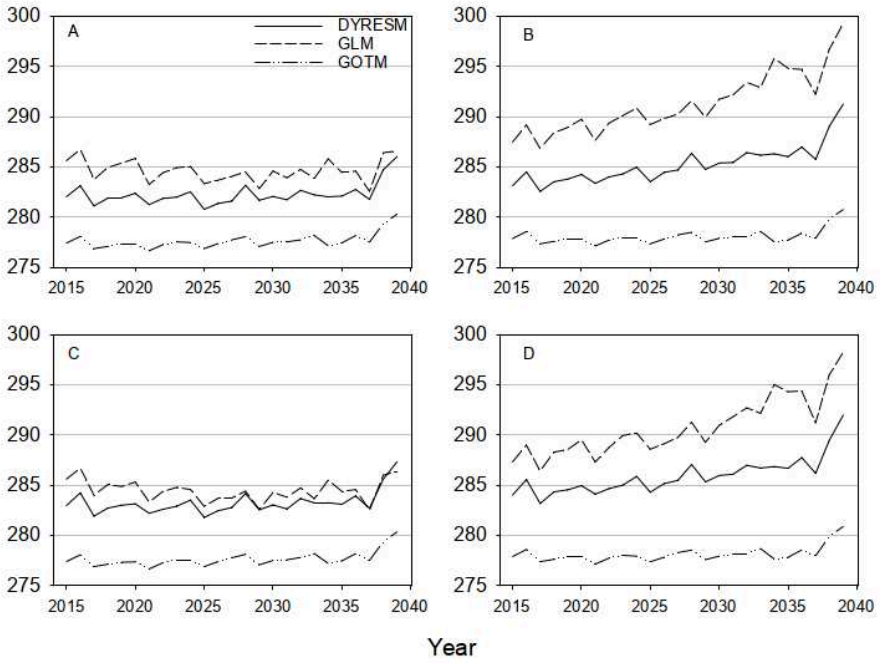


Figure 6.4: Mean stratification lengths in days per year for *Stale* (A), *Gradual* (B), *Spicy* (C) and *Spicy-Gradual* (D). Stratification in this study means that the temperature difference between the mean of the upper 10m and the mean of 30-40m depth is above 2 °C. As published in (Gal, Gilboa, et al., 2020).

Figure 6.4 also shows that heat waves have a negligible effect on stratification lengths. However when they occur, they do warm the upper layer of the lake during summer (see lower panels in figure 6.3). In relation to the *Gradual* scenario, *Spicy* results in nearly no change in mean of the upper layer summer temperatures, but it skews their distribution towards higher values. The year 2010, which featured an exceptional heat wave, is displayed separately in figure 6.3. Years with events at least as extreme as the ones in 2010 are most probable under the *Spicy* and *Spicy-Gradual* scenario and much less so for *Gradual*.

Figure 6.5 shows depth profiles of water temperatures for one representative realization. Heat waves are easily visible by increasing temperatures in the uppermost layers. With a heat wave ending, the onset of the diurnal cycle with strong wind speeds in the afternoon causes a mixing of the warm water down to the thermocline. While the epilimnion warms up towards the end of the summer under the *Gradual* scenario, such warming is only temporary in *Spicy*. However, combining *Gradual* and *Spicy*, the heat waves cause a

summer-lasting warming of the whole epilimnion much earlier than *Gradual* alone.

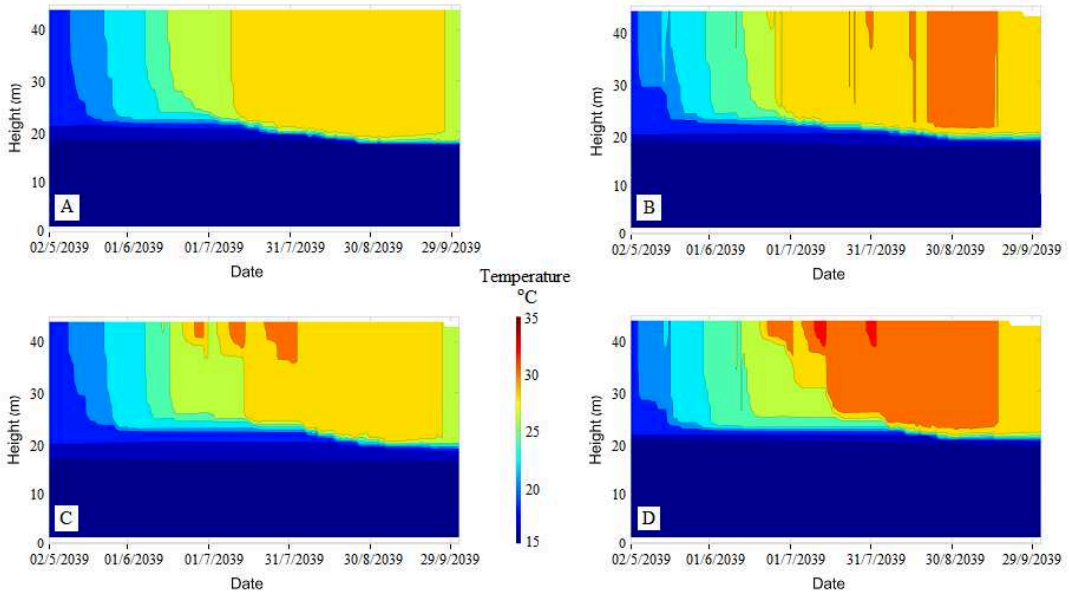


Figure 6.5: Representative water temperature profiles of GLM simulations for *Stale* (A), *Gradual* (B), *Spicy* (C) and *Spicy-Gradual* (D). As published in (Gal, Gilboa, et al., 2020).

Increased water temperatures at Lake Kinneret can cause shifts in zooplankton composition (Ninio et al., 2020), while lake warming in general is linked to proliferation of cyanobacteria and associated harmful algal blooms (Paerl and Paul, 2012).

6.4 Summary and Discussion

The lakes studied proved to be susceptible to warmer and to more variable weather. The details, however, depend on the lake and on what variability in weather entails. Warmer weather resulted most directly in increased water temperatures at the surface. How this heat spreads to deeper levels is up to the stratification, which in turn depends on the longer evolution of meteorologic conditions, and wind forcing to initiate mixing.

A combination of warming and increased variability can have more than the additive effects of their parts. At Lake Kinneret, the gradual change accompanied by heat waves triggered an earlier summer-lasting warming of the complete epilimnion.

VG is the first WG that was used in a peer-reviewed publication concerning lake modelling. It has proven its worth with regard to estimating possible climate impacts on such eco-hydraulical systems as a tool to produce what-if scenarios defined by the lake-modeller.

The results presented in this chapter have to be interpreted as projections based on past linear correlation structures. As shown in the extrapolation test in section 3.9, changes in

air temperature might not be accompanied by the changes in other variables as expected by their statistical relation. Still, the method of balanced scenarios is arguably more statistically and meteorologically plausible than scenarios with isolated, univariate changes.

Lake Constance and Kinneret are very different in size and are located in different climate zones with distinct meteorological conditions. VG can be configured to account for different climates, lakes and lake models. Having designed VG myself made it possible to adapt this tool in a targeted way. The close collaboration with lake modelers made it possible to identify such needs.

7 Outlook

Apart from addressing the weak points of the proposed WGs (overestimation of variability, use of simplified vines, daily generation), a few other, tangential research trajectories are sketched in the following.

7.1 Non-Gaussian Temporal Dependence

In the aim of having few free parameters, WeatherCop models temporal dependence by phase randomization, which preserves linear dependence. This could be overcome by employing phase annealing, which optimizes phase values by simulated annealing in order to minimize an objective function that characterizes deviations from possibly non-gaussian properties (Hörning, 2016). It is flexible as the objective function can be formulated freely, but the additional optimization step, that becomes part of the simulation process, adds computational cost.

The phase randomization technique relies on the fact that the autocorrelation function is only related to the power spectrum and not the phases. Hence, "surrogate" time series with the same autocorrelation as a source time series can be generated by randomizing the phases. Other, higher-order statistics do depend on the phases such as the so-called bispectrum (Collis, White, and Hammond, 1998). The bispectrum arises when conducting spectral analysis of cubed signals and, as the power spectrum is a decomposition of the power of a signal, the bispectrum decomposes the skewness. A possible next step would be to find ways to randomize phases without changing the bispectrum and other higher-order spectra too much, thus maintaining more statistics during phase randomization. If this can be done without an on-line optimization step, it would be faster than phase annealing, however lacking its flexibility. If the mathematical tools could be found, this kind of exploration could extend to "auto-bispectra", which has the potential to keep temporal asymmetry (the skewness of the difference of a signal with itself lagged) that is usually present in discharge time series.

7.2 Coupling Multi-Site and Wind Field Generation for 3D Lake Modelling

The horizontal wind speed distribution has an influence on flow patterns and mixing in lakes. It is a feature requested repeatedly by limnologists who are employing 3D lake models. A path towards wind fields may be a gridded version of WeatherCop that interpolates

amplitude spectra of the decorrelated variables spatially. Physical conservation laws that govern flow and validation using output of regional atmospheric models pose challenges, though. Simplifications can possibly be found by determining necessary properties of the wind fields for 3D lake modelling.

7.3 Variability Changes Through Amplitude Spectrum Adjustment

An aim in designing the enhanced variability scenario in section 3.8.2 was to obtain variability changes not just in the marginal distributions, but on a time scale of about a week. Increasing the amplitude of a range of frequencies close to the weekly might achieve the same thing and would be achievable in a precise manner with the help of the Fourier Transform.

7.4 Extending the Model-Chain to Rainfall-Runoff-Models

Rivers also play an important role in lake ecology. The model-chain of WG and hydrodynamic model could be extended by a rainfall-runoff model. Modelling the watershed in together with lakes is nothing new (Cremona et al., 2017), but a multi-site WG feeding both the river- and the lake model would be. Such a WG should be able to be consistent across longer temporal and spatial scales.

Bibliography

Aas, Kjersti, Claudia Czado, Arnoldo Frigessi, and Henrik Bakken (2009). "Pair-copula constructions of multiple dependence." In: *Insurance: Mathematics and Economics* 44.2, pp. 182–198. ISSN: 01676687. DOI: 10.1016/j.insmatheco.2007.02.001. URL: http://linkinghub.elsevier.com/retrieve/pii/S0167668707000194.

Agilan, V. and N. V. Umamahesh (Sept. 2019). "Rainfall Generator for Nonstationary Extreme Rainfall Condition." In: *Journal of Hydrologic Engineering* 24.9, p. 04019027. ISSN: 1084-0699. DOI: 10.1061/(ASCE)HE.1943-5584.0001821. URL: http://ascelibrary.org/doi/10.1061/%7B%5C%%7D28ASCE%7B%5C%%7D29HE.1943-5584.0001821.

Antenucci, Jason P. and Jörg Imberger (2003). "The seasonal evolution of wind/internal wave resonance in Lake Kinneret." In: *Limnology and Oceanography* 48.5, pp. 2055–2061. ISSN: 00243590. DOI: 10.4319/lo.2003.48.5.2055. URL: http://www.aslo.org/lo/toc/vol%7B%5C_%7D48/issue%7B%5C_%7D5/2055.html.

Bárdossy, András and Geoffrey G. S. Pegram (Feb. 2016). "Space-time conditional disaggregation of precipitation at high resolution via simulation." In: *Water Resources Research* 52.2, pp. 920–937. ISSN: 00431397. DOI: 10.1002/2015WR018037. URL: http://doi.wiley.com/10.1002/2015WR018037.

Baxevani, Anastassia and Jan Lennartsson (June 2015). "A spatiotemporal precipitation generator based on a censored latent Gaussian field." In: *Water Resources Research* 51.6, pp. 4338–4358. ISSN: 00431397. DOI: 10.1002/2014WR016455. URL: http://doi.wiley.com/10.1002/2014WR016455.

Behnel, Stefan, Robert Bradshaw, Craig Citro, Lisandro Dalcin, Dag Sverre Seljebotn, and Kurt Smith (Mar. 2011). "Cython: The Best of Both Worlds." In: *Computing in Science & Engineering* 13.2, pp. 31–39. ISSN: 1521-9615. DOI: 10.1109/MCSE.2010.118. URL: http://ieeexplore.ieee.org/document/5582062/.

Brown, James H., James F. Gillooly, Andrew P. Allen, Van M. Savage, and Geoffrey B. West (July 2004). "TOWARD A METABOLIC THEORY OF ECOLOGY." In: *Ecology* 85.7, pp. 1771–1789. ISSN: 0012-9658. DOI: 10.1890/03-9000. URL: http://doi.wiley.com/10.1890/03-9000.

Buishand, T a and R Wójcik (2003). "Simulation of 6-hourly rainfall and temperature by two resampling schemes." In: *Journal of Hydrology* 273.1-4, pp. 69–80.

Buishand, T. A. and T. Brandsma (2001). "Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbor resampling." In: *Water Resources Research* 37.11, pp. 2761–2776. ISSN: 00431397. DOI: 10.1029/2001WR000291.

Caskey, James E. (Jan. 1962). "A Markov chain model for daily rainfall occurrence at Tel Aviv." In: *Quarterly Journal of the Royal Meteorological Society* 88.375, pp. 90–95. ISSN:

00359009. DOI: 10.1002/qj.49708837511. URL: http://doi.wiley.com/10.1002/qj.49708837511.

Chen, Jie, François P. Brissette, and Robert Leconte (July 2010). "A daily stochastic weather generator for preserving low-frequency of climate variability." In: *Journal of Hydrology* 388.3-4, pp. 480–490. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2010.05.032. URL: http://linkinghub.elsevier.com/retrieve/pii/S0022169410003082.

Clark, Martyn, Subhrendu Gangopadhyay, Lauren Hay, Balaji Rajagopalan, and Robert Wilby (Feb. 2004). "The Schaake Shuffle: A Method for Reconstructing Space–Time Variability in Forecasted Precipitation and Temperature Fields." In: *Journal of Hydrometeorology* 5.1, pp. 243–262. ISSN: 1525-755X. DOI: 10.1175/1525-7541(2004)005<0243:TSSAMF>2.0.CO;2. URL: http://journals.ametsoc.org/doi/abs/10.1175/1525-7541%7B%5C%%7D282004%7B%5C%%7D29005%7B%5C%%7D3C0243%7B%5C%%7D3ATSSAMF%7B%5C%%7D3E2.0.CO%7B%5C%%7D3B2.

Collis, W. B., P. R. White, and J. K. Hammond (1998). "Higher-order spectra: The bispectrum and trispectrum." In: *Mechanical Systems and Signal Processing* 12.3, pp. 375–394. ISSN: 08883270. DOI: 10.1006/mssp.1997.0145.

Cooke, Roger M. (1997). "Markov and Entropy Properties of Tree and Vine-Dependent Variables." In: *American Statistical Association - Proceedings of the Section on Bayesian Statistical Science*.

Cremona, Fabien, Sirje Vilbaste, Raoul-Marie Couture, Peeter Nõges, and Tiina Nõges (Mar. 2017). "Is the future of large shallow lakes blue-green? Comparing the response of a catchment-lake model chain to climate predictions." In: *Climatic Change* 141.2, pp. 347–361. ISSN: 0165-0009. DOI: 10.1007/s10584-016-1894-8. URL: http://link.springer.com/10.1007/s10584-016-1894-8.

Czado, Claudia, Ulf Schepsmeier, and Aleksey Min (June 2012). "Maximum likelihood estimation of mixed C-vines with application to exchange rates." In: *Statistical Modelling: An International Journal* 12.3, pp. 229–255. ISSN: 1471-082X. DOI: 10.1177/1471082X1101200302. URL: http://journals.sagepub.com/doi/10.1177/1471082X1101200302.

Dißmann, J., E. C. Brechmann, C. Czado, and D. Kurowicka (2013). "Selecting and estimating regular vine copulae and application to financial returns." In: *Computational Statistics and Data Analysis* 59.1, pp. 52–69. ISSN: 01679473. DOI: 10.1016/j.csda.2012.08.010. arXiv: 1202.2002.

Dubrovský, Martin, Josef Buchtele, and Zdeněk Žalud (Mar. 2004). "High-Frequency and Low-Frequency Variability in Stochastic Daily Weather Generator and Its Effect on Agricultural and Hydrologic Modelling." In: *Climatic Change* 63.1/2, pp. 145–179. ISSN: 0165-0009. DOI: 10.1023/B:CLIM.0000018504.99914.60. URL: http://www.springerlink.com/openurl.asp?id=doi:10.1023/B:CLIM.0000018504.99914.60.

Duin (Nov. 1976). "On the Choice of Smoothing Parameters for Parzen Estimators of Probability Density Functions." In: *IEEE Transactions on Computers* C-25.11, pp. 1175–1179. ISSN: 0018-9340. DOI: 10.1109/TC.1976.1674577. URL: http://ieeexplore.ieee.org/document/1674577/.

Eddy, W. F. (1982). "Convex Hull Peeling." In: *COMPSTAT 1982 5th Symposium held at Toulouse 1982*. Heidelberg: Physica-Verlag HD, pp. 42–47. ISBN: 3705100025. DOI: 10.

1007/978-3-642-51461-6_4. URL: http://link.springer.com/10.1007/978-3-642-51461-6%7B%5C_%7D4.

Eder, MM Maria Magdalena (2013). "Climate sensitivity of a large lake." PhD thesis. Holzgartenstr. 16, 70174 Stuttgart: Universität Stuttgart. ISBN: 978-3-942036-30-6. URL: http://elib.uni-stuttgart.de/opus/volltexte/2014/8885.

Fenocchi, Andrea, Michela Rogora, Stefano Sibilla, Marzia Ciampittiello, and Claudia Dresti (2018). "Forecasting the evolution in the mixing regime of a deep subalpine lake under climate change scenarios through numerical modelling (Lake Maggiore, Northern Italy/Southern Switzerland)." In: *Climate Dynamics* 51.9-10, pp. 3521–3536. ISSN: 14320894. DOI: 10.1007/s00382-018-4094-6. URL: http://dx.doi.org/10.1007/s00382-018-4094-6.

Fowler, H. J., C. G. Kilsby, and P. E. O'Connell (2000). "A stochastic rainfall model for the assessment of regional water resource systems under changed climatic condition." In: *Hydrology and Earth System Sciences* 4.2, pp. 263–281. ISSN: 1607-7938. DOI: 10.5194/hess-4-263-2000. URL: http://www.hydrol-earth-syst-sci.net/4/263/2000/.

Gal, Gideon, Yael Gilboa, Noam Schachar, Moshe Estroti, and Dirk Schlabing (2020). "Ensemble Modeling of the Impact of Climate Warming and Increased Frequency of Extreme Climatic Events on the Thermal Characteristics of a Sub-Tropical Lake." In: *Water* 12.7, p. 1982. DOI: 10.3390/w12071982.

Gal, Gideon, Jorg Imberger, Tamar Zohary, Jason Antenucci, Ayal Anis, and Tzahi Rosenberg (Apr. 2003). "Simulating the thermal dynamics of Lake Kinneret." In: *Ecological Modelling* 162.1-2, pp. 69–86. ISSN: 03043800. DOI: 10.1016/S0304-3800(02)00380-0. URL: http://linkinghub.elsevier.com/retrieve/pii/S0304380002003800.

Gómez-Giraldo, Andrés, Jörg Imberger, and Jason P Antenucci (2006). "Spatial structure of the dominant basin-scale internal waves in Lake Kinneret." In: *Limnology and Oceanography* 51.1, pp. 229–246. URL: http://www.jstor.org/stable/4499577.

Hao, Zengchao and Vijay P. Singh (2016). "Review of dependence modeling in hydrology and water resources." In: *Progress in Physical Geography* 40.4, pp. 549–578. ISSN: 0309-1333. DOI: 10.1177/0309133316632460. URL: http://ppg.sagepub.com/cgi/doi/10.1177/0309133316632460.

Harrold, Timothy I. (2003). "A nonparametric model for stochastic generation of daily rainfall occurrence." In: *Water Resources Research* 39.10. ISSN: 0043-1397. DOI: 10.1029/2003WR002182. URL: http://www.agu.org/pubs/crossref/2003/2003WR002182.shtml.

Hörning, Sebastian (2016). "Process-oriented modeling of spatial random fields using copulas." PhD thesis. ISBN: 9783942036504.

Hughes, J. P., P Guttorp, and S. P. Charles (Feb. 1999). "A non-homogeneous hidden Markov model for precipitation occurrence." In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 48.1, pp. 15–30. ISSN: 0035-9254. DOI: 10.1111/1467-9876.00136. URL: http://doi.wiley.com/10.1111/1467-9876.00136.

IPCC (2007). "Climate change 2007: synthesis report." In: *Change* November, pp. 12–17. URL: http://scholar.google.com/scholar?hl=en%7B%5C&%7DbtnG=Search%

7B%5C&%7Dq=intitle:Climate+change+2007:+synthesis+report%7B%5C#%7D2.

Joe, Harry (Mar. 1994). "Multivariate extreme-value distributions with applications to environmental data." In: *Canadian Journal of Statistics* 22.1, pp. 47–64. ISSN: 03195724. DOI: 10.2307/3315822. URL: http://doi.wiley.com/10.2307/3315822.

— (2015). *Dependence Modeling with Copulas*. CRC press. ISBN: 9781466583221.

Khazaei, Mohammad Reza, Shahin Ahmadi, Bahram Saghafian, and Bagher Zahabiyoun (2013). "A new daily weather generator to preserve extremes and low-frequency variability." In: *Climatic Change* 119.3-4, pp. 631–645. ISSN: 01650009. DOI: 10.1007/s10584-013-0740-5.

Kim, Y, Rw Katz, B Rajagopalan, Gp Podestá, and Em Furrer (May 2012). "Reducing overdispersion in stochastic weather generators using a generalized linear modeling approach." In: *Climate Research* 53.1, pp. 13–24. ISSN: 0936-577X. DOI: 10.3354/cr01071. URL: http://www.int-res.com/abstracts/cr/v53/n1/p13-24/.

King, Leanna M., a. Ian McLeod, and Slobodan P. Simonovic (Oct. 2015). "Improved Weather Generator Algorithm for Multisite Simulation of Precipitation and Temperature." In: *JAWRA Journal of the American Water Resources Association* 51.5, pp. 1305–1320. ISSN: 1093474X. DOI: 10.1111/1752-1688.12307. URL: http://doi.wiley.com/10.1111/1752-1688.12307.

Kleiber, William, Richard W. Katz, and Balaji Rajagopalan (Jan. 2012). "Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes." In: *Water Resources Research* 48.1, W01523. ISSN: 0043-1397. DOI: 10.1029/2011WR011105. URL: http://www.agu.org/pubs/crossref/2012/2011WR011105.shtml.

Kobler, Ulrike Gabriele, Alfred Wüest, and Martin Schmid (Dec. 2018). "Combined effects of pumped-storage operation and climate change on thermal structure and water quality." In: *Climatic Change*. ISSN: 0165-0009. DOI: 10.1007/s10584-018-2340-x. URL: http://link.springer.com/10.1007/s10584-018-2340-x.

Kucharik, Christopher J. (2003). "Evaluation of a Process-Based Agro-Ecosystem Model (Agro-IBIS) across the U.S. Corn Belt: Simulations of the Interannual Variability in Maize Yield." In: *Earth Interactions* 7.14, pp. 1–33. ISSN: 1087-3562. DOI: 10.1175/1087-3562(2003)007<0001:eoapam>2.0.co;2.

Kumaraswamy, P. (1980). "A generalized probability density function for double- bounded random processes." In: *Journal of Hydrology* 46, pp. 79–88.

Kupisch, Moritz, Sylvia Moenickes, Jeanette Schlief, Marieke Frassl, and Otto Richter (2012). "Temperature-dependent consumer-resource dynamics: A coupled structured model for Gammarus pulex (L.) and leaf litter." In: *Ecological Modelling* 247, pp. 157–167. ISSN: 03043800. DOI: 10.1016/j.ecolmodel.2012.07.037. URL: http://dx.doi.org/10.1016/j.ecolmodel.2012.07.037.

Lall, Upmanu and Ashish Sharma (Mar. 1996). "A Nearest Neighbor Bootstrap For Resampling Hydrologic Time Series." In: *Water Resources Research* 32.3, pp. 679–693. ISSN: 00431397. DOI: 10.1029/95WR02966. URL: http://doi.wiley.com/10.1029/95WR02966.

Lee, Taesam (2017). "Climate change inspector with intentionally biased bootstrapping (CCIIBB ver. 1.0) – methodology development." In: *Geoscientific Model De-*

velopment 10.2, pp. 525–536. ISSN: 1991-9603. DOI: 10.5194/gmd-10-525-2017. URL: http://www.geosci-model-dev.net/10/525/2017/.

Lee, Taesam and Vijay P. Singh (2019). "Discrete k-nearest neighbor resampling for simulating multisite precipitation occurrence and model adaption to climate change." In: *Geoscientific Model Development* 12.3, pp. 1189–1207. ISSN: 19919603. DOI: 10.5194/gmd-12-1189-2019.

Lütkepohl, Helmut (Oct. 2006). *NEW INTRODUCTION TO MULTIPLE TIME SERIES ANALYSIS*. 1st ed. Vol. 22. 05. Berlin: Springer-Verlag, p. 764. ISBN: 978-3540262398. URL: http://www.journals.cambridge.org/abstract%7B%5C_%7DS0266466606000442.

Mavromatis, Theodoruos and James W Hansen (Sept. 2001). "Interannual variability characteristics and simulated crop response of four stochastic weather generators." In: *Agricultural and Forest Meteorology* 109.4, pp. 283–296. ISSN: 01681923. DOI: 10.1016/S0168-1923(01)00272-6. URL: http://linkinghub.elsevier.com/retrieve/pii/S0168192301002726.

Mehrotra, R. and Ashish Sharma (Oct. 2007). "Preserving low-frequency variability in generated daily rainfall sequences." In: *Journal of Hydrology* 345.1-2, pp. 102–120. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2007.08.003. URL: http://linkinghub.elsevier.com/retrieve/pii/S0022169407004416.

Mehrotra, R., R. Srikanthan, and Ashish Sharma (Nov. 2006). "A comparison of three stochastic multi-site precipitation occurrence generators." In: *Journal of Hydrology* 331.1-2, pp. 280–292. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2006.05.016. URL: http://linkinghub.elsevier.com/retrieve/pii/S0022169406002861.

Meurer, Aaron, Christopher P Smith, Mateusz Paprocki, Ondřej Čertík, Sergey B Kirpichev, Matthew Rocklin, Amit Kumar, Sergiu Ivanov, Jason K Moore, Sartaj Singh, Thilina Rathnayake, Sean Vig, Brian E Granger, Richard P Muller, Francesco Bonazzi, and Harsh Gupta (2017). "SymPy : symbolic computing in Python." In: pp. 1–27. DOI: 10.7717/peerj-cs.103.

Morales-Nápoles, O (Dec. 2010). "Counting Vines." In: *Dependence Modeling*. WORLD SCIENTIFIC. Chap. 9, pp. 189–218. DOI: 10.1142/9789814299886_0009. URL: https://www.worldscientific.com/doi/abs/10.1142/9789814299886%7B%5C_%7D0009%20http://www.worldscientific.com/doi/abs/10.1142/9789814299886%7B%5C_%7D0009.

Ninio, Shira, Achsa Lupu, Yehudit Viner-Mozzini, Tamar Zohary, and Assaf Sukenik (2020). "Multiannual variations in Microcystis bloom episodes – Temperature drives shift in species composition." In: *Harmful Algae* 92.August, p. 101710. ISSN: 18781470. DOI: 10.1016/j.hal.2019.101710. URL: https://doi.org/10.1016/j.hal.2019.101710.

O'Reilly, Catherine M., Sapna Sharma, Derek K. Gray, Stephanie E. Hampton, Jordan S. Read, Rex J. Rowley, Philipp Schneider, John D. Lenters, Peter B. McIntyre, Benjamin M. Kraemer, Gesa A. Weyhenmeyer, Dietmar Straile, Bo Dong, Rita Adrian, Mathew G. Allan, Orlane Anneville, Lauri Arvola, Jay Austin, John L. Bailey, Jill S. Baron, Justin D. Brookes, Elvira Eyto, Martin T. Dokulil, David P. Hamilton, Karl Havens, Amy L. Hetherington, Scott N. Higgins, Simon Hook, Lyubov R. Izmest'eva, Klaus D. Joehnk, Kulli Kangur, Peter Kasprzak, Michio Kumagai, Esko Kuusisto, George Leshkevich, David M.

Livingstone, Sally MacIntyre, Linda May, John M. Melack, Doerthe C. Mueller-Navarra, Mikhail Naumenko, Peeter Noges, Tiina Noges, Ryan P. North, Pierre-Denis Plisnier, Anna Rigosi, Alon Rimmer, Michela Rogora, Lars G. Rudstam, James A. Rusak, Nico Salmaso, Nihar R. Samal, Daniel E. Schindler, S. Geoffrey Schladow, Martin Schmid, Silke R. Schmidt, Eugene Silow, M. Evren Soylu, Katrin Teubner, Piet Verburg, Ari Voutilainen, Andrew Watkinson, Craig E. Williamson, and Guoqing Zhang (Dec. 2015). "Rapid and highly variable warming of lake surface waters around the globe." In: *Geophysical Research Letters* 42.24, pp. 1–9. ISSN: 0094-8276. DOI: 10.1002/2015GL066235. URL: https://onlinelibrary.wiley.com/doi/10.1002/2015GL066235.

Olson, Tim (2017). *Applied Fourier Analysis*. New York, NY: Springer New York. ISBN: 978-1-4939-7391-0. DOI: 10.1007/978-1-4939-7393-4. URL: http://link.springer.com/10.1007/978-1-4939-7393-4.

Onof, C., R. E. Chandler, a. Kakou, P. Northrop, H. S. Wheater, and V. Isham (Nov. 2000). "Rainfall modelling using Poisson-cluster processes: a review of developments." In: *Stochastic Environmental Research and Risk Assessment* 14.6, pp. 0384–0411. ISSN: 14363240. DOI: 10.1007/s004770000043. URL: http://link.springer.com/10.1007/s004770000043.

Pachauri, Rajendra K, Myles R Allen, Vicente R Barros, John Broome, Wolfgang Cramer, Renate Christ, John A Church, Leon Clarke, Qin Dahe, Purnamita Dasgupta, et al. (2014). *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Ipcc.

Paerl, Hans W. and Valerie J. Paul (2012). "Climate change: Links to global expansion of harmful cyanobacteria." In: *Water Research* 46.5, pp. 1349–1363. ISSN: 18792448. DOI: 10.1016/j.watres.2011.08.002. URL: http://dx.doi.org/10.1016/j.watres1.2011.08.002.

Peleg, Nadav, Simone Fatichi, Athanasios Paschalis, Peter Molnar, and Paolo Burlando (2017). "An advanced stochastic weather generator for simulating 2-D high-resolution climate variables." In: *Journal of Advances in Modeling Earth Systems* 9.3, pp. 1595–1627. ISSN: 19422466. DOI: 10.1002/2016MS000854. URL: http://doi.wiley.com/10.1002/2016MS000854.

Prichard and Theiler (1994). "Generating Surrogate Data for Time Series with Several Simultaneously Measured Variables." In: *Physical Review Letters* 73.7, pp. 951–954.

Racsko, P, L Szeidl, and M. Semenov (Oct. 1991). "A serial approach to local stochastic weather models." In: *Ecological Modelling* 57.1-2, pp. 27–41. ISSN: 03043800. DOI: 10.1016/0304-3800(91)90053-4. URL: http://linkinghub.elsevier.com/retrieve/pii/0304380091900534.

Rajagopalan, Balaji and Upmanu Lall (1999). "A k -nearest-neighbor simulator for daily precipitation and other weather variables." In: *Water Resources Research* 35.10, p. 3089. ISSN: 0043-1397. DOI: 10.1029/1999WR900028. URL: http://www.agu.org/pubs/crossref/1999/1999WR900028.shtml.

Richardson, C. W. and D. A. Wright (1984). "WGEN : A Model for Generating Daily Weather Variables." In: *U. S. Department of Agriculture, Agricultural Research Service* ARS-8, p. 83. ISSN: 01697722. URL: http://soilphysics.okstate.edu/software/cmls/WGEN.pdf.

Richardson, C. W. CW (1981). "Stochastic simulation of daily precipitation, temperature, and solar radiation." In: *Water Resources Research* 17.1, p. 182. ISSN: 0043-1397. DOI: 10.1029/WR017i001p00182. URL: http://www.agu.org/pubs/crossref/1981/WR017i001p00182.shtml%20http://onlinelibrary.wiley.com/doi/10.1029/WR017i001p00182/full.

Rodriguez-Iturbe, I (1987). "Some models for rainfall based on stochastic point processes." In: *Proceedings Of The Royal Society Of London Series A, Mathematical and Physical Sciences* 410.1839, pp. 269–288. URL: http://rspa.royalsocietypublishing.org/content/410/1839/269.short.

Rosenblatt, Murray (1952). "REMARKS ON A MULTIVARIATE TRANSFORMATION." In: *The Annals of Mathematical Statistics* 23.3, pp. 470–472.

Schirmacher, Doris and Ernesto Schirmacher (2008). "Multivariate Dependence Modeling Using Pair-Copulas." In: *2008 ERM Symposium*, pp. 1–52. URL: http://www.ermsymposium.org/2008/index.php.

Schlabing, D., M.a. Frassl, M.M. Eder, K. Rinke, and A. Bárdossy (July 2014). "Use of a weather generator for simulating climate change effects on ecosystems: A case study on Lake Constance." In: *Environmental Modelling & Software*, pp. 1–13. ISSN: 13648152. DOI: 10.1016/j.envsoft.2014.06.028. URL: http://linkinghub.elsevier.com/retrieve/pii/S1364815214002059.

Schwefel, Robert, Adrien Gaudard, Alfred Wüest, and Damien Bouffard (Nov. 2016). "Effects of climate change on deepwater oxygen and winter mixing in a deep lake (Lake Geneva): Comparing observational findings and modeling." In: *Water Resources Research* 52.11, pp. 8811–8826. ISSN: 00431397. DOI: 10.1002/2016WR019194. URL: http://doi.wiley.com/10.1002/2016WR019194.

Semenov, M A (2002). "LARS-WG: a stochastic weather generator for use in climate impact studies." In: *User Manual Version*. URL: http://scholar.google.com.sg/scholar?hl=en%7B%5C&%7Dq=LARS-wg%7B%5C&%7DbtnG=Search%7B%5C&%7Das%7B%5C_%7Dsdt=0,5%7B%5C&%7Das%7B%5C_%7Dylo=%7B%5C&%7Das%7B%5C_%7Dvis=0%7B%5C#%7D4.

Shannon, C. E. (July 1948). "A Mathematical Theory of Communication." In: *Bell System Technical Journal* 27.3, pp. 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x. URL: https://academic.oup.com/jas/article/36/5/986-993/4667403%20http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6773024.

Sharif, Mohammed and Donald H Burn (June 2006). "Simulating climate change scenarios using an improved K-nearest neighbor model." In: *Journal of Hydrology* 325.1-4, pp. 179–196. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2005.10.015. URL: http://linkinghub.elsevier.com/retrieve/pii/S0022169405005421.

Shilo, Elad, Baruch Ziv, Eylon Shamir, and Alon Rimmer (2015). "Evaporation from Lake Kinneret, Israel, during hot summer days." In: *Journal of Hydrology* 528, pp. 264–275. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2015.06.042. URL: http://dx.doi.org/10.1016/j.jhydrol.2015.06.042.

Singh, Vijay P. (1997). "The use of entropy in hydrology and water resources." In: *Hydrological Processes* 11.1996, pp. 587–626. ISSN: 08856087. DOI: 10.1002/(SICI)1099-1085(199705)11:6<587::AID-HYP479>3.0.CO;2-P.

Sklar, A (1959). "Fonctions de répartition à n dimensions et leurs marges." In: *Publ. Inst. Statist. Univ. Paris* 8, pp. 229–231. DOI: 10.1007/978-3-642-33590-7.

Steinschneider, Scott and Casey Brown (2013). "A semiparametric multivariate, multisite weather generator with low-frequency variability for use in climate risk assessments." In: *Water Resources Research* 49.11, pp. 7205–7220. ISSN: 00431397. DOI: 10.1002/wrcr.20528.

Te Linde, A. H., J. C.J.H. Aerts, A. M.R. Bakker, and J. C.J. Kwadijk (2010). "Simulating low-probability peak discharges for the Rhine basin using resampled climate modeling data." In: *Water Resources Research* 46.3, pp. 1–19. ISSN: 00431397. DOI: 10.1029/2009WR007707.

Theiler, J, S Eubank, A Longtin, B Galdrikian, and J D Farmer (1992). "Testing for nonlinearity in time series: the method of surrogate data." In: *Physica D* 58.1-4, pp. 77–94. ISSN: 01672789. DOI: 10.1016/0167-2789(92)90102-S. arXiv: 9909037 [chao-dyn].

Trolle, Dennis, David P. Hamilton, Conrad a. Pilditch, Ian C. Duggan, and Erik Jeppesen (Apr. 2011). "Predicting the effects of climate change on trophic status of three morphologically varying lakes: Implications for lake restoration and management." In: *Environmental Modelling & Software* 26.4, pp. 354–370. ISSN: 13648152. DOI: 10.1016/j.envsoft.2010.08.009. URL: http://linkinghub.elsevier.com/retrieve/pii/S1364815210002562.

Tukey, JW (1975). "Mathematics and the picturing of data." In: *Proceedings of the international congress of mathematicians*, pp. 523–532. URL: http://www.mathunion.org/ICM/ICM1974.2/Main/icm1974.2.0523.0532.ocr.pdf.

Wiener, Norbert (1930). "Generalized harmonic analysis." In: *Acta Mathematica* 55, pp. 117–258. ISSN: 0001-5962. DOI: 10.1007/BF02546511. URL: http://projecteuclid.org/euclid.acta/1485887877.

Wilby, R.L. and T.M.L. Wigley (Dec. 1997). "Downscaling general circulation model output: a review of methods and limitations." In: *Progress in Physical Geography: Earth and Environment* 21.4, pp. 530–548. ISSN: 0309-1333. DOI: 10.1177/030913339702100403. URL: http://journals.sagepub.com/doi/10.1177/030913339702100403.

Wilks, D.S. (Sept. 1998). "Multisite generalization of a daily stochastic precipitation generation model." In: *Journal of Hydrology* 210.1-4, pp. 178–191. ISSN: 00221694. DOI: 10.1016/S0022-1694(98)00186-3. URL: http://linkinghub.elsevier.com/retrieve/pii/S0022169498001863.

Wilks, Daniel S (2015). "Multivariate ensemble Model Output Statistics using empirical copulas." In: April, pp. 945–952. DOI: 10.1002/qj.2414.

— (Sept. 1992). "Adapting stochastic weather generation algorithms for climate change studies." In: *Climatic Change* 22.1, pp. 67–84. ISSN: 0165-0009. DOI: 10.1007/BF00143344. URL: http://www.springerlink.com/index/10.1007/BF00143344.

— (Aug. 1999). "Simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at multiple sites in complex terrain." In: *Agricultural and Forest Meteorology* 96.1-3, pp. 85–101. ISSN: 01681923. DOI: 10.1016/S0168-1923(99)00037-4. URL: http://linkinghub.elsevier.com/retrieve/pii/S0168192399000374.

Wilks, Daniel S. and R. L. Wilby (July 1999). "The weather generation game: a review of stochastic weather models." In: *Progress in Physical Geography* 23.3, pp. 329–357. ISSN: 0309-1333. DOI: 10.1177/030913339902300302. URL: http://ppg.sagepub.com/cgi/doi/10.1177/030913339902300302.

Woolway, R. Iestyn and Christopher J. Merchant (2019). "Worldwide alteration of lake mixing regimes in response to climate change." In: *Nature Geoscience* 12.4, pp. 271–276. ISSN: 1752-0894. DOI: 10.1038/s41561-019-0322-x. URL: http://dx.doi.org/10.1038/s41561-019-0322-x.

Woolway, R. Iestyn, Gesa A. Weyhenmeyer, Martin Schmid, Martin T. Dokulil, Elvira de Eyto, Stephen C. Maberly, Linda May, and Christopher J. Merchant (2019). "Substantial increase in minimum lake surface temperatures under climate change." In: *Climatic Change*, pp. 81–94. ISSN: 0165-0009. DOI: 10.1007/s10584-019-02465-y.

Yates, David (2003). "A technique for generating regional climate scenarios using a nearest-neighbor algorithm." In: *Water Resources Research* 39.7, p. 1199. ISSN: 0043-1397. DOI: 10.1029/2002WR001769. URL: http://www.agu.org/pubs/crossref/2003/2002WR001769.shtml.

Yvon-Durocher, Gabriel, Jane M. Caffrey, Alessandro Cescatti, Matteo Dossena, Paul Del Giorgio, Josep M. Gasol, José M. Montoya, Jukka Pumpanen, Peter A. Staehr, Mark Trimmer, Guy Woodward, and Andrew P. Allen (2012). "Reconciling the temperature dependence of respiration across timescales and ecosystem types." In: *Nature* 487.7408, pp. 472–476. ISSN: 00280836. DOI: 10.1038/nature11205.

Ziv, B, H Saaroni, and P Alpert (2011). "Temperature, heat-stress and rainfall fluctuations and trends in Israel during the last 30 years: is it an evidence for climatic change." In: *Climatological–synoptic study, report*, pp. 8–810.

A Tables for VG Extrapolation Test

Table A.2 shows the changed means between calibration and validation period for all simulated variables. The uppermost section shows these changes strictly for the observed data.

The second section in table A.2 shows a comparison between the means of observed and simulated values during the calibration period.

Table A.1: Means in calibration and validation periods including observed and simulated data (cal.: calibration, val.: validation). Based on the regression method for dryness estimation.

	cal. observed	val. observed	val.-cal.	val.-cal. [%]
θ [$^{\circ}C$]	8.783	10.087	1.304	14.848
R [mm]	2.430	2.491	0.061	2.493
Q_{sw} [W/m^2]	125.580	127.501	1.921	1.530
$Q_{lw(inc.)}$ [W/m^2]	304.558	314.776	10.217	3.355
ϕ [-]	0.781	0.797	0.016	2.017
u [m/s]	0.594	0.663	0.069	11.573
v [m/s]	0.209	0.071	-0.139	-66.203
	cal. observed	cal. simulated	cal.-cal.	cal.-cal. [%]
θ [$^{\circ}C$]	8.783	8.805	0.022	0.255
R [mm]	2.430	2.800	0.369	15.195
Q_{sw} [W/m^2]	125.580	127.247	1.667	1.328
$Q_{lw(inc.)}$ [W/m^2]	304.558	304.671	0.112	0.037
ϕ [-]	0.781	0.780	-0.001	-0.154
u [m/s]	0.594	0.604	0.010	1.648
v [m/s]	0.209	0.220	0.010	4.909
	cal. simulated	val. simulated	val.-cal.	val.-cal. [%]
θ [$^{\circ}C$]	8.805	10.111	1.306	14.831
R [mm]	2.800	2.565	-0.235	-8.386
Q_{sw} [W/m^2]	127.247	131.795	4.547	3.573
$Q_{lw(inc.)}$ [W/m^2]	304.671	310.572	5.901	1.937
ϕ [-]	0.780	0.772	-0.008	-1.017
u [m/s]	0.604	0.615	0.011	1.795
v [m/s]	0.220	0.276	0.056	25.584
	cal. observed	val. simulated	val.-cal.	val.-cal. [%]
θ [$^{\circ}C$]	8.783	10.111	1.328	15.123
R [mm]	2.430	2.565	0.135	5.535
Q_{sw} [W/m^2]	125.580	131.795	6.214	4.949
$Q_{lw(inc.)}$ [W/m^2]	304.558	310.572	6.014	1.975
ϕ [-]	0.781	0.772	-0.009	-1.169
u [m/s]	0.594	0.615	0.021	3.472
v [m/s]	0.209	0.276	0.066	31.749

Table A.2: Means in calibration and validation periods including observed and simulated data (cal.: calibration, val.: validation). Based on the distance method for dryness estimation.

	cal. observed	val. observed	val.-cal.	val.-cal. [%]
θ [$^{\circ}C$]	8.783	10.087	1.304	14.848
R [mm]	2.430	2.491	0.061	2.493
Q_{sw} [W/m^2]	125.580	127.501	1.921	1.530
$Q_{lw(inc.)}$ [W/m^2]	304.558	314.776	10.217	3.355
ϕ [-]	0.781	0.797	0.016	2.017
u [m/s]	0.594	0.663	0.069	11.573
v [m/s]	0.209	0.071	-0.139	-66.203
	cal. observed	cal. simulated	cal.-cal.	cal.-cal. [%]
θ [$^{\circ}C$]	8.783	8.806	0.023	0.260
R [mm]	2.430	2.872	0.441	18.163
Q_{sw} [W/m^2]	125.580	127.199	1.619	1.289
$Q_{lw(inc.)}$ [W/m^2]	304.558	304.672	0.114	0.037
ϕ [-]	0.781	0.780	-0.001	-0.152
u [m/s]	0.594	0.603	0.009	1.457
v [m/s]	0.209	0.218	0.009	4.243
	cal. simulated	val. simulated	val.-cal.	val.-cal. [%]
θ [$^{\circ}C$]	8.806	10.112	1.306	14.830
R [mm]	2.872	3.023	0.151	5.260
Q_{sw} [W/m^2]	127.199	131.741	4.542	3.571
$Q_{lw(inc.)}$ [W/m^2]	304.672	310.574	5.902	1.937
ϕ [-]	0.780	0.772	-0.008	-1.015
u [m/s]	0.603	0.614	0.011	1.810
v [m/s]	0.218	0.275	0.056	25.810
	cal. observed	val. simulated	val.-cal.	val.-cal. [%]
θ [$^{\circ}C$]	8.783	10.112	1.329	15.129
R [mm]	2.430	3.023	0.592	24.379
Q_{sw} [W/m^2]	125.580	131.741	6.161	4.906
$Q_{lw(inc.)}$ [W/m^2]	304.558	310.574	6.015	1.975
ϕ [-]	0.781	0.772	-0.009	-1.166
u [m/s]	0.594	0.614	0.020	3.294
v [m/s]	0.209	0.275	0.065	31.148

B Tables for WeatherCop Extrapolation Test

Tables B.1 and B.2 show results from the extrapolation test in section 4.6.2 depending on dryness probability estimation method.

Table B.1: Means in calibration and validation periods including observed and simulated data (cal.: calibration, val.: validation). Based on the regression method for dryness estimation.

	cal. observed	val. observed	val.-cal.	val.-cal. [%]
theta [$^{\circ}C$]	8.509	9.302	0.793	9.317
R [mm]	0.110	0.105	-0.006	-5.087
sun [min]	11.228	11.437	0.209	1.858
rh [-]	0.786	0.789	0.003	0.383
	cal. simulated	val. simulated	val.-cal.	val.-cal. [%]
theta [$^{\circ}C$]	8.523	9.316	0.793	9.302
R [mm]	0.111	0.107	-0.003	-2.949
sun [min]	11.094	11.762	0.667	6.016
rh [-]	0.790	0.781	-0.009	-1.150
	cal. observed	val. simulated	val.-cal.	val.-cal. [%]
theta [$^{\circ}C$]	8.509	9.316	0.807	9.480
R [mm]	0.110	0.107	-0.003	-2.597
sun [min]	11.228	11.762	0.533	4.749
rh [-]	0.786	0.781	-0.006	-0.719

Table B.2: Means in calibration and validation periods including observed and simulated data (cal.: calibration, val.: validation). Based on the distance method for dryness estimation.

	cal. observed	val. observed	val.-cal.	val.-cal. [%]
theta [$^{\circ}C$]	8.509	9.302	0.793	9.317
R [mm]	0.110	0.105	-0.006	-5.087
sun [min]	11.228	11.437	0.209	1.858
rh [-]	0.786	0.789	0.003	0.383
	cal. simulated	val. simulated	val.-cal.	val.-cal. [%]
theta [$^{\circ}C$]	8.523	9.316	0.793	9.302
R [mm]	0.111	0.107	-0.003	-2.949
sun [min]	11.094	11.762	0.667	6.016
rh [-]	0.790	0.781	-0.009	-1.150
	cal. observed	val. simulated	val.-cal.	val.-cal. [%]
theta [$^{\circ}C$]	8.509	9.316	0.807	9.480
R [mm]	0.110	0.107	-0.003	-2.597
sun [min]	11.228	11.762	0.533	4.749
rh [-]	0.786	0.781	-0.006	-0.719

C Bivariate Copulas

Table C.1: Bivariate copula cumulative distribution functions (*cdf*s) used as candidates in Vine construction. Subscripts (e.g. $_{90}$) refer to rotated versions obtained using the substitutions from section 4.5. Not all rotated copulas passed automated tests (problematic were often the inverse conditional *cdf*s) and were thus discarded.

Name	Copula cdf
Ali-Mikail-Haq	$-\frac{uv}{\theta(u-1)(v-1)-1}$
Ali-Mikail-Haq$_{180}$	$\frac{uv(\theta u+\theta v-\theta-1)}{\theta uv-1}$
Ali-Mikail-Haq$_{270}$	$\frac{uv(\theta(v-1)+1)}{\theta u(v-1)+1}$
Ali-Mikail-Haq$_{90}$	$\frac{uv(\theta(u-1)+1)}{\theta v(u-1)+1}$
Clayton	$\left(-1+v^{-\theta}+u^{-\theta}\right)^{-\frac{1}{\theta}}$
Clayton$_{180}$	$u+v-1+\left(-1+(1-v)^{-\theta}+(1-u)^{-\theta}\right)^{-\frac{1}{\theta}}$
Clayton$_{270}$	$v-\left(-1+(1-u)^{-\theta}+v^{-\theta}\right)^{-\frac{1}{\theta}}$
Gaussian	$\Phi(\Phi^{-1}(u),\Phi^{-1}(v),\rho)$
Gumbel	$e^{-\left((-\log(u))^{\theta}+(-\log(v))^{\theta}\right)^{\frac{1}{\theta}}}$
Gumbel$_{180}$	$u+v-1+e^{-\left((-\log(1-u))^{\theta}+(-\log(1-v))^{\theta}\right)^{\frac{1}{\theta}}}$
Gumbel$_{270}$	$v-e^{-\left((-\log(v))^{\theta}+(-\log(1-u))^{\theta}\right)^{\frac{1}{\theta}}}$
Gumbel$_{90}$	$u-e^{-\left((-\log(u))^{\theta}+(-\log(1-v))^{\theta}\right)^{\frac{1}{\theta}}}$
Gumbel-Barnett	$uve^{-\theta\log(u)\log(v)}$
Independence	uv
Joe	$1-\left(-(1-u)^{\theta}(1-v)^{\theta}+(1-u)^{\theta}+(1-v)^{\theta}\right)^{\frac{1}{\theta}}$
Nelsen10	$2^{\frac{1}{\theta}}\left(2-2v^{-\theta}-2u^{-\theta}+4u^{-\theta}v^{-\theta}\right)^{-\frac{1}{\theta}}$
Nelsen10$_{180}$	$2^{\frac{1}{\theta}}\left(2-2(1-v)^{-\theta}-2(1-u)^{-\theta}+4(1-u)^{-\theta}(1-v)^{-\theta}\right)^{-\frac{1}{\theta}}+u+v-1$
Nelsen10$_{270}$	$-2^{\frac{1}{\theta}}\left(2-2(1-u)^{-\theta}-2v^{-\theta}+4v^{-\theta}(1-u)^{-\theta}\right)^{-\frac{1}{\theta}}+v$
Nelsen10$_{90}$	$-2^{\frac{1}{\theta}}\left(2-2(1-v)^{-\theta}-2u^{-\theta}+4u^{-\theta}(1-v)^{-\theta}\right)^{-\frac{1}{\theta}}+u$

Nelsen12	$\frac{1}{\left(\left(\frac{1-u}{u}\right)^{\theta}+\left(\frac{1-v}{v}\right)^{\theta}\right)^{\frac{1}{\theta}}+1}$
Nelsen12$_{180}$	$u+v-1+\frac{1}{\left(\left(-\frac{u}{u-1}\right)^{\theta}+\left(-\frac{v}{v-1}\right)^{\theta}\right)^{\frac{1}{\theta}}+1}$
Nelsen12$_{270}$	$v-\frac{1}{\left(\left(-\frac{u}{u-1}\right)^{\theta}+\left(-\frac{v-1}{v}\right)^{\theta}\right)^{\frac{1}{\theta}}+1}$
Nelsen12$_{90}$	$u-\frac{1}{\left(\left(-\frac{u-1}{u}\right)^{\theta}+\left(-\frac{v}{v-1}\right)^{\theta}\right)^{\frac{1}{\theta}}+1}$
Nelsen13	$e^{1-\left((1-\log(u))^{\theta}+(1-\log(v))^{\theta}-1\right)^{\frac{1}{\theta}}}$
Nelsen13$_{180}$	$u+v+e^{1-\left((1-\log(1-u))^{\theta}+(1-\log(1-v))^{\theta}-1\right)^{\frac{1}{\theta}}}-1$
Nelsen13$_{270}$	$v-e^{1-\left((1-\log(v))^{\theta}+(1-\log(1-u))^{\theta}-1\right)^{\frac{1}{\theta}}}$
Nelsen13$_{90}$	$u-e^{1-\left((1-\log(u))^{\theta}+(1-\log(1-v))^{\theta}-1\right)^{\frac{1}{\theta}}}$
Nelsen14	$\left(\frac{1}{\left(\left(-1+u^{-\frac{1}{\theta}}\right)^{\theta}+\left(-1+v^{-\frac{1}{\theta}}\right)^{\theta}\right)^{\frac{1}{\theta}}+1}\right)^{\theta}$
Nelsen14$_{180}$	$u+v+\left(\frac{1}{\left(\left(-1+(1-u)^{-\frac{1}{\theta}}\right)^{\theta}+\left(-1+(1-v)^{-\frac{1}{\theta}}\right)^{\theta}\right)^{\frac{1}{\theta}}+1}\right)^{\theta}-1$
Nelsen14$_{270}$	$v-\left(\frac{1}{\left(\left(-1+v^{-\frac{1}{\theta}}\right)^{\theta}+\left(-1+(1-u)^{-\frac{1}{\theta}}\right)^{\theta}\right)^{\frac{1}{\theta}}+1}\right)^{\theta}$
Nelsen14$_{90}$	$u-\left(\frac{1}{\left(\left(-1+u^{-\frac{1}{\theta}}\right)^{\theta}+\left(-1+(1-v)^{-\frac{1}{\theta}}\right)^{\theta}\right)^{\frac{1}{\theta}}+1}\right)^{\theta}$
Plackett	$\frac{(\theta-1)(u+v)-\sqrt{-4\theta uv(\theta-1)+((\theta-1)(u+v)+1)^2}+1}{2(\theta-1)}$

D Resampler Output Statistics

Table D.1: Some basic statistics to compare source and generated data in a univariate sense. The marginals of the source data where transformed to standard-normal an the resampled and VAR-generated time series are not back-transformed into the measurement dimensions. All generated time series (K-Nearest Neighbors (KNN), VAR, VG, WeatherCop) have a length of 1000 years. VAR refers to results from a VAR-process with gaussian noise, whereas VG uses phase-randomized residuals.

Statistic	Variable	Source	KNN	VAR	VG	WCop
mean	Temperature	0.002	-0.077	-0.000	0.002	0.002
	Precipitation	0.034	0.067	-0.003	0.066	0.069
	Short-wave radiation	0.025	-0.031	0.001	0.049	0.015
	Incident long-wave radiation	0.000	-0.016	-0.001	-0.000	0.018
	Relative humidity	-0.004	0.129	-0.000	-0.003	0.020
	Eastward wind speed	0.008	0.017	-0.003	0.011	0.016
	Northward wind speed	0.003	0.052	-0.007	-0.000	0.008
std. deviation	Temperature	0.960	0.817	0.958	0.947	0.960
	Precipitation	1.033	0.878	1.032	1.040	1.028
	Short-wave radiation	0.964	0.861	0.967	0.965	0.957
	Incident long-wave radiation	0.980	0.830	0.980	0.987	0.985
	Relative humidity	0.988	0.842	0.985	0.969	0.950
	Eastward wind speed	0.959	0.801	0.961	0.961	0.958
	Northward wind speed	0.965	0.827	0.968	0.967	0.967
skewness	Temperature	-0.129	-0.063	-0.002	-0.045	-0.005
	Precipitation	0.001	0.073	0.002	-0.016	0.004
	Short-wave radiation	-0.012	0.032	-0.002	0.010	0.034
	Incident long-wave radiation	-0.163	-0.155	-0.008	-0.039	-0.021
	Relative humidity	0.003	-0.021	-0.000	-0.009	-0.064
	Eastward wind speed	0.050	0.188	0.001	-0.007	-0.041
	Northward wind speed	0.023	0.141	0.004	-0.002	0.016
kurtosis	Temperature	**-0.202**	**-0.206**	**0.020**	**0.023**	**0.022**
	Precipitation	**-0.130**	**-0.049**	**0.063**	0.002	0.012
	Short-wave radiation	**-0.385**	**-0.307**	**0.018**	-0.008	**0.074**
	Incident long-wave radiation	**-0.348**	**-0.228**	-0.002	**0.020**	-0.001
	Relative humidity	**-0.169**	**-0.131**	**0.027**	0.006	**0.063**
	Eastward wind speed	0.018	**0.152**	-0.009	0.011	-0.001
	Northward wind speed	0.061	**0.180**	-0.011	0.007	**-0.048**
range	Temperature	6.991	6.991	9.065	8.980	9.817
	Precipitation	7.651	7.651	9.965	9.722	9.040

	Short-wave radiation	6.265	6.265	8.919	9.908	9.657
	Incident long-wave radiation	6.192	6.192	9.222	9.062	9.296
	Relative humidity	7.256	7.256	9.677	9.126	8.955
	Eastward wind speed	6.102	6.102	8.943	8.584	9.091
	Northward wind speed	6.110	6.110	9.454	8.780	9.441
interq. range	Temperature	1.343	1.129	1.288	1.278	1.294
	Precipitation	1.449	1.242	1.385	1.401	1.386
	Short-wave radiation	1.379	1.201	1.301	1.304	1.282
	Incident long-wave radiation	1.379	1.161	1.322	1.329	1.330
	Relative humidity	1.360	1.160	1.326	1.307	1.274
	Eastward wind speed	1.289	1.084	1.298	1.299	1.294
	Northward wind speed	1.239	1.056	1.310	1.303	1.315
minimum	Temperature	-3.965	-3.965	-4.566	-4.705	-4.162
	Precipitation	-3.823	-3.823	-4.929	-4.863	-4.457
	Short-wave radiation	-2.861	-2.861	-4.597	-5.080	-4.846
	Incident long-wave radiation	-3.310	-3.310	-4.433	-4.563	-4.636
	Relative humidity	-4.133	-4.133	-4.918	-4.815	-4.728
	Eastward wind speed	-3.055	-3.055	-4.404	-4.413	-4.724
	Northward wind speed	-3.030	-3.030	-4.228	-4.366	-4.609
lower quartile	Temperature	-0.646	-0.629	-0.644	-0.634	-0.645
	Precipitation	-0.675	-0.519	-0.695	-0.633	-0.624
	Short-wave radiation	-0.661	-0.629	-0.649	-0.603	-0.629
	Incident long-wave radiation	-0.677	-0.582	-0.661	-0.662	-0.644
	Relative humidity	-0.678	-0.445	-0.663	-0.655	-0.610
	Eastward wind speed	-0.625	-0.531	-0.651	-0.638	-0.627
	Northward wind speed	-0.590	-0.466	-0.662	-0.652	-0.651
median	Temperature	0.021	-0.075	0.001	0.006	0.004
	Precipitation	0.014	0.018	-0.005	0.070	0.066
	Short-wave radiation	0.028	-0.036	0.002	0.047	0.012
	Incident long-wave radiation	0.038	0.008	-0.002	0.006	0.021
	Relative humidity	-0.024	0.123	-0.000	-0.002	0.029
	Eastward wind speed	-0.038	-0.047	-0.003	0.013	0.024
	Northward wind speed	-0.031	-0.002	-0.009	0.002	0.006
upper quartile	Temperature	0.697	0.500	0.644	0.644	0.649
	Precipitation	0.775	0.723	0.690	0.768	0.762
	Short-wave radiation	0.719	0.572	0.652	0.701	0.652
	Incident long-wave radiation	0.702	0.579	0.661	0.667	0.686
	Relative humidity	0.682	0.716	0.663	0.652	0.663
	Eastward wind speed	0.664	0.553	0.647	0.661	0.667
	Northward wind speed	0.649	0.589	0.648	0.651	0.664
maximum	Temperature	3.026	3.026	4.499	4.275	5.656
	Precipitation	3.828	3.828	5.036	4.858	4.583
	Short-wave radiation	3.404	3.404	4.322	4.828	4.811
	Incident long-wave radiation	2.882	2.882	4.789	4.500	4.660
	Relative humidity	3.123	3.123	4.759	4.311	4.227
	Eastward wind speed	3.047	3.047	4.538	4.171	4.367
	Northward wind speed	3.080	3.080	5.227	4.414	4.832

Institut für Wasser- und Umweltsystemmodellierung
Universität Stuttgart

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

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

Vorstand *(Stand 21.05.2021)*
Prof. Dr. rer. nat. Dr.-Ing. A. Bárdossy
Prof. Dr.-Ing. R. Helmig
Prof. Dr.-Ing. W. Nowak
Prof. Dr.-Ing. S. Wieprecht
Prof. Dr. J.A. Sander Huisman
Jürgen Braun, PhD
apl. Prof. Dr.-Ing. H. Class
PD Dr.-Ing. Claus Haslauer
Stefan Haun, PhD
apl. Prof. Dr.-Ing. Sergey Oladyshkin
Dr. rer. nat. J. Seidel
Dr.-Ing. K. Terheiden

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

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

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

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

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

***VEGAS,* Versuchseinrichtung zur Grundwasser- und Altlastensanierung**
Leiter: Jürgen Braun, PhD
PD Dr.-Ing. Claus Haslauer

Verzeichnis der Mitteilungshefte

1 Röhnisch, Arthur: *Die Bemühungen um eine Wasserbauliche Versuchsanstalt an der Technischen Hochschule Stuttgart*, und
Fattah Abouleid, Abdel: *Beitrag zur Berechnung einer in lockeren Sand gerammten, zweifach verankerten Spundwand*, 1963

2 Marotz, Günter: *Beitrag zur Frage der Standfestigkeit von dichten Asphaltbelägen im Großwasserbau*, 1964

3 Gurr, Siegfried: *Beitrag zur Berechnung zusammengesetzter ebener Flächentrag-werke unter besonderer Berücksichtigung ebener Stauwände, mit Hilfe von Rand-wert- und Lastwertmatrizen*, 1965

4 Plica, Peter: *Ein Beitrag zur Anwendung von Schalenkonstruktionen im Stahlwasserbau*, und
Petrikat, Kurt: *Möglichkeiten und Grenzen des wasserbaulichen Versuchswesens*, 1966

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

58 Kirchenbaur, Klaus: *Mikroprozessorgesteuerte Erfassung instationärer Druckfelder am Beispiel seegangsbelasteter Baukörper*, 1985, ISBN 3-921694-58-2

59 Kobus, Helmut (Hrsg.): *Modellierung des großräumigen Wärme- und Schadstofftransports im Grundwasser*, Tätigkeitsbericht 1984/85 (DFG-Forschergruppe an den Universitäten Hohenheim, Karlsruhe und Stuttgart), 1985, ISBN 3-921694-59-0, *vergriffen*

60 Spitz, Karlheinz: *Dispersion in porösen Medien: Einfluß von Inhomogenitäten und Dichteunterschieden*, 1985, ISBN 3-921694-60-4, *vergriffen*

61 Kobus, Helmut: *An Introduction to Air-Water Flows in Hydraulics*, 1985, ISBN 3-921694-61-2

62 Kaleris, Vassilios: *Erfassung des Austausches von Oberflächen- und Grundwasser in horizontalebenen Grundwassermodellen*, 1986, ISBN 3-921694-62-0

63 Herr, Michael: *Grundlagen der hydraulischen Sanierung verunreinigter Porengrundwasserleiter*, 1987, ISBN 3-921694-63-9

64 Marx, Walter: *Berechnung von Temperatur und Spannung in Massenbeton infolge Hydratation*, 1987, ISBN 3-921694-64-7

65 Koschitzky, Hans-Peter: *Dimensionierungskonzept für Sohlbelüfter in Schußrinnen zur Vermeidung von Kavitationsschäden*, 1987, ISBN 3-921694-65-5

66 Kobus, Helmut (Hrsg.): *Modellierung des großräumigen Wärme- und Schadstofftransports im Grundwasser*, Tätigkeitsbericht 1986/87 (DFG-Forschergruppe an den Universitäten Hohenheim, Karlsruhe und Stuttgart) 1987, ISBN 3-921694-66-3

67 Söll, Thomas: *Berechnungsverfahren zur Abschätzung anthropogener Temperaturanomalien im Grundwasser*, 1988, ISBN 3-921694-67-1

68 Dittrich, Andreas; Westrich, Bernd: *Bodenseeufererosion, Bestandsaufnahme und Bewertung*, 1988, ISBN 3-921694-68-X, *vergriffen*

69 Huwe, Bernd; van der Ploeg, Rienk R.: *Modelle zur Simulation des Stickstoffhaushaltes von Standorten mit unterschiedlicher landwirtschaftlicher Nutzung*, 1988, ISBN 3-921694-69-8, *vergriffen*

70 Stephan, Karl: *Integration elliptischer Funktionen*, 1988, ISBN 3-921694-70-1

71 Kobus, Helmut; Zilliox, Lothaire (Hrsg.): *Nitratbelastung des Grundwassers, Auswirkungen der Landwirtschaft auf die Grundwasser- und Rohwasserbeschaffenheit und Maßnahmen zum Schutz des Grundwassers*. Vorträge des deutsch-französischen Kolloquiums am 6. Oktober 1988, Universitäten Stuttgart und Louis Pasteur Strasbourg (Vorträge in deutsch oder französisch, Kurzfassungen zweisprachig), 1988, ISBN 3-921694-71-X

72 Soyeaux, Renald: *Unterströmung von Stauanlagen auf klüftigem Untergrund unter Berücksichtigung laminarer und turbulenter Fließzustände*,1991, ISBN 3-921694-72-8

73 Kohane, Roberto: *Berechnungsmethoden für Hochwasserabfluß in Fließgewässern mit überströmten Vorländern*, 1991, ISBN 3-921694-73-6

74 Hassinger, Reinhard: *Beitrag zur Hydraulik und Bemessung von Blocksteinrampen in flexibler Bauweise*, 1991, ISBN 3-921694-74-4, vergriffen

75 Schäfer, Gerhard: *Einfluß von Schichtenstrukturen und lokalen Einlagerungen auf die Längsdispersion in Porengrundwasserleitern*, 1991, ISBN 3-921694-75-2

76 Giesecke, Jürgen: *Vorträge, Wasserwirtschaft in stark besiedelten Regionen; Umweltforschung mit Schwerpunkt Wasserwirtschaft*, 1991, ISBN 3-921694-76-0

77 Huwe, Bernd: *Deterministische und stochastische Ansätze zur Modellierung des Stickstoffhaushalts landwirtschaftlich genutzter Flächen auf unterschiedlichem Skalenniveau*, 1992, ISBN 3-921694-77-9, vergriffen

78 Rommel, Michael: *Verwendung von Kluftdaten zur realitätsnahen Generierung von Kluftnetzen mit anschließender laminar-turbulenter Strömungsberechnung*, 1993, ISBN 3-92 1694-78-7

79 Marschall, Paul: *Die Ermittlung lokaler Stofffrachten im Grundwasser mit Hilfe von Einbohrloch-Meßverfahren*, 1993, ISBN 3-921694-79-5, vergriffen

80 Ptak, Thomas: *Stofftransport in heterogenen Porenaquiferen: Felduntersuchungen und stochastische Modellierung*, 1993, ISBN 3-921694-80-9, vergriffen

81 Haakh, Frieder: *Transientes Strömungsverhalten in Wirbelkammern,* 1993, ISBN 3-921694-81-7

82 Kobus, Helmut; Cirpka, Olaf; Barczewski, Baldur; Koschitzky, Hans-Peter: *Versuchseinrichtung zur Grundwasser- und Altlastensanierung VEGAS, Konzeption und Programmrahmen*, 1993, ISBN 3-921694-82-5

83 Zang, Weidong: *Optimaler Echtzeit-Betrieb eines Speichers mit aktueller Abflußregenerierung*, 1994, ISBN 3-921694-83-3, vergriffen

84 Franke, Hans-Jörg: *Stochastische Modellierung eines flächenhaften Stoffeintrages und Transports in Grundwasser am Beispiel der Pflanzenschutzmittelproblematik*, 1995, ISBN 3-921694-84-1

85 Lang, Ulrich: *Simulation regionaler Strömungs- und Transportvorgänge in Karstaquiferen mit Hilfe des Doppelkontinuum-Ansatzes: Methodenentwicklung und Parameteridentifikation*, 1995, ISBN 3-921694-85-X, vergriffen

86 Helmig, Rainer: *Einführung in die Numerischen Methoden der Hydromechanik*, 1996, ISBN 3-921694-86-8, vergriffen

87 Cirpka, Olaf: *CONTRACT: A Numerical Tool for Contaminant Transport and Chemical Transformations - Theory and Program Documentation -*, 1996, ISBN 3-921694-87-6

88 Haberlandt, Uwe: *Stochastische Synthese und Regionalisierung des Niederschlages für Schmutzfrachtberechnungen*, 1996, ISBN 3-921694-88-4

89 Croisé, Jean: *Extraktion von flüchtigen Chemikalien aus natürlichen Lockergesteinen mittels erzwungener Luftströmung*, 1996, ISBN 3-921694-89-2, vergriffen

90 Jorde, Klaus: *Ökologisch begründete, dynamische Mindestwasserregelungen bei Ausleitungskraftwerken*, 1997, ISBN 3-921694-90-6, vergriffen

91 Helmig, Rainer: *Gekoppelte Strömungs- und Transportprozesse im Untergrund - Ein Beitrag zur Hydrosystemmodellierung-*, 1998, ISBN 3-921694-91-4, vergriffen

92 Emmert, Martin: *Numerische Modellierung nichtisothermer Gas-Wasser Systeme in porösen Medien,* 1997, ISBN 3-921694-92-2

93 Kern, Ulrich: *Transport von Schweb- und Schadstoffen in staugeregelten Fließgewässern am Beispiel des Neckars*, 1997, ISBN 3-921694-93-0, vergriffen

94 Förster, Georg: *Druckstoßdämpfung durch große Luftblasen in Hochpunkten von Rohrleitungen 1997*, ISBN 3-921694-94-9

95 Cirpka, Olaf: *Numerische Methoden zur Simulation des reaktiven 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

Die Mitteilungshefte ab der Nr. 134 (Jg. 2005) stehen als pdf-Datei über die Homepage des Instituts: www.iws.uni-stuttgart.de zur Verfügung.