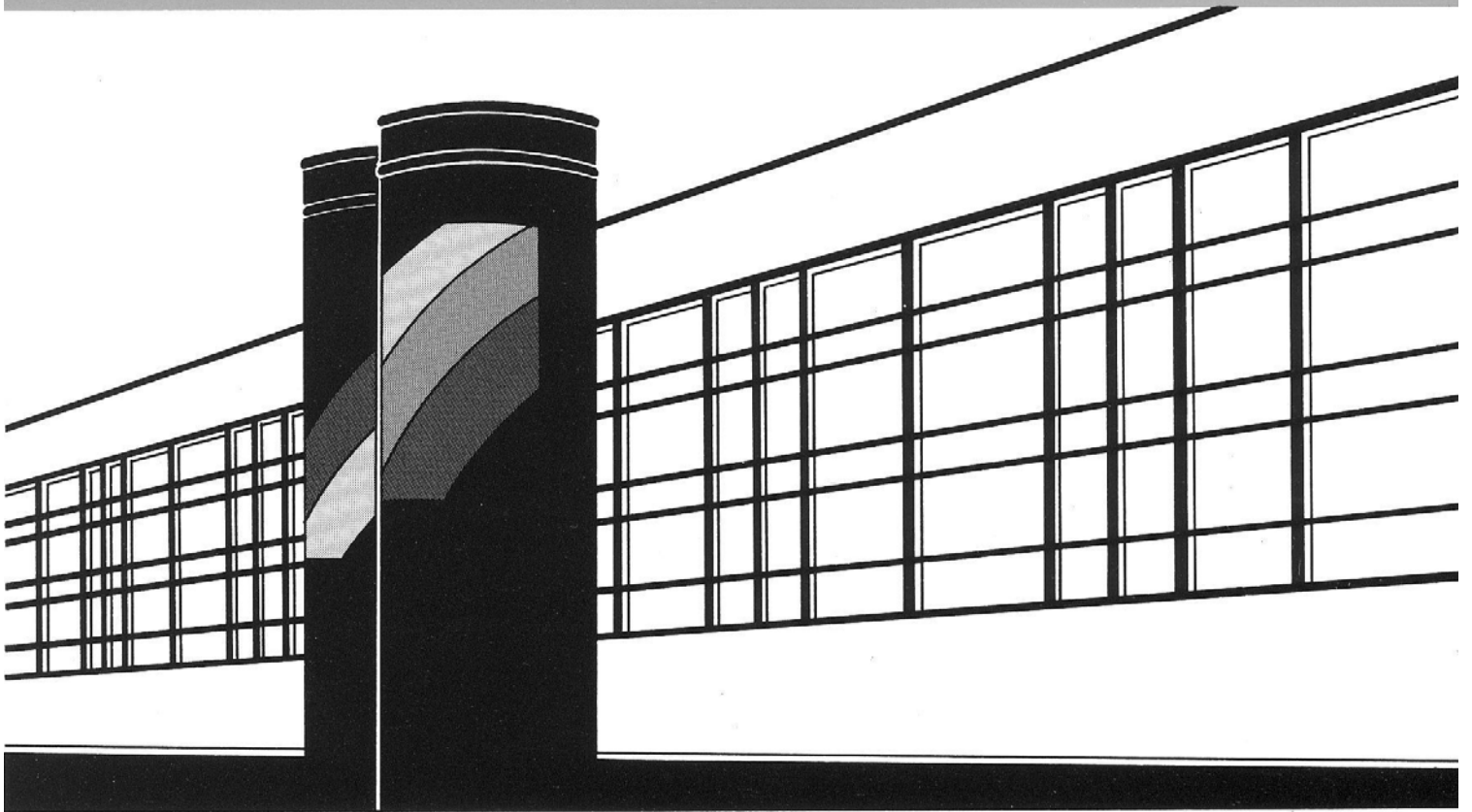


Institut für Wasserbau · Universität Stuttgart

# *Mitteilungen*



Heft 187    Jing Li

Application of Copulas as a New  
Geostatistical Tool





# **Application of Copulas as a New Geostatistical Tool**

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

Vorgelegt von  
**Jing Li**  
aus Shanghai, China

Hauptberichter: (Prof. Dr. rer. nat. Dr.-Ing. András Bárdossy)  
Mitberichter: (Prof. Dr. Ir. Sjeord van der Zee)  
(Prof. Dr. Insa Neuweiler)

Tag der mündlichen Prüfung: 15. Januar 2010

Institut für Wasserbau der Universität Stuttgart  
2010





Heft 187    Application of Copulas as a  
New Geostatistical Tool

von  
Dr.-Ing.  
Jing Li

## D93 Application of Copulas as a New Geostatistical Tool

### **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

Li, Jing:

Application of Copulas as a New Geostatistical Tool / von Jing Li. Institut für Wasserbau, Universität Stuttgart. - Stuttgart: Inst. für Wasserbau, 2010

(Mitteilungen / Institut für Wasserbau, Universität Stuttgart: H. 187)

Zugl.: Stuttgart, Univ., Diss., 2010

ISBN 978-3-933761-91-0

NE: Institut für Wasserbau <Stuttgart>: Mitteilungen

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

Herausgegeben 2010 vom Eigenverlag des Instituts für Wasserbau  
Druck: Document Center S. Kästl, Ostfilder

# Acknowledgement

First of all, I would like to express my heartfelt thanks to Prof. András Bárdossy, who has been illuminating the way of my study. His wealth of unique ideas always motivates me during my PhD work. His guidance in capturing the essence of a problem in a relatively simple assumptions has been a lasting influence on my way of thinking and solving the research problems. His optimistic attitude gives me a lot of inspiration. I also appreciate his support and encouragement to me for various academic exchanges and his concern for my future career.

My special thanks to Prof. Sjoerd van der Zee from University of Wageningen for his warm-hearted support and valuable suggestions during my PhD work and his kind recommendations and help for my future. I sincerely thank Prof. Insa Neuweiler from university of Hannover. She also gave me a lot of valuable input for my research. Her patient explanations for all my questions is really a great help for my study.

I would like to acknowledge the International Research Training Group (NUPUS) which is a joint research project funded by the German Research Foundation (DFG) and the Netherlands Organization for Scientific Research (NWO). It offers me a wonderful platform for academic and research exchanges. Hereby, I would like to say a lot of thanks to Prof. Rainer Helmig who is one of the sponsors of the NUPUS project. I am very grateful for all his helps and supports during both my Master and PhD study.

I also would like to acknowledge ENWAT International Doctoral Program of University of Stuttgart for providing a very good academic framework for my study. A lot of thanks to Dr.-Ing. Gabriele M. Hartmann for the warmhearted helps she offered to me.

I would like to extend my thanks to Dr. -Ing. Holger Class, who is the supervisor of my Master's thesis. His brilliant guidance helped in paving my way towards the following research work.

I also wish to thank Dr. Ger de Rooij from University of Wageningen, who helped me in acquiring the dataset of Las Cruces Trench Site which acts as an important practical part of this thesis.

My thanks further extend to all those unmentioned colleagues and friends at the Institut für Wasserbau who gave me helpful advices for my work and made this pleasant working atmosphere.

Last but not the least, I should say that my husband Min also contributes a lot. His unconditional love, understanding and tolerance always encourage me during any hard time. I thank my parents, my aunts and my uncles in China for their consistent love and support.



# Contents

<b>1</b>	<b>General Introduction</b>	<b>1</b>
1.1	Background and Motivations . . . . .	1
1.2	Scope and Organization of the Thesis . . . . .	4
<b>2</b>	<b>Introduction to Copulas</b>	<b>6</b>
<b>3</b>	<b>Methodology of using Copulas to investigate Spatial Dependence</b>	<b>11</b>
3.1	Basic Hypothesis of Spatial Copulas . . . . .	11
3.2	Empirical Bivariate Spatial Copulas . . . . .	12
3.3	Measure of Dependence . . . . .	14
3.3.1	Spearman's rho . . . . .	15
3.3.2	Measure of asymmetry . . . . .	16
3.3.3	Euler Characteristics . . . . .	20
<b>4</b>	<b>Theoretical Spatial Copulas</b>	<b>25</b>
4.1	Gaussian Copulas . . . . .	26
4.2	V-transformed Normal Copulas . . . . .	27
4.3	Maximum Normal Copulas . . . . .	31
<b>5</b>	<b>Model Inference</b>	<b>37</b>
5.1	Model Inference . . . . .	37
5.2	Stability and Reliability Test . . . . .	40
<b>6</b>	<b>Interpolation using Copulas</b>	<b>42</b>
6.1	Procedure of Interpolation using Copulas . . . . .	43
6.1.1	Example of interpolation using different copula models . . . . .	46
6.1.2	Neighborhood selection . . . . .	50
6.1.3	Properties of the copula based interpolation . . . . .	50
6.2	Application . . . . .	51
6.2.1	Empirical copulas . . . . .	53
6.2.2	Parameter estimation and statistical test . . . . .	55
6.2.3	Interpolation . . . . .	61
6.2.4	Confidence intervals . . . . .	63
6.2.5	Cross validation and split sampling results . . . . .	65
6.3	Summary of Chapter 6 . . . . .	68

---

<b>7</b>	<b>Simulation of Random Fields using non-Gaussian Dependence</b>	<b>70</b>
7.1	Unconditional simulation . . . . .	72
7.2	Sequential conditional simulation . . . . .	77
7.3	Parameter Optimization for Correlation Structure . . . . .	81
7.4	Dataset Investigation . . . . .	84
7.5	Summary of Chapter 7 . . . . .	98
<b>8</b>	<b>Observation Network Design</b>	<b>99</b>
8.1	The Problem of Network Design . . . . .	99
8.2	Description of the proposed methodology . . . . .	100
8.3	One Synthetical Example . . . . .	104
8.4	Application to Water Quality parameters . . . . .	106
8.5	Summary of Chapter 8 . . . . .	108
<b>9</b>	<b>Summary and Outlook</b>	<b>110</b>
9.1	Summary . . . . .	110
9.2	Outlook . . . . .	112

# List of Figures

2.1	Scatter plots of joint probability densities (upper two) of two pairs of random variables and the corresponding copula densities (lower two) . . . . .	8
2.2	Bivariate Gaussian copula (left) and corresponding copula density with correlation equal to 0.85. . . . .	9
3.1	Procedure of calculating the empirical copula from a dataset. . . . .	13
3.2	Empirical copulas of chloride concentration for the separation lengths of 3 km (left), 6 km (middle) and 9 km (right). . . . .	14
3.3	Variogram (left) and rank correlation over distance curve (right) of chloride. . . . .	16
3.4	The asymmetry over distance curve for chloride concentration. . . . .	17
3.5	Bivariate copula densities with different asymmetry properties. . . . .	18
3.6	Two hydraulic conductivity fields sharing the same marginal distribution and rank correlation structure. . . . .	19
3.7	Marginal distributions and rank correlation structures of the two hydraulic conductivity fields. . . . .	19
3.8	Assymetry over distance curves of the two hydraulic conductivity fields. . . . .	20
3.9	Empirical bivariate copula densities for the shortest distance class of the hydraulic conductivity fields shown in Figure 3.6. . . . .	20
3.10	Binary images according to different threshold values $t$ of a spatial field. Black color indicates values above $t$ , while white color indicates values below $t$ . . . . .	22
3.11	The original spatial field (left panel) of the binary images shown in Figure 3.10 and its Euler characteristic against threshold plot (right). . . . .	22
3.12	Comparison of empirical copulas (right) and Euler characteristic curves (left) of two spatial fields. . . . .	23
4.1	Bivariate Gaussian copula densities with different correlation $\rho$ values . . . . .	27
4.2	V-shaped transformations. . . . .	28
4.3	Consequence of the V-shaped transformatoin with parameters: $m = 0.0$ , $k = 3.0$ and $\alpha = 1.0$ (left chart); $m = 0.0$ , $k = 2.5$ and $\alpha = 0.5$ (middle), $m = 1.0$ , $k = 2.5$ and $\alpha = 0.5$ (right) . . . . .	28
4.4	Bivariate copula densities of the random variables after the transformations in Figure 4.3 corresponding to the $\rho = 0.8$ of the underlying bivariate normal distribution. . . . .	29
4.5	Bivariate copula density of $v$ -transformed normal copula with different parameter values. . . . .	31
4.6	Bivariate copula density of a $v$ -transformed normal copula (left) and its reversed version (right). . . . .	32



4.7	<i>Bivariate copula densities of maximum normal copula model with difference parameter settings.</i>	34
4.8	<i>Examples of realizations of the maximum normal copulas with different parameters in the copula space.</i>	34
4.9	<i>Contour plot of asymmetry against the correlations of two underlying bivariate Gaussian variables for maximum normal copula with parameters: <math>m = 0.5</math> and <math>\sigma = 0.5</math>.</i>	35
4.10	<i>Bivariate copula densities of maximum normal copula with parameters: <math>m = 0.5</math> and <math>\sigma = 0.5</math> for different <math>\rho</math> combinations.</i>	36
5.1	<i>Discretization of the studied domain for parameterization purpose</i>	38
5.2	<i>An example of the distribution of the sizes of the delineated subsets for parameterization purpose.</i>	38
5.3	<i>Contour plots of likelihoods for parameter estimation of a simulated field. Right panel corresponds to the result when the domain is divided into subsets containing 3 nodes, middle one is for subsets of 6 nodes while the left one is for subsets of 10 nodes.</i>	41
6.1	<i>Heteroskedasticity</i>	43
6.2	<i>Configuration of the test field. The dots are the observation locations, while the cross is interpolation location. The values listed near to the dots are the cdf values of the observations (right panel for the first synthetic observation set and left panel for the second one).</i>	47
6.3	<i>Estimation distribution density of the cdf values at the target point using a Gaussian copula (left panel) and a <math>v</math>-transformed normal copula (right panel).</i>	47
6.4	<i>Conditioning values after imposing a normal marginal distribution (<math>N(1,1)</math>).</i>	48
6.5	<i>Conditioning values after imposing an exponential marginal distribution (<math>E(1)</math>).</i>	48
6.6	<i>Estimation distribution density of the cdf values at the target point using a Gaussian copula (left panel) and a <math>v</math>-transformed normal copula (right panel).</i>	49
6.7	<i>Estimated distribution density of the cdf values (left panel) and the real values (right panel) at the selected interpolation point in the studied domain using different number of neighboring observations for interpolation.</i>	51
6.8	<i>Location of the domain under study and the observation wells.</i>	52
6.9	<i>Empirical copulas of the five groundwater quality parameters.</i>	54
6.10	<i>Contour plots of log likelihood values against the <math>v</math>-transformed normal copula parameters <math>m</math> and <math>k</math> for nitrate concentration. The left panel is the result corresponding to the partition of the domain into subsets containing 6 points, while the right one corresponds to the partition of the domain into subsets containing 11 points.</i>	56
6.11	<i>Statistical boundaries of the rank correlation over distance structures of the Gaussian realizations compared with the empirical ones.</i>	57
6.12	<i>Statistical boundaries of the rank correlation over distance structures of the <math>v</math>-transformed normal copula realizations compared with the empirical ones.</i>	58
6.13	<i>Statistical boundaries of asymmetry over distance structures of the Gaussian realizations compared with the empirical ones.</i>	59
6.14	<i>Statistical boundaries of asymmetry over distance structures of the <math>v</math>-transformed normal copula realizations compared with the empirical ones.</i>	60

6.15	<i>Interpolated mean values of nitrate concentration using a Gaussian copula (left) and a <math>v</math>-transformed normal copula (right).</i> . . . . .	61
6.16	<i>Interpolated means of sulfate concentration obtained from the <math>v</math>-transformed normal copula (left) and Ordinary Kriging (right).</i> . . . . .	62
6.17	<i>Interpolated means of pH obtained from the <math>v</math>-transformed normal copula (left) and Ordinary Kriging (right).</i> . . . . .	63
6.18	<i>Difference maps between the interpolated means obtained from <math>v</math>-transformed normal copula and Ordinary Kriging for pH (left) and sulfate concentration (right).</i> . . . . .	64
6.19	<i>Zoomed maps of pH around Tübingen. Left one is the difference map, middle one is the <math>v</math>-copula interpolated map and right one is the kriging map.</i> . . . . .	64
6.20	<i>Interpolated mean values (left) and median values (right) of dissolved oxygen concentration using a <math>v</math>-transformed normal copula.</i> . . . . .	65
6.21	<i>90% confidence intervals of dissolved oxygen concentration (left) and pH value (right) obtained from the <math>v</math>-transformed normal copula.</i> . . . . .	66
7.1	<i>Empirical copulas of a multi-Gaussian realization for the separation lengths of 11%, 33% and 56% correlation length (from left to right).</i> . . . . .	73
7.2	<i>A spatial multi-Gaussian realization in copula space.</i> . . . . .	73
7.3	<i>An example of non-monotonic transformation (right upper corner) and resultant spatial realization (left upper corner) and its corresponding empirical copulas (bottom row) for the separation lengths of 12.5%, 18.75% and 31.25% correlation length.</i> . . . . .	75
7.4	<i>Another example of non-monotonic transformation (right upper corner) and resultant spatial realization (left upper corner) and its corresponding empirical copulas (bottom row) for the separation lengths of 13%, 19% and 32% correlation length.</i> . . . . .	76
7.5	<i>Example of realizations from combined processes.</i> . . . . .	77
7.6	<i>Realizations of conditional simulation using <math>v</math>-transformed normal copula.</i> . . . . .	80
7.7	<i>Example of fitting a polynomial to the relationship <math>\rho = f(r)</math> for a <math>v</math>-transformed normal copula under <math>m = 0.0</math>, <math>k = 2.5</math> and <math>\alpha = 0.5</math>.</i> . . . . .	83
7.8	<i>Comparison of Pearson's rho over distance structures calculated from step 6 and from step 7 before (left) and after several iterations for optimization (right).</i> . . . . .	83
7.9	<i>Classified post plot of the field saturated hydraulic conductivity.</i> . . . . .	84
7.10	<i>Histogram of the logarithm of the saturated hydraulic conductivity.</i> . . . . .	85
7.11	<i>Empirical copulas of saturated hydraulic conductivity for omni-direction (upper row) and horizontal direction (bottom row). The separation lengths from left to right are 0.5 m, 1.0 m, 1.5 m, 2.0 m and 2.5 m.</i> . . . . .	85
7.12	<i>Rank correlation over distance curves (left) and asymmetry over distance (right) curves of the saturated hydraulic conductivity field.</i> . . . . .	86
7.13	<i>Histogram of the radii of the circumcircles of the subsets.</i> . . . . .	87
7.14	<i>Comparison between the theoretical bivariate copulas of the three parameterized models: a Gaussian copula model (first row), a <math>v</math>-transformed normal copula model (second row), a maximum normal copula model (third row), and the empirical copulas along the horizontal direction of the saturated hydraulic conductivity field.</i> . . . . .	88
7.15	<i>Comparison between the theoretical rank correlation (left) and asymmetry (right) over distance structures of the three parameterized copula models and the empirical ones of the saturated hydraulic conductivity field along the horizontal direction.</i> . . . . .	89

7.16	<i>Rotation of coordinate system (new system is denoted by ' ). . . . .</i>	90
7.17	<i>Discretization of the quadrant to calculate the omnidirectional separation lengths in the new coordinate system. . . . .</i>	91
7.18	<i>Theoretical omnidirectional rank correlation over distance curves (left) and theoretical omnidirectional asymmetry over distance curves (right) compared with the empirical ones. . . . .</i>	91
7.19	<i>Statistical hypothesis test of rank correlation structure for the horizontal direction (left column) and the omnidirection (right column). . . . .</i>	93
7.20	<i>Statistical hypothesis test of asymmetry over distance structure for the horizontal direction (left column) and the omnidirection (right column). . . . .</i>	94
7.21	<i>Comparison of realizations from the three copula models. Top one is from the Gaussian model, the middle one is from the <math>v</math>-transformed normal copula and bottom one is from the maximum normal copula. . . . .</i>	95
7.22	<i>Rank correlation over distance curves of the realizations shown in the above figure. . . . .</i>	96
7.23	<i>80% statistical boundaries and the mean behaviour of Euler's characteristic curves calculated from 100 realizations of the three copula models. . . . .</i>	97
8.1	<i>Synthetical example for network expansion. Closed symbols are monitoring locations. Open symbols are candidate locations. Grid points are new estimation locations . . . . .</i>	104
8.2	<i>Maps of percentage of positive decisions resulted from the Gaussian copula (left) and the <math>v</math>-transformed normal copula (right) after adding the optimal candidate to the measurements (<math>p(\theta = \theta_0)=0.8, k_{00} = 0.0, k_{01} = -2.0, k_{10} = -1.0, k_{11} = 0.0</math>). . . . .</i>	105
8.3	<i>Maps of percentage of positive decisions resulted from the Gaussian copula (left) and the <math>v</math>-transformed normal copula (right) after adding the optimal candidate to the measurements (<math>p(\theta = \theta_0)=0.7, k_{00} = 0.0, k_{01} = -2.0, k_{10} = -1.0, k_{11} = 0.0</math>). . . . .</i>	106
8.4	<i>Maps of percentage of positive decisions resulted from the Gaussian copula (left) and the <math>v</math>-transformed normal copula (right) after adding the optimal candidate to the measurements (<math>p(\theta = \theta_0)=0.7, k_{00} = 0.0, k_{01} = -4.0, k_{10} = -1.0, k_{11} = 0.0</math>). . . . .</i>	107
8.5	<i>Old measurements and new candidate points in the sub-region of Baden-Württemberg in Germany. Different symbols represent different concentration levels for the old measurements and utility levels for the candidates. Small dots represent the interpolation grid. . . . .</i>	108
9.1	<i>Horizontal planes from the X-ray tomography of the bulk density of a soil column [13].</i>	112
9.2	<i>An example of more complicated non-monotonic transformation with transformation parameters to be estimated from datasets. . . . .</i>	113

# List of Tables

5.1	<i>Parameters to be estimated for the three theoretical copulas . . . . .</i>	40
5.2	<i>Parameterization results of one realization of a <math>v</math>-transformed normal copula compared with the pre-defined values used in the simulation. . . . .</i>	41
6.1	<i>Statistics of the five groundwater quality parameters . . . . .</i>	53
6.2	<i>Estimated model parameters for <math>v</math>-transformed normal copula . . . . .</i>	56
6.3	<i>Results of crossvalidation and split sampling for different interpolation methods . . . . .</i>	67
6.4	<i>Results of crossvalidation for the confidence intervals of the different interpolation methods . . . . .</i>	68
7.1	<i>Optimized parameters of a Gaussian copula from the saturated hydraulic conductivity measurements. . . . .</i>	87
7.2	<i>Optimized parameters of a <math>v</math>-transformed normal copula from the saturated hydraulic conductivity measurements. . . . .</i>	87
7.3	<i>Optimized parameters of a maximum normal copula from the saturated hydraulic conductivity measurements. . . . .</i>	88
8.1	<i>Utility matrix . . . . .</i>	101
8.2	<i>Results of optimization of synthetical study field with threshold <math>\beta = 0.8</math>, coefficients of utility function: <math>k_{00} = 0.0, k_{01} = -2.0, k_{10} = -1.0, k_{11} = 0.0</math>. . . . .</i>	105
9.1	<i>Illustration of binary decoding for assigning the signs of the components in the vector <math>\zeta_i^T</math>. . . . .</i>	117

# Abstract

In most geostatistical analysis, the spatial variability is solely described with a variogram or a covariance function. These measures have three major problems. First of all, they imply a Gaussianity assumption for the spatial dependence structure. This is, however, often violated in reality as manifested by the dataset investigations in this study. Second, they are sensitive to outliers and thus can be easily polluted by measurement anomalies. Third, they describe the spatial dependence as an integral over the whole marginal distribution of the parameter values. The change of the dependence strength with different quantiles of the parameter is not reflected by these measures, hence not considered in the subsequent interpolation or simulation procedures. But this aspect can be of vital importance for some prediction and design purposes. For example, to predict the flow or transport behavior in a heterogeneous subsurface, where cracks or connected paths, whose spatial continuity deviates largely from the mean trend, exist. Or to extend a monitoring network for non-compliance with environmental standards, where analysis of estimation uncertainty plays a central role and the difference in the uncertainties for different parameter values cannot be overlooked.

To overcome the above addressed problems, the concept of copulas is borrowed in this study as an alternative to the traditional geostatistical tools for spatial description and modeling. As a counterpart of the shortcomings of variogram/covariance function, the main advantages of using copula are also threefold. First of all, it captures the *pure* dependence among the random variables separately from their univariate distributions and thus the influence of measurement outliers or very skewed distributions vanishes. Second, since it describes the dependence as a full distribution instead of the mean behavior, the variation of dependence strength for different quantiles is revealed, which considerably improves the estimation and prediction quality. Last but not the least, non-Gaussian theoretical copulas which are suitable for spatial modeling can be developed so that the Gaussianity assumption is no more a must and the real dependence structure can be mimicked better.

In this study, methodology of using copulas for spatial modeling is established, including making the basic hypothesis, defining empirical copulas as a substitute for variogram/covariance, adopting and devising scale invariant measures for quantification of spatial dependence. Theoretical non-Gaussian copulas which are suitable for spatial modeling are derived and the model inference approach which is a combination of maximum likelihood and multiple-point statistics is proposed as well. The application of the methodology breaks down into three parts. The first part focuses on spatial interpolation, where the procedure of interpolation based on conditional copula is developed. An example of spatial interpolation of groundwater parameters in Baden-Württemberg (Germany) shows that

the copula based approach gives better cross validation results than the ordinary and indicator kriging methods. Validation of the confidence intervals estimated from conditional copulas indicates that they are more realistic than the estimation variances obtained from ordinary and indicator kriging. The second part deals with the topic of spatial simulations. In this part, simulation algorithms of generating realizations of multivariate non-Gaussian dependence are developed for both unconditional and conditional cases. Spatial analysis of the hydraulic conductivity measurements of Las Cruces Trench Site shows that the spatial dependence of this dataset exhibit clear anisotropic and non-Gaussian behavior. Statistical tests of the realizations from three copula models parameterized on this dataset, i.e., the Gaussian copula, the  $v$ -transformed normal copula and the maximum normal copula, also indicate that the Gaussian copula is most likely to be rejected, while the maximum normal copula is proved to be the most suitable one. The significance of multivariate dependence structure described by copulas to the flow and transport behavior is studied indirectly by investigating the topological/connectivity characteristics of the realizations from different copula models. In the third part, the approach of spatial modeling based on copulas is applied to facilitate observation network design where the estimated conditional copulas describing the probabilistic structure of the values at the unsampled locations are embedded into the utility function which acts as the objective function to be maximized in order to select the optimal location for new measurement. The application to expand the observation network of groundwater quality parameters in a sub-region of Baden-Württemberg demonstrates the potentialities of the methodology.

# Kurzfassung

In den meisten geostatistischen Analysen wird die räumliche Variabilität allein mit Hilfe eines Variogramms oder einer Kovarianzfunktion beschrieben. Diese Werte haben drei große Probleme: Erstens, sie nehmen implizit an, daß die räumliche Abhängigkeitsstruktur normalverteilt sei. Diese Annahme ist oft nicht gerechtfertigt, wie in den Analysen von Datensätzen in dieser Studie gezeigt wird. Zweitens sind diese Werte sensitiv gegenüber Ausreißern, weswegen sie leicht von Anomalien in den Messwerten beeinflusst werden. Drittens beschreiben diese Werte die räumliche Abhängigkeit als ein Integral über die gesamte Randverteilung der Parameterwerte. Eine Veränderung des Grades der Abhängigkeit für verschiedene Quantile des Parameters wird von diesen Werten nicht berücksichtigt, und auch nicht in darauf folgenden Interpolationen oder Simulationen. Dieser Aspekt kann jedoch wichtig für einige Vorhersagen oder für Design werden. Zum Beispiel um die Strömung oder das Transportverhalten in einem porösen Medium in dem zum Beispiel bevorzugte Fließwege existieren können, deren räumliche Verbundenheit signifikant von der mittleren Tendenz abweichen. Oder um ein Messnetz für die Überwachung von Umweltstandards zu erweitern, wo die Analyse von Schätzunsicherheit eine zentrale Rolle spielt und die Unterschiede in Unsicherheiten für verschiedene Parameterwerte nicht vernachlässigt werden können.

Um die oben angesprochenen Probleme zu adressieren, wurde das Konzept von Copulas in dieser Studie angewendet als Alternative zu traditionellen geostatistischen Werkzeugen für räumliche Beschreibung und Modellierung. Als Gegenpart zu den Mängeln der Variogram/Kovarianz-funktion gibt es hauptsächlich drei Vorteile bei der Benutzung von Copulas: Erstens erfassen Copulas die pure Abhängigkeit der Zufallsvariablen unabhängig von deren univariaten Verteilungen. Deshalb entschwindet der Einfluss von Ausreißern oder sehr schiefen Verteilungen. Zweitens, da die Copula die Abhängigkeit als eine komplette Verteilung beschreibt, anstatt einem mittleren Verhalten, wird der veränderliche Grad der Abhängigkeit für verschiedene Quantile offengelegt, was die Schätzung und Vorhersagequalität erheblich verbessert. Nicht zuletzt können theoretische nicht-Gauss Copulas entwickelt werden, die für räumliche Modellierung passend sind. Dadurch werden Gauss-Annahmen nicht mehr zwingend, und die wirkliche Zusammenhangstruktur besser abgebildet.

In dieser Studie wird die Methodik eingeführt, wie man Copulas für räumliches Modellieren benutzt, als Ersatz für Variogramm/Kovariogram. Dabei werden skalenunabhängige Maße zur Quantifizierung der räumlichen Abhängigkeit eingesetzt und eingeführt. Theoretische nicht-Gauss Copulas die für räumliche Modellierung adäquat sind werden abgeleitet und ein Ansatz zur Model-Inferenz der eine Kombination aus Maximum Likelihood und

multiple-point Statistik ist, wird vorgeschlagen. Die Anwendung der Methodik gliedert sich in drei Teile:

Der erste Teil behandelt räumliche Interpolation, wobei eine Prozedur der Interpolation, basierend auf bedingten Copulas entwickelt wird. Ein Beispiel der räumlichen Interpolation von Grundwasserparametern in Baden-Württemberg (Deutschland) zeigt, dass der Copula-basierte Ansatz bessere Kreuzvalidierungs-Ergebnisse liefert als herkömmliche- und Indikator-Kriging-Methoden. Die Validierung der Konfidenzintervalle die basierend auf konditionellen Copulas geschätzt wurden zeigt, dass sie realistischer sind als die Schätzvarianz basierend auf Ordinary Kriging und auf Indikator Kriging.

Der zweite Teil beinhaltet das Thema räumlicher Simulationen. In diesem Teil werden Simulations-Algorithmen entwickelt mit dem Ziel, Realisationen multivariater nicht-Gauss-Abhängigkeit zu generieren, sowohl für nicht-bedingte als auch für bedingte Simulation. Räumliche Analyse von Messungen hydraulischer Leitfähigkeit der Las Cruces Trench Site zeigt, dass die räumliche Abhängigkeitsstruktur deutlich anisotrop und nicht-gauss ist. Statistische Tests der Realisationen aus den drei Copula-Modellen die basierend auf diesem Datensatz parameterisiert wurden – der Gauss-Copula, der v-transformierten Copula und der maximalen normalen Copula- zeigen ebenfalls, dass die Gauss-Copula am wahrscheinlichsten abgelehnt wird, während bewiesen wird, dass die maximale normale Copula die passendste ist. Die Bedeutung der multivariaten Abhängigkeitsstruktur für Strömungs- und Transportverhalten, wie sie von Copulas beschrieben wird, wird beispielhaft indirekt dargelegt durch die Analyse der Topologie und Verbundenheit der Realisationen von verschiedenen Copulamodellen.

Im dritten Abschnitt wird der Ansatz der räumlichen Modellierung basierend auf Copulas angewandt auf Messnetzplanung wobei die geschätzten bedingten Copulas die Wahrscheinlichkeitsstruktur der Werte an nicht-beprobten Orten in die zu maximierende Nutzenfunktion eingebettet wird um den optimalen Ort für eine Messung zu wählen. Die Anwendung auf eine Vergrößerung des Grundwasserqualität-Messnetzes in einer Teilregion Baden-Württembergs verdeutlicht das Potential dieser Methode.



# 1 General Introduction

All natural physical processes are subject to variability. It is difficult and expensive to collect sufficient field observations which are needed by an environmental engineer or hydrogeologist to answer an engineering or scientific question. As a consequence, one must make the best use of available data to estimate the needed information. This requires the experts who analyze the data to use their knowledge and understanding of the process to fill in the gaps. However, the complexity of most natural processes precludes a complete understanding and thus a unique or precise answer. Due to this fact, statistical estimation methods come into play to complement process understanding and to support in making rational decisions. When spatial distributed parameters are under concern, one branch of probabilistic approach called 'Geostatistics' is applied to solve the estimation problem.

Geostatistics offers a collection of deterministic and statistical tools aimed at modeling spatial variability [24]. It was first used by the mining industry, as high costs of drillings made the spatial interpolation of the available data an imperative technique. Its origin dates back to the early fifties when in South-Africa Danie. G Krige and his colleagues started to apply statistical techniques to ore reserve estimation [7]. In the sixties, Krige's empirical work was formalised by the french mathematician Georges Matheron. Since then it is popularized in diverse fields ranging from mining and petroleum industry to hydrology, meteorology, oceanography, environmental control, landscape ecology and agriculture. The contribution of this thesis to the geostatistical theory is the development of a novel spatial analysis tool using the concept of *Copulas*.

## 1.1 Background and Motivations

Geostatistics focuses on the study of natural phenomena that fluctuate in space and/or time [69]. Its axis is the *theory of regionalized variables* [54]. *When a phenomenon spreads in space and exhibits a certain spatial structure, we shall say that it is regionalized* [64]. For example, the altitude in a horizontal space, the heavy metal concentration in the groundwater, the gold grade in a deposit, etc. Due to the inherent randomness and consequent unpredictability of the natural phenomena, the quantity of the natural parameter measured at a certain location is usually considered as an outcome of a *random variable* [55], abbreviated as RV hereafter. Then for each location  $s$  in the studied domain, a corresponding random variable  $Z(s)$  is assigned. The collection of the RVs associated to all the points within the domain under study is called a *random function*, or shortly RF. A *regionalized variable* is the realization of a RF [7]. This means that the set of measured values of the parameter under interest is

interpreted as one realization of the RF. This interpretation expresses both the random and the structured aspects of the spatial parameter [52].

A RF is described by its multidimensional distribution function. For any set of points  $\mathbf{s}_1, \dots, \mathbf{s}_n$  in the studied domain, a cumulative distribution function  $F_{\mathbf{s}_1, \dots, \mathbf{s}_n}$  is assigned. Then the joint probability of the set of possible values  $w_1, \dots, w_n$  at those points can be found:

$$P\left(Z(\mathbf{s}_1) < w_1, \dots, Z(\mathbf{s}_n) < w_n\right) = F_{\mathbf{s}_1, \dots, \mathbf{s}_n}(w_1, \dots, w_n) \quad (1.1)$$

This is useful in the sense that the value at certain unsampled location can be evaluated by the conditional probability at this location conditioned on the neighbouring measurements. Unfortunately, the direct assessment of the distribution function described in Equation 1.1 is infeasible. Because there are usually only one observation available at each location, which means one realization of the RF is known. Even in the case that repeated measurements are made for the parameter under study, there are not enough measurements to assess the above distribution function [7]. The strategy to tackle this problem is to trade the unavailable replication at location  $\mathbf{s}$  for another replication available somewhere else in space [24]. The formal term of this trade is called the *hypothesis of stationarity*. Under this assumption, the RV is invariant under any translation of the coordinates:

$$F(\mathbf{s}_1, \dots, \mathbf{s}_n; w_1, \dots, w_n) = F(\mathbf{s}_1 + \mathbf{h}, \dots, \mathbf{s}_n + \mathbf{h}; w_1, \dots, w_n) \quad (1.2)$$

where  $\mathbf{h}$  denotes the translation vector. Invariance of the multivariate distribution function entails invariance of any lower order distribution functions [24]. This assumption is key to the traditional geostatistics and also essential to the new methodology developed in this thesis. It allows the inference of the distribution function. For example the univariate distribution of the RV at a location or the bivariate distribution of the RVs at two locations. The latter one plays a central role in describing the spatial dependence in traditional geostatistics. And the application of hypothesis of stationarity to the univariate and bivariate case in the expected sense has a special name: *second order stationarity*:

$$E[Z(\mathbf{s})] = m \quad (1.3)$$

and

$$E\left[(Z(\mathbf{s} + \mathbf{h}) - m)(Z(\mathbf{s}) - m)\right] = C(\mathbf{h}) \quad (1.4)$$

In practice, the assumption of the *second order stationarity* is relaxed to the so called *intrinsic hypothesis*, where the first condition remains the same, while the second one is modified to:

$$\frac{1}{2}Var\left[Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})\right] = \frac{1}{2}E\left[(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}))^2\right] = \gamma(\mathbf{h}) \quad (1.5)$$

where the function  $\gamma(\mathbf{h})$  is called *semivariogram* or briefly *variogram* which describes the variance of the increment corresponding to two locations as a function of the separation vector

h between the two locations. In traditional geostatistics, variogram acts as the main description and modeling tool for the spatial dependence structure which channels into the various procedures of evaluation and prediction (e.g., interpolation and simulation). However, several drawbacks exist for the variogram as the description of spatial dependence which are listed below:

1. It is limited to two-point statistics.
2. It only takes the configuration of the data points into consideration but not the state of the nature of the parameter, and thus describes the spatial dependence as an integral over the whole marginal distribution of the parameter values [8].
3. It is susceptible to measurement anomalies and skewed distributions.
4. It entails the assumption of Gaussian dependence which in many cases cannot be fulfilled in nature.

These restrictions sometimes preclude an appropriate description of the underlying spatial dependence, hence lead to estimations and predictions that are far away from reality. For instance, the prediction of flow and transport fate in heterogeneous porous media is considerably influenced by the field characterization of the relevant spatial parameters. The spatial continuity of the extreme hydraulic conductivity values is crucial to this problem. As in many heterogeneous fields, very irregular spatial patterns such as channels, fractures, aquitards are present. The spatial continuity of those geological formations can deviate greatly from the mean trend. Under such circumstances, the description of the dependence by the variogram is not sufficient and the implicit Gaussianity assumption is not appropriate.

Besides the local or global estimation of the studied parameter from the available observations, another problem that a geologist or a project engineer are faced with is the assessment of the confidence that can be given to these estimations. In traditional geostatistics, this confidence is expressed with the Kriging variance, which is merely a function of the observation density and the imposed variogram model [11]. However, the estimation uncertainty can vary with varying quantiles of the marginal distribution of the parameter values. This aspect can be of prime importance for monitoring network design, where the impact of additional measurements on error reduction is anticipated.

All the problems of the traditional geostatistical approach posed by the above discussions give the genesis of this thesis. Introducing a new contribution to the geostatistical approach, copulas offer an interesting and preferable possibility to describe and model the dependence structures for multivariate distributions. Copulas are standardized multivariate distributions with univariate marginals being uniformly distributed. Copula describes the dependence structure among random variables separately from their univariate distributions and thus captures the *pure expression* of the dependence without the influence of the marginal distribution [8]. In other words, copula represents the scale-invariant dependence and therefore unlike the variogram values it is not sensitive to the outliers in the measurements. Another

benefit of this property is that the frequently faced dilemma, i.e. whether to transform the data (e.g., taking the natural logarithms or applying a normal score transformation) or not is out of the question. In addition, as the definition of copula is a multivariate distribution, it is able to express whether the corresponding dependence is different for different quantiles of the parameter values or not. As mentioned before, this aspect is not appropriately considered in the traditional geostatistic approach but has great significance for estimation and prediction of some natural phenomena.

Most applications of copulas are in the financial sector (e.g., *Embrechts et al.* [2001] [25], *Rachev et al.* [2005] [71] and *Cherubini et al.* [2004] [17]) where dependence of extreme values plays an important role [11]. There are a few hydrological applications, most of which are also related to the analysis of extremes (e.g., *Salvadori et al.* [2007] [76]; *Favre et al.* [2004] [30]; *Poulin et al.* [2007] [70]). In *Bárdossy* [2006] [8], the concept of copulas is first introduced into the spatial context, where the empirical variogram is replaced by the empirical copulas to investigate the spatial dependence structure of datasets. Translation invariant spatial copula functions are developed as the random function used for interpolation purpose. The methodology presented in this paper acts as a precursor to this PhD study.

## 1.2 Scope and Organization of the Thesis

Based on the work of *Bárdossy* [2006] [8], the scope of this study is to establish systematically a copula based geostatistical method aiming to improve the spatial analysis, spatial interpolation, spatial simulation and decision making procedures. This thesis is organized as follows.

Before beginning any theoretical development, the mathematical definitions of copula and its main properties are briefly reviewed in Chapter 2. In Chapter 3, the methodology of using copulas to investigate spatial dependence is described, where the basic hypothesis are listed and the calculation of empirical copulas from observations is explained. In addition, alternative dependence measures, i.e., the rank correlation, the measure of asymmetry and the Euler characteristics are introduced, among which, the second one is a newly developed measure. The advantages of using the empirical copulas and the selected measures compared with the empirical variogram or covariance structure are demonstrated as well in this chapter. In Chapter 4, three theoretical copula models including the Gaussian copula and two newly developed non-Gaussian copulas which are suitable for spatial dependence modeling are introduced. The properties of each copula models are examined. In the next step, the theoretical copulas have to be parameterized so that they can be used to model the spatial dependence of the real world. To this end, Chapter 5 presents the model inference method which combines the method of maximum likelihood and multiple-point statistics.

The remainder of the thesis is devoted to the applications of the methodology to three different purposes: spatial interpolation, spatial simulation and observation network design. Chapter 6 deals with the interpolation problem. In this chapter, the interpolation scheme based on the estimation of conditional copula at the unsampled locations conditioned on the available measurements are developed and described in details. This also provides the

fundamentals of the conditional simulation algorithm and the utility function used in the network design which are discussed in Chapter 7 and Chapter 8, respectively. The application of the interpolation approach to a large scale groundwater quality observation network in Baden-Württemberg, Germany is demonstrated. Chapter 7 presents the stochastic simulation algorithms on simulation of multivariate non-Gaussian realizations in both the unconditional and conditional cases. In order to illuminate the effect of spatial dependence, aside from the univariate distribution and covariance structure, on the flow and transport behaviour, comparison should be done between realizations of different dependence models which share the same or similar histogram and covariance function. The third section of Chapter 7 provides an adhoc procedure to generate realizations sharing the same aforementioned statistics but differing in their multivariate dependence structures. In the fourth section, the three theoretical copulas are parameterized based on the dataset of the hydraulic conductivities in Las Cruces Trench Site and the corresponding spatial simulations are carried out aiming at the comparison of spatial structures generated by different copula models. Statistical tests are conducted to check the goodness-of-fit of the models. Significance of difference in the spatial dependence on the flow and transport behaviour is investigated indirectly by studying the Euler characteristics, which quantifies the connectivity property of porous media, of the realizations from different copula models. The application of spatial copulas can be further extended to facilitate decision making process. In Chapter 8, an approach for environmental network design using the framework of spatial modeling with copulas are presented followed by an application on the expansion of the observation network of groundwater quality parameters in Baden-Württemberg, Germany, showing the potentialities of the methodology.

At the end of this thesis, a summary of the study and an outlook of the future steps are outlined.

## 2 Introduction to Copulas

This chapter does not provide a rigorous mathematical exposition of the concept of copula. Instead, a few basic definitions and theorems that are used in this thesis work are recalled. More details about the concept of copula can be found in Joe [1997] [40] and Nelsen [1999] [66].

Copulas are standardized multivariate distributions whose support is the  $n$  dimensional unit hypercube,

$$C : [0, 1]^n \rightarrow [0, 1] \quad (2.1)$$

with all one dimensional marginals being uniformly distributed on  $[0,1]$ :

$$C(\mathbf{u}^{(i)}) = u_i \quad \text{if} \quad \mathbf{u}^{(i)} = (1, \dots, 1, u_i, 1, \dots, 1) \quad (2.2)$$

The copula becomes zero if any of the arguments is zero:

$$C(\mathbf{u}) = 0 \quad \text{if} \quad \mathbf{u} = (u_1, \dots, 0, \dots, u_n) \quad (2.3)$$

For every  $n$  dimensional hypercube in the unit hypercube the corresponding probability has to be non negative:

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

if  $0 \leq u_i \leq u_i + \Delta_i \leq 1$  and  $i = \sum_{k=0}^{n-1} j_k 2^k, j_k = 0, 1$

For the bivariate case, the above properties are summarized as: (1)  $C(u, 1) = C(1, v) = u$ ; (2)  $C(u, 0) = C(0, v) = 0$ ; (3)  $C(u_2, v_2) + C(u_1, v_1) - C(u_2, v_1) - C(u_1, v_2) \geq 0$  if  $u_2 \geq u_1$  and  $v_2 \geq v_1$ . The third property ensures that the probability assigned to any rectangle in the unit square is nonnegative.

*Sklar* [80] proved that each multivariate distribution  $F(x_1, \dots, x_n)$  can be represented together with its univariate marginal distributions:

$$F(x_1, \dots, x_n) = C(F_{x_1}(x_1), \dots, F_{x_n}(x_n)) \quad (2.4)$$

where  $F_{x_i}(x)$  represents the  $i$ -th one dimensional marginal distribution of the multivariate distribution. If the distribution is continuous then the copula  $C$  in Equation (2.4) is unique. If the random variables are independent of each other, the copula of the joint distribution function is:

$$C(u_1, u_2, \dots, u_n) = u_1 \cdot u_2 \dots u_n = \prod^n$$

where  $u_i = F(x_i)$ ,  $\prod^n$  denotes the  $n$ -dimensional independent copula.

In case of full dependence, the copula becomes:

$$C(u_1, u_2, \dots, u_n) = \min(u_1, u_2, \dots, u_n) = M^n(\mathbf{u})$$

where  $M^n(\mathbf{u})$  denotes the *Fréchet-Hoeffding upper bound* [66].

The boundaries of any  $n$ -dimensional copula, for every vector  $\mathbf{u} = (u_1, \dots, u_n)$  are:

$$W^n(\mathbf{u}) \leq C(\mathbf{u}) \leq M^n(\mathbf{u}) \quad (2.5)$$

where  $W^n(\mathbf{u}) = \max(u_1 + u_2 + \dots + u_n - n + 1, 0)$  and denotes the *Fréchet-Hoeffding lower bound* [66].

It is clear that copulas are functions that 'couple' multivariate distribution functions to their one dimensional marginal distribution functions [66]. Copulas also provides a way of studying 'scale-invariant' measure of dependence. In other words, under strictly monotone transformations of the random variables, the copula is invariant while the margins may be changed at will. It is precisely the copula which captures the properties of the joint distribution which are invariant under almost surely increasing transformations (*Schweizer and Wolff* [1981] [78]). Thus the frequently applied data transformation (for example taking the natural logarithms or normal score transformation) has no influences on the copula. Figure 2.1 shows the scatter plots of the joint probability densities of two pairs of random variables and the corresponding copula densities. We can see that different joint distributions can have the same copula. It is the marginal distributions that disturb the analysis of dependence structure.

On the other hand, copulas can be constructed from joint distribution functions, as described by ([66]):

$$C(\mathbf{u}) = C(u_1, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)) \quad (2.6)$$

Copulas offers an interesting way of simulating multivariate distributions. This topic is explored in Chapter 7.

If the copula  $C$  is continuous, the copula density can be written as:

$$c(u_1, \dots, u_n) = \frac{\partial^n C(u_1, \dots, u_n)}{\partial u_1 \dots \partial u_n} \quad (2.7)$$

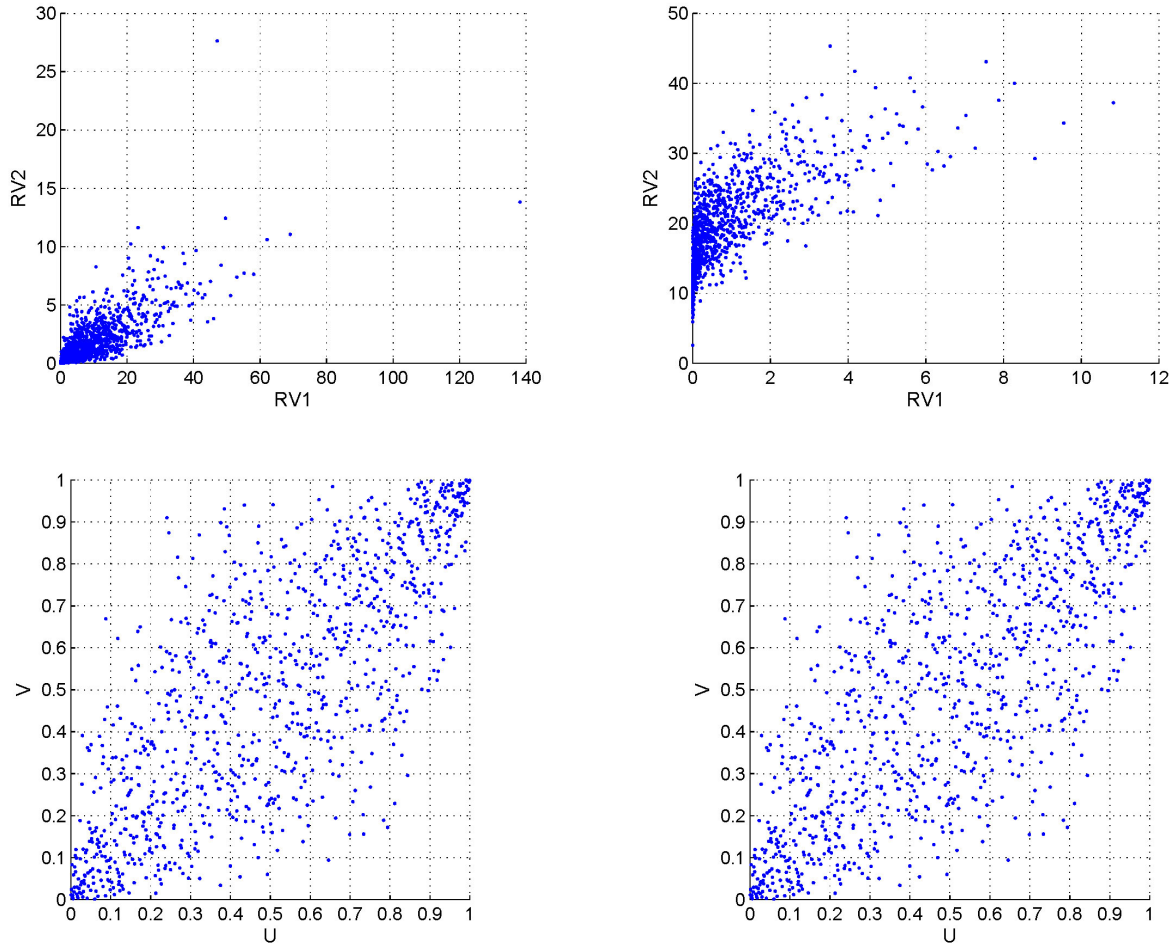


Figure 2.1: Scatter plots of joint probability densities (upper two) of two pairs of random variables and the corresponding copula densities (lower two)

For the bivariate case, the copula density should fulfill:

$$c(u, v) \geq 0 \quad (2.8)$$

$$\int_0^1 c(u, v) dv = c(u) = 1 \quad (2.9)$$

$$\int_0^1 c(u, v) du = c(v) = 1 \quad (2.10)$$

It can be proved that the conditional copula is:

$$C(u|U_1 = u_1, \dots, U_n = u_n) = \frac{\partial^n C(u, u_1, \dots, u_n)}{\partial u_1 \dots \partial u_n} \cdot \frac{1}{c(u_1, \dots, u_n)} \quad (2.11)$$

As mentioned before, copula describes the dependence as a full distribution, therefore it can express whether the corresponding dependence is different for different quantiles of



the variables. For example, high quantiles might exhibit stronger dependence than the low quantiles.

A bivariate copula expresses a symmetrical dependence if:

$$C(u, v) = C(1 - u, 1 - v) - 1 + u + v \quad (2.12)$$

which means that the copula density is symmetrical with respect to the secondary diagonal  $u = 1 - v$  of the unit square and the copula density  $c$  fulfills:

$$c(u, v) = c(1 - u, 1 - v) \quad (2.13)$$

A typical family of copulas which describes a symmetrical dependence is the Gaussian copula, which is often used in hydrology for example in regressions, or if the normal score transformation is applied [8]. Figure 2.2 shows the Gaussian bivariate copula (on the left hand side) and the corresponding bivariate copula density with correlation equal to 0.85. The horizontal and vertical axis stand for the cumulative distribution values of the two random variables, respectively. The scale of the left chart stands for the bivariate copula values. As described in Equation (2.2) and (2.3), the copula values range between 0 and 1. The minimum and maximum values, i.e.,  $-0.5$  and  $1.05$ , indicated in the legend are due to the fact that the values are calculated on a 100 by 100 grid of the unit square and any values in between are interpolated/extrapolated by the graphic software, which induces the exceedance of the boundaries. The same applies for the negative minimum value shown in the legend of the chart on the right hand side. The scale of the chart on the right hand side stands for the bivariate copula density, which fulfils the properties described in Equation (2.10).

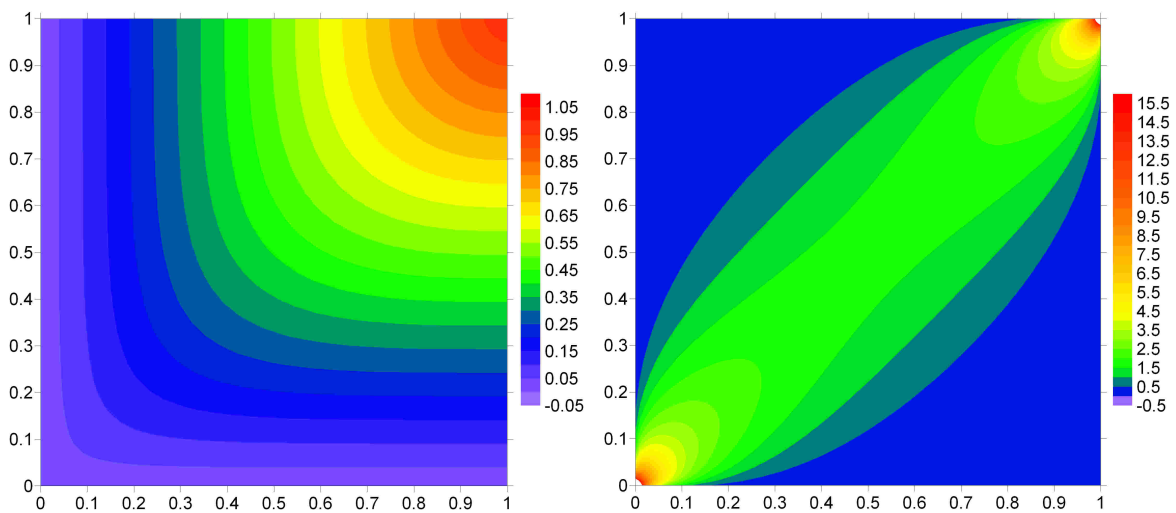


Figure 2.2: *Bivariate Gaussian copula (left) and corresponding copula density with correlation equal to 0.85.*

In order to apply the concept of copulas to investigate spatial dependence, some assumptions have to be made. Based on these assumptions, spatial copulas can be defined. The corresponding scale-invariant measures can be derived from the spatial copulas. The next chapter is composed to address these points.

## 3 Methodology of using Copulas to investigate Spatial Dependence

In the spatial statistical framework, copulas can be used to describe the joint multivariate distribution corresponding to the variables associated with the points of the studied domain (Bardossy [2006] [8]). For this purpose, several assumptions has to be made for spatial copulas. This chapter begins with the basic hypothesis needed in this context. As a visualization tool to reveal the dependence structure as a function of both data configuration and data values of the natural variables, the concept of empirical copulas is introduced in Section 3.2 of this chapter. In order to quantify the dependence structure, the copula related 'scale-invariant' measures are preferred. In this study, the *Spearman's rho* [66] [56] [60] and the measure of asymmetry are used and discussed in details in Section 3.3. In addition, the concept of Euler characteristic is applied to investigate the flow and transport related properties of the spatial field. It is also introduced in Section 3.3.

### 3.1 Basic Hypothesis of Spatial Copulas

As mentioned in Chapter 1 the assumption of a stationary random function is key to the traditional geostatistics. Since in the geostatistical framework, the spatial observation of a natural parameter of interest is assumed to be a realization of a *random function* RF, which is defined as a set of usually dependent random variables  $Z(\mathbf{s})$ , one for each location in the studied domain  $\mathcal{A}$ ,  $\{Z(\mathbf{s}), \forall \mathbf{s} \in \mathcal{A}\}$  [35]. The spatial observation is interpreted as a *regionalized variable* which is one realization of the prior RF. Due to the fact that usually only a single realization of the RF is observed at a limited number of points, the direct assessment of the multivariate RF is impossible, which entail the stationary hypothesis on the RF (see Equation 1.2).

Similarly for the new methodology, the basic hypothesis is to assume a stationary/translation invariant spatial copula:

*For each set of locations  $\{\mathbf{s}_1, \dots, \mathbf{s}_k\} \subset \mathcal{A}$ , and for each set of possible values  $w_1, \dots, w_k$ , the spatial copula  $\mathbf{C}_{\mathbf{S}}$  remains the same if all the points in this set is shifted by the same vector  $\mathbf{h}$ :*

$$\begin{aligned} \mathbf{C}_{\mathbf{S}} &= P(F_Z(Z(\mathbf{s}_1)) < u_1, \dots, F_Z(Z(\mathbf{s}_k)) < u_k) \\ &= P(F_Z(Z(\mathbf{s}_1 + \mathbf{h})) < u_1, \dots, F_Z(Z(\mathbf{s}_k) + \mathbf{h}) < u_k) \\ &= C(u_1, \dots, u_k) \end{aligned} \tag{3.1}$$

where  $F_Z$  denotes the marginal distribution of the random variable which is assumed to be the same at all locations in the studied domain, and  $u_i = F_Z^{-1}(w_i)$  is the probability distribution value corresponding to a certain value  $w_i$  of the random variable at the  $i$ -th location in the set  $\{\mathbf{s}_1, \dots, \mathbf{s}_k\}$ . The invariance of the multivariate copula entails the invariance of any lower order marginals. This implies that for any two selected points at locations  $\mathbf{s}$  and  $\mathbf{s} + \mathbf{h}$  in the investigated domain, the bivariate spatial copula corresponding to the random variables associated to these points is a function of the vector  $\mathbf{h}$  separating the two points and independent of the locations of the points:

$$\begin{aligned} \mathbf{C}_{\mathbf{S}}(\mathbf{h}, u, v) &= P(F_Z(Z(\mathbf{s})) < u, F_Z(Z(\mathbf{s} + \mathbf{h})) < v) \\ &= C(F_Z(Z(\mathbf{s})), F_Z(Z(\mathbf{s} + \mathbf{h}))) \end{aligned} \quad (3.2)$$

The assumption about the translation invariant bivariate marginal of the spatial copulas enables the assessment of empirical bivariate spatial copulas from the observations, which offers an insight into the spatial dependence structure of the natural variable of interest and is introduced in the following section.

### 3.2 Empirical Bivariate Spatial Copulas

Empirical copulas can be used to explore the spatial dependence of a dataset. Based on the stationarity assumption, a copula can be assigned to a given configuration of data points. The simplest and visible case is the construction of bivariate empirical copulas. Higher dimensional empirical copulas are difficult to visualize and require a large number of observations. However for the case of training images considered in multipoint geostatistical methods ([45]) higher dimensional empirical copulas can also be assessed. [11]

According to *Bardossy* [2006] [8], the estimation of the empirical bivariate copulas from observations  $z(\mathbf{s}_1), \dots, z(\mathbf{s}_n)$  can be done with the following procedure:

1. Calculate the empirical marginal distribution function  $F_Z(z)$  from the observations.
2. For any given vector  $\mathbf{h}$ , calculate the set of pairs  $S(\mathbf{h})$  which consists of the empirical distribution function values of the interested variables at locations separated by the vector  $\mathbf{h}$ :

$$S(\mathbf{h}) = \left\{ \left( F_Z(z(\mathbf{s}_i)), F_Z(z(\mathbf{s}_j)) \right) \mid |\mathbf{s}_i - \mathbf{s}_j| \approx \mathbf{h} \right\} \quad (3.3)$$

3. Divide the unit square into a regular  $k \times k$  grid. Since  $S(\mathbf{h})$  is a set of points in the unit square, the empirical copula density of a given vector  $\mathbf{h}$  can be constructed by calculating the empirical frequencies on the grid by:

$$\begin{aligned} q_{i,j} &= c^* \left( \frac{2i-1}{2k}, \frac{2j-1}{2k} \right) \\ &= \frac{k^2}{|S(\mathbf{h})|} \cdot \left| (u, v) \in S(\mathbf{h}); \frac{i-1}{k} < u < \frac{i}{k} \quad \text{and} \quad \frac{j-1}{k} < v < \frac{j}{k} \right| \end{aligned} \quad (3.4)$$

where  $|S(\mathbf{h})|$  stands for the total number of pairs of data points which are separated by  $\mathbf{h}$ . The theory underpins Equation (3.4) is that the bivariate density centered at a cell in the  $k \times k$  grid approximates to the ratio between the probability whose support is the cell and the area of the cell.

4. If the set  $S(\mathbf{h})$  contains sufficient number of pairs, the bivariate density can be approximated using kernel smoothing. Then the corresponding contour plot can be obtained for visualization.

Figure (3.1) illustrates the procedure of calculating the empirical copulas from a certain dataset. The left chart shows the data domain with black dots indicating the sampling locations. The double-arrow-ended lines connect the points which are separated by the required vector  $\mathbf{h}$ . The values listed next to two of the sampling points are the empirical distribution values of the measurements at the two locations. The right chart is the scattergram of the empirical copula density corresponding to the separation vector  $\mathbf{h}$ . It is worth mentioning that the scattergram of  $S(\mathbf{h})$  or contour plot of the empirical copula density is symmetrical with respect to the major axis  $u = v$  of the unit square, since by definition, if  $(u, v) \in S(\mathbf{h})$ , then  $(v, u) \in S(\mathbf{h})$ . In other words, if a pair of points is selected for a given  $\mathbf{h}$ , it is counted for both cells into which the pairs of the distribution values, i.e.  $(u, v)$  and  $(v, u)$ , of the two points fall on the  $k \times k$  grid of the unit square.

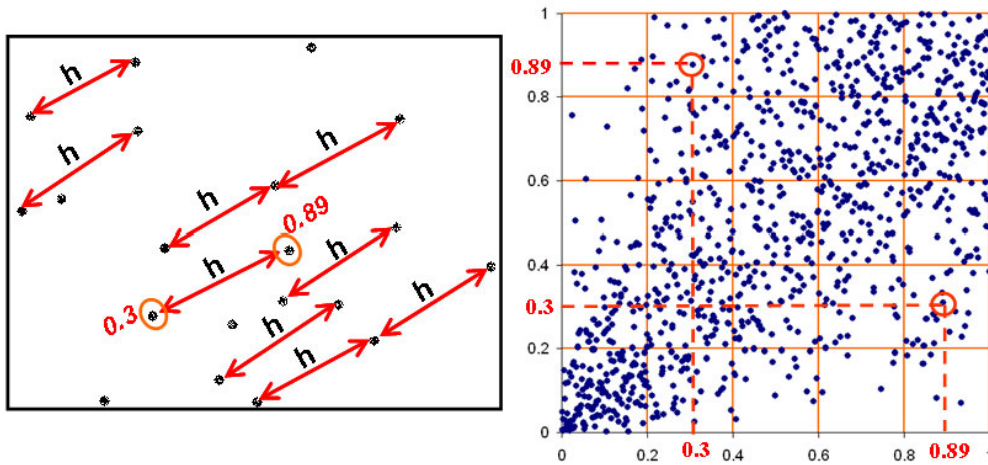


Figure 3.1: Procedure of calculating the empirical copula from a dataset.

Alike the variogram, for each  $\mathbf{h}$ , an empirical copula of the dataset can be calculated. The difference is that a full distribution is obtained for each  $\mathbf{h}$  instead of a single value. Therefore the empirical copula is more informative, which reveals how the dependence strength varies with different quantiles of the data values. This difference might be of vital importance in interpolation or simulation procedure for decision making purposes where extremes play a crucial role and the dependence of which differs prominently from the central behavior. Figure 3.2 shows the empirical copulas of chloride concentration in the groundwater of Baden-Württemberg for the separation length of 3 km, 6 km and 9 km from left to right.

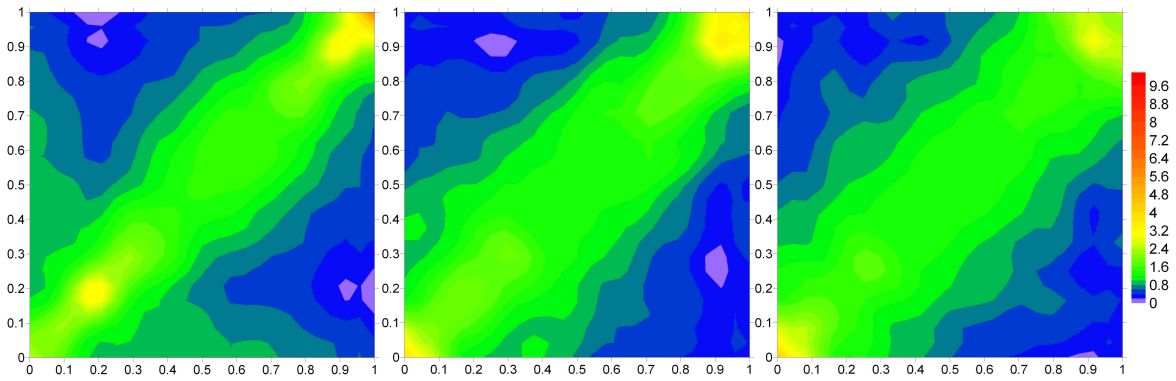


Figure 3.2: Empirical copulas of chloride concentration for the separation lengths of 3 km (left), 6 km (middle) and 9 km (right).

It can be inferred from Figure 3.2 that it is very likely to find a point with a cumulative distribution value of 0.2 at the location which is separated by 3 km from another location with the same cumulative distribution value. Because there is a peak in the density plot for the separation length of 3 km. But it is very unlikely to find the other point with a cumulative distribution value of 0.9 which is 3 km away from the data point with a cumulative distribution value of 0.2, because the density of this combination is very low. In addition, as the separation length increases, the shape of the empirical copula becomes less and less specific, which indicates the decaying of the dependence as the points are getting far away apart.

Empirical copulas enable us to visualize the dependence structure and gives us an insight into how the dependence varies with different data values and with increasing separation lengths. However, in order to get a 'crisp' description of dependence, one still needs measures which are discussed in the following section.

### 3.3 Measure of Dependence

As mentioned before that it is the copula which captures the essential dependence without the influence of the marginal distributions, the 'scale-invariant' measures of dependence can be expressed in terms of the copula of the random variables. The most widely known scale-invariant measures of dependence or association are Kendall's tau [56] [42] [61] and Spearman's rho [56] [60], both of which measure a form of dependence known as concordance. Since in this work, the latter one is applied, only the Spearman's rho is discussed in this section. A newly developed measure called 'measure of asymmetry' which measures the degree of departure of a bivariate dependence from a symmetrical dependence described in Equation (2.12) and (2.13) is introduced. In addition, the Euler characteristic is adopted in this study to investigate the connectivity property of porous media and the concept of which is reviewed at the end of this section.

### 3.3.1 Spearman's rho

The Spearman's rho measures the concordance of a pair of random variables. *Informally, a pair of random variables are concordant if large values of one tend to be associated with large values of the other and small values of one with small values of the other (Nelsen [1999] [66])*. If the relationship is inversed, then the pair of random variables are called *discordant*. The population version of Spearman's rho is defined to be proportional to the probability of concordance minus the probability of discordance for two vectors  $(X_1, Y_1)$  and  $(X_2, Y_3)$ , which have the same margins. The former vector has distribution function  $H$ , while the components of the latter one are independent:

$$\rho_{X,Y} = 3(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0]) \quad (3.5)$$

It can be proved that the Spearman's rho of two continuous random variables  $X$  and  $Y$  can be calculated with their copula  $C$  as following [66]:

$$\begin{aligned} \rho_{X,Y} &= 12 \int \int_{\mathbf{I}^2} uv dC(u, v) - 3 \\ &= 12 \int \int_{\mathbf{I}^2} C(u, v) dudv - 3 \end{aligned} \quad (3.6)$$

The sample version of Spearman's rho corresponds to the rank correlation coefficient. Suppose  $x$  and  $y$  are observations from two random variables  $X$  and  $Y$  with individual distribution functions  $F$  and  $G$ . Then the ranks of  $x$  and  $y$  are given by  $u = F(x)$  and  $v = G(y)$ . Note that  $u$  and  $v$  are observations from the uniform random variable  $U = F(X)$  and  $V = G(Y)$  both on  $[0, 1]$ , whose joint distribution function is described by their copula  $C$ . Because the expectations of  $U$  and  $V$  each are  $1/2$  and variances are  $1/12$ ,  $\rho_{X,Y}$  can expressed as:

$$\begin{aligned} \rho_{X,Y} &= 12 \int \int_{\mathbf{I}^2} uv dC(u, v) - 3 = 12E(U, V) - 3 \\ &= \frac{E(UV) - 1/4}{1/12} = \frac{E(UV) - E(U)E(V)}{\sqrt{Var(U)}\sqrt{Var(V)}} \end{aligned} \quad (3.7)$$

As a consequence, Spearman's rho for a pair of random variables is identical to the rank correlation. Like the variogram, for any separation vector  $\mathbf{h}$ , a rank correlation can be calculated. The rank correlation describes the strength of dependence independently of the marginal distribution, therefore it is a better alternative to variogram. Figure 3.3 shows the variogram (left) of chloride concentration in the groundwater of Baden-Württemberg, Germany as well as the rank correlation over distance curve (right). It shows that the rank correlation indicates a clear spatial dependence while the variogram does not show any structure. The main reason is that the chloride dataset has a highly skewed distribution which destroyed the variogram structure, but has no influence on the copulas and rank correlations.

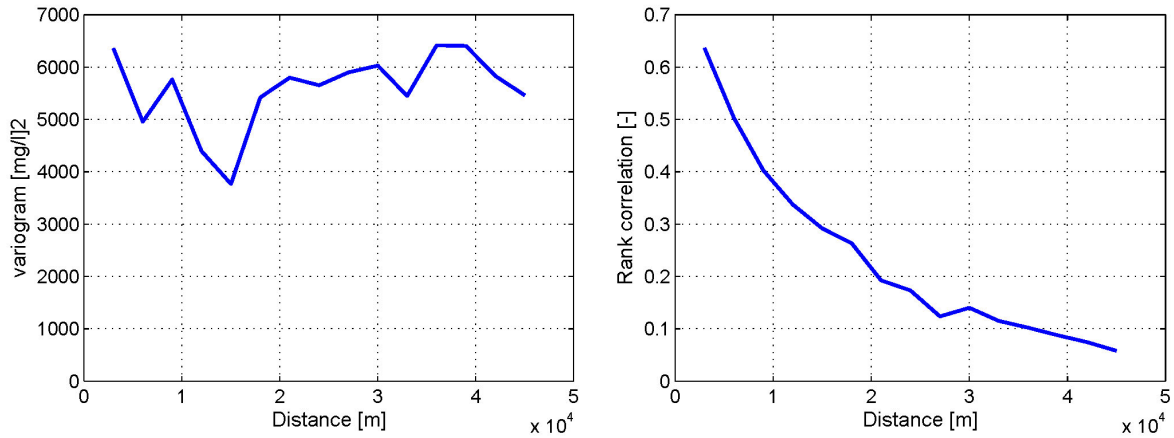


Figure 3.3: Variogram (left) and rank correlation over distance curve (right) of chloride.

### 3.3.2 Measure of asymmetry

The traditional geostatistics relies on two-point statistics, which requires the assumption of multi-Gaussian dependence of the spatial random function. As mentioned before, this assumption is often violated in reality. *Bardossy* [2006] [8], *Bardossy and Li* [2008] [11] show through the study of empirical copulas that the spatial dependence of the five groundwater quality parameters in Baden-Württemberg exhibit clear non-Gaussian behaviour. *Zinn and Harvey* [2003] [95] also pointed out that some conductivity fields which share the same histogram and covariance structure lead to very different flow and transport properties. This implies that some spatial conductivity fields can not be characterized by the conventional geostatistics which is only up to the second moment statistics and is only sufficient for the characterization of fields with Gaussian dependence.

To attack this problem, a new measure which is called measure of asymmetry of the underlying bivariate copula is invented. In *Bardossy* [2006] [8], the asymmetry is defined as the ratio of the joint probability of not exceeding a quantile  $u$  and exceeding the quantile  $1 - u$ :

$$A(u) = \frac{2u - 1 + C(1 - u, 1 - u)}{C(u, u)} \quad (3.8)$$

This ratio shows whether the high values have a stronger dependence than the low values ( $A(u) > 1$ ) or the opposite ( $A(u) < 1$ ), or if the dependence is the same ( $A(u) \approx 1$ ) (*Bardossy* [2006] [8]). The Gaussian copula is fully symmetrical, therefore  $A(u) = 1$ , which means the high quantiles have the same dependence as the mirrored low quantiles with respect to the median. This measure can be very important for spatial statistics where the locations of variable values that exceed certain thresholds are under interest. Using this measure, one can find out whether the high extremes behave in the same way as the low extremes or not. It has been used in the investigation of the five groundwater quality parameters in Baden-Württemberg. Like the spatial copulas, the asymmetry measure is independent of the locations  $s$ , but is a function of the separation vector  $\mathbf{h}$ .



Figure 3.4 shows the asymmetry over distance curves for the chloride concentration. The ratio calculated in Equation (3.8) is between the upper 10% and the lower 10% copulas.

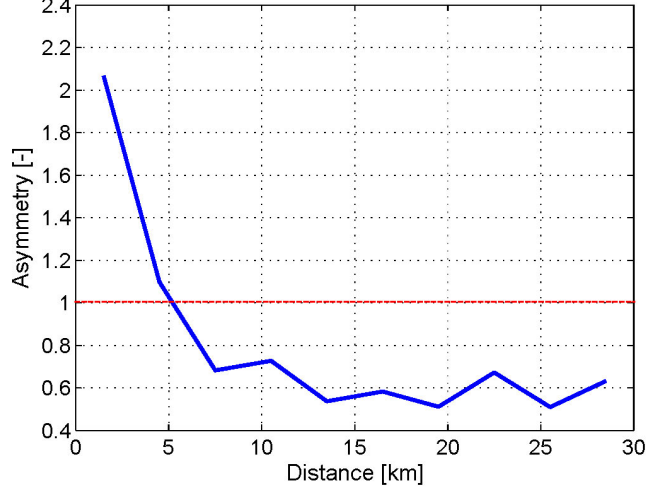


Figure 3.4: *The asymmetry over distance curve for chloride concentration.*

Figure 3.4 shows that for the short distances, the upper 10% quantiles have stronger dependence than the lower 10% quantiles. While for the long distances, this relationship is reversed.

The disadvantage of this measure is that it only counts for a part of the dependence and is very much dependent on the choice of  $u$  in Equation (3.8). An improved version is developed that compares the dependence of all the quantiles below the median and all the quantiles above the median. Formally,

$$A(\mathbf{h}) = E \left[ \left( F_z(Z(\mathbf{x})) - 0.5 \right)^2 \cdot \left( F_z(Z(\mathbf{x} + \mathbf{h})) - 0.5 \right) + \left( F_z(Z(\mathbf{x})) - 0.5 \right) \cdot \left( F_z(Z(\mathbf{x} + \mathbf{h})) - 0.5 \right)^2 \right] \quad (3.9)$$

and can be calculated via the bivariate copulas by:

$$A(\mathbf{h}) = \int \int_{\mathbf{I}^2} \left( (u - 0.5)^2 (v - 0.5) + (u - 0.5) (v - 0.5)^2 \right) dC(u, v; \rho) \quad (3.10)$$

where  $\rho$  refers to the correlation (between the random variables  $U$  and  $V$ ) corresponding a certain separation vector  $\mathbf{h}$ .

The measure of asymmetry gives a value of 0, if the dependence is symmetrical (e.g. Gaussian dependence) or there is no dependence (uniform dependence). It takes on positive values if the high quantiles have stronger dependence than the low quantiles and vice versa. Figure 3.5 shows the bivariate copula densities corresponding to the three cases. The upper

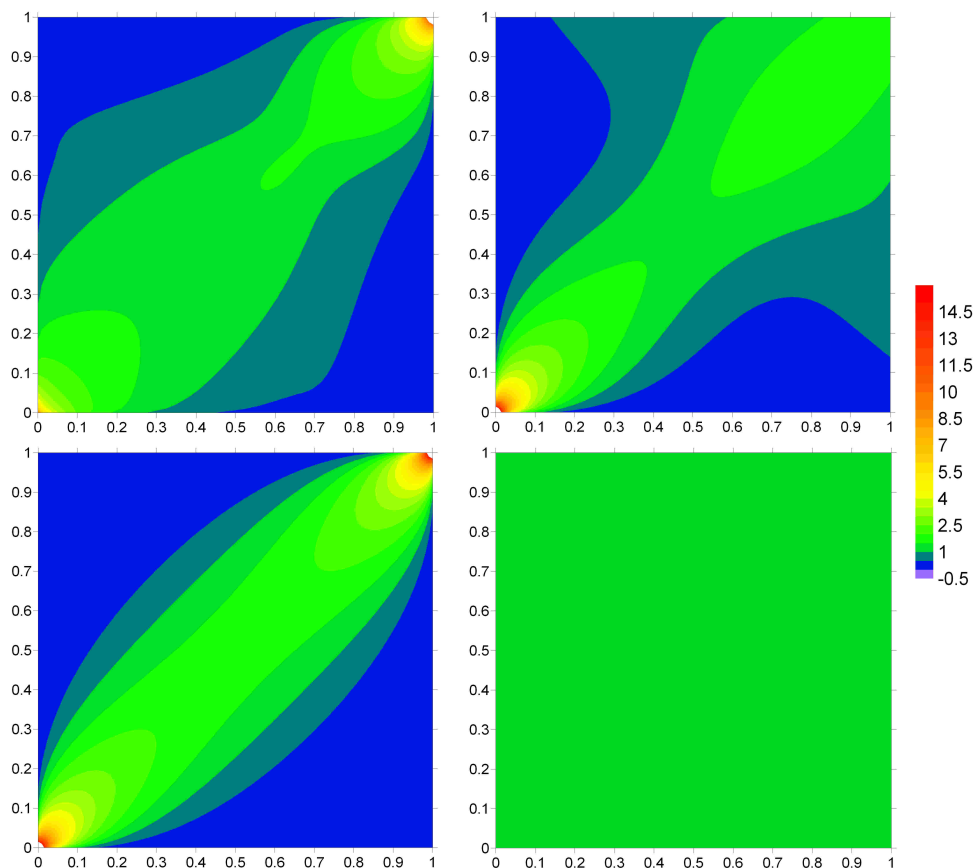


Figure 3.5: *Bivariate copula densities with different asymmetry properties.*

left chart shows a copula with positive asymmetry, the upper right one shows a copula with negative asymmetry, and the lower two charts are all corresponding to 0 asymmetry.

As shown in Equation (3.9), the new version of asymmetry is a third moment statistics. Therefore it is able to distinguish spatial patterns which can not be differentiated by the conventional descriptors, i.e. histograms and covariance structures. Figure 3.6 shows two hydraulic conductivity fields with exactly the same marginal distribution (left chart of Figure 3.7) and approximately the same rank correlation structure (right chart of Figure 3.7) which, however, have clearly distinct spatial patterns. The difference between the two fields can be detected immediately after studying the asymmetry over distance curves.

Figure 3.8 shows the corresponding asymmetry over distance curves and unravels the difference between the two fields. It does not mean that the asymmetry is sufficient to characterize any kind of spatial field. But it demonstrates that it is more powerful than the traditional variograms/covariance functions which are only good for Gaussian dependence. In some cases it can be used to detect whether the dependence is Gaussian or non-Gaussian. But in order to find out what kind of non-Gaussian dependence it is, the study of empirical copulas provides an insight into this question. Figure 3.9 displays the empirical bivariate copula

densities of these two fields for the shortest separation class which also reveals the clear differences in the spatial dependences. From Figure 3.9, we can see that the dependence of the first field is close to Gaussian, while the dependence of the second one is skewed to the extreme high value side. In addition, a secondary peak in the middle to high quantiles appears in the dependence of the second field.

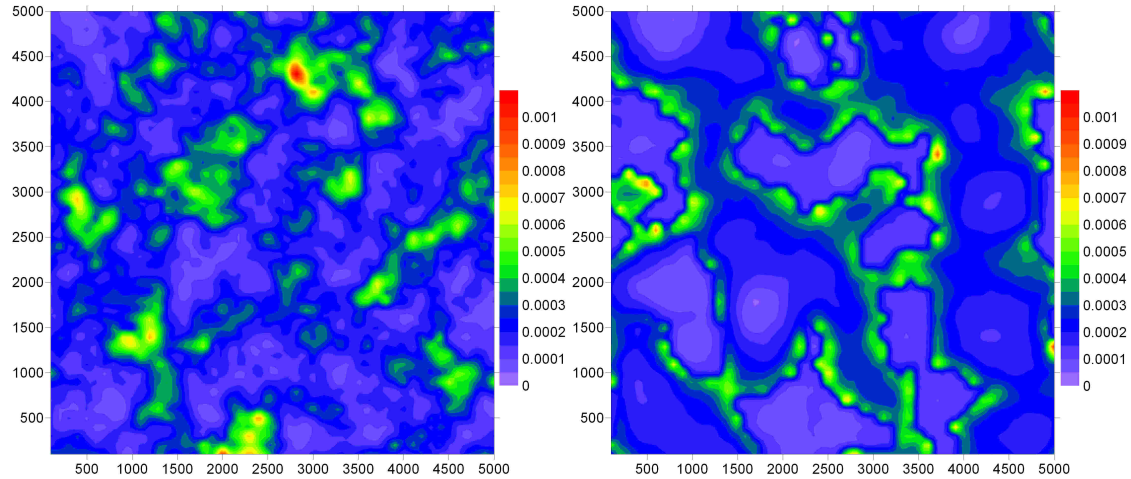


Figure 3.6: *Two hydraulic conductivity fields sharing the same marginal distribution and rank correlation structure.*

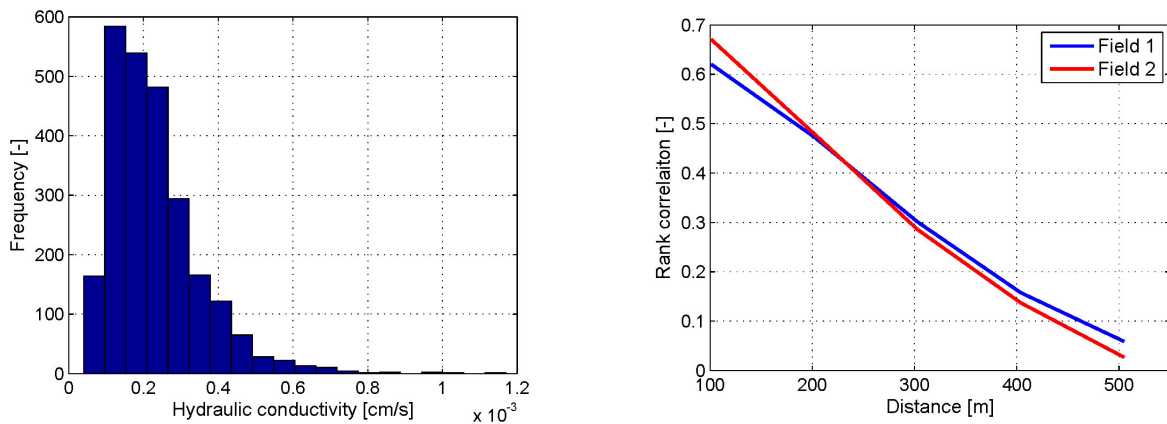


Figure 3.7: *Marginal distributions and rank correlation structures of the two hydraulic conductivity fields.*

The examples demonstrated above shows that the natural spatial variables often exhibit non-symmetrical non-Gaussian dependence. In order to model these variables, theoretical copulas should be constructed that are able to model non-Gaussian dependence. In the following chapter, three types of theoretical copulas, including the Gaussian copulas, are introduced.

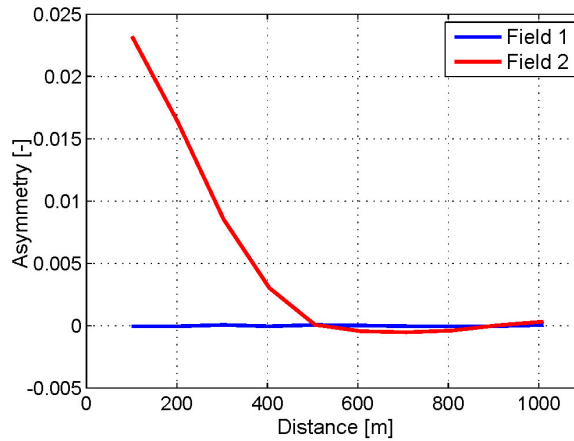


Figure 3.8: *Asymmetry over distance curves of the two hydraulic conductivity fields.*

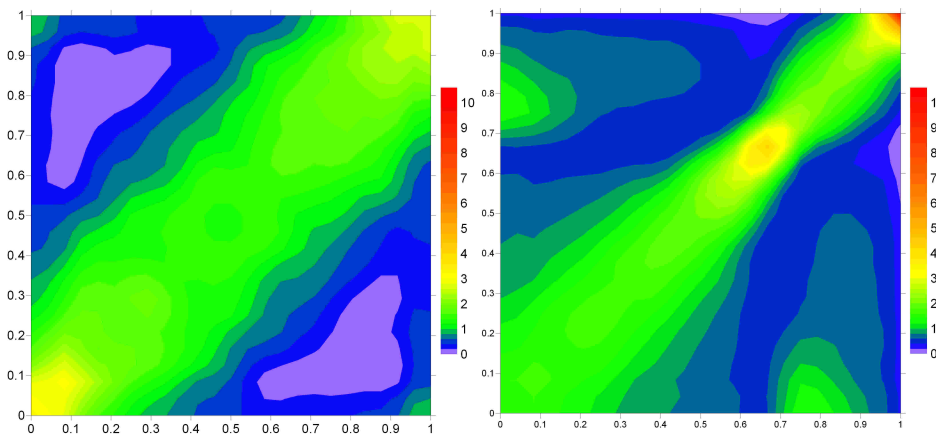


Figure 3.9: *Empirical bivariate copula densities for the shortest distance class of the hydraulic conductivity fields shown in Figure 3.6.*

### 3.3.3 Euler Characteristics

When flow and transport problem in porous media is concerned, it is attractive to predict flow and transport behaviour based on investigations of spatial structures which are directly observable and requires no expensive experiments. The most relevant property of the porous media is, at the pore scale, the way how the pores are interconnected, or, at the large continuum scale, the connectivity of various conductive regions [87]. Thus the topological characteristics of the porous media play an important role in this framework.

Different approaches are available towards the quantification of topology in porous media, as measure of percolation probabilities [37] [38] and Euler characteristics [86] [87] [77].The

former one evaluates the probability of finding a continuous path through a sample of a given size. This approach can be successfully applied to provide the required topological information for isotropic structures such as sandstone [39] [91]. However, the problem of this approach is that it entails the knowledge of the complete structure or at least a representative subsample, which is hard to get for highly heterogeneous porous media where various structural units, e.g., root channels, cracks and earthworm barrows are present. On the contrary, an unbiased estimation of the Euler characteristic is possible from local measurements and the edge effects as well as the shape of the studied sample are irrelevant. *Hans-Jörg Vogel* has demonstrated this fact in his paper entitled *Topological Characterization of Porous Media* [87]. Because of its advantage compared with the former one, the later approach is applied in this thesis work to investigate the connectivity of spatial realizations of hydraulic conductivity fields drawn from different copula models (see Chapter 7).

In the following, the concept of using Euler characteristics to quantify topology of porous media is briefly reviewed.

The Euler characteristic  $\chi$  combines the basic topological measures: the number  $N$  of isolated objects (for example pores), the number of redundant connections  $C$  within the pore space and, in three dimensions, the number of completely enclosed cavities  $H$  [73]. The mathematical expression is:

$$\chi = N - C + H \quad (3.11)$$

It is a measure of connectivity which takes on positive values for poorly connected structures where  $N > C$  (in two dimensional cases) and vice versa. In continuous fields, the Euler characteristic  $\chi$  is calculated as a function of threshold values  $t$ . For each  $t$ , a binary image can be obtained by excluding the values above the  $t$  and retaining the rest or vice versa. Based on the binary images of all the thresholds, the corresponding Euler characteristics and thus the connectivity function  $\chi(t)$  can be calculated. Figure 3.10 shows the binary images for different threshold values of a spatial field with multi-Gaussian dependence. Figure 3.11 shows the original spatial field (left) and the corresponding plot of the Euler characteristics against the threshold values.

Comparing Figure 3.10 and 3.11, one can see that at the very beginning, i.e.,  $t = 0.1$ , numerous isolated blobs emerge which lead to a high positive value of Euler's number. As the threshold increases, the area of the white blobs grows and at the same time some connections among the isolated blobs take shape. This causes the Euler's number to drop. Once the whole white structure becomes connected, the Euler's number drops to negative values. As the threshold further increases, the loops of redundant connections are erased and the curve of Euler characteristic shoots up again. When the threshold reaches its maximum, all values will become below  $t$  and the binary image will totally turn white. In this case, the number of objects  $N$  equals to 1 and the number of connections  $C$  is 0, which results in 1 as the final value of Euler's number. From the plot of the original field, it is also clear that the greenish and yellowish regions representing the middle quantiles are most connected and are corresponding to the sharp decrease in the Euler characteristic curve. While the purple

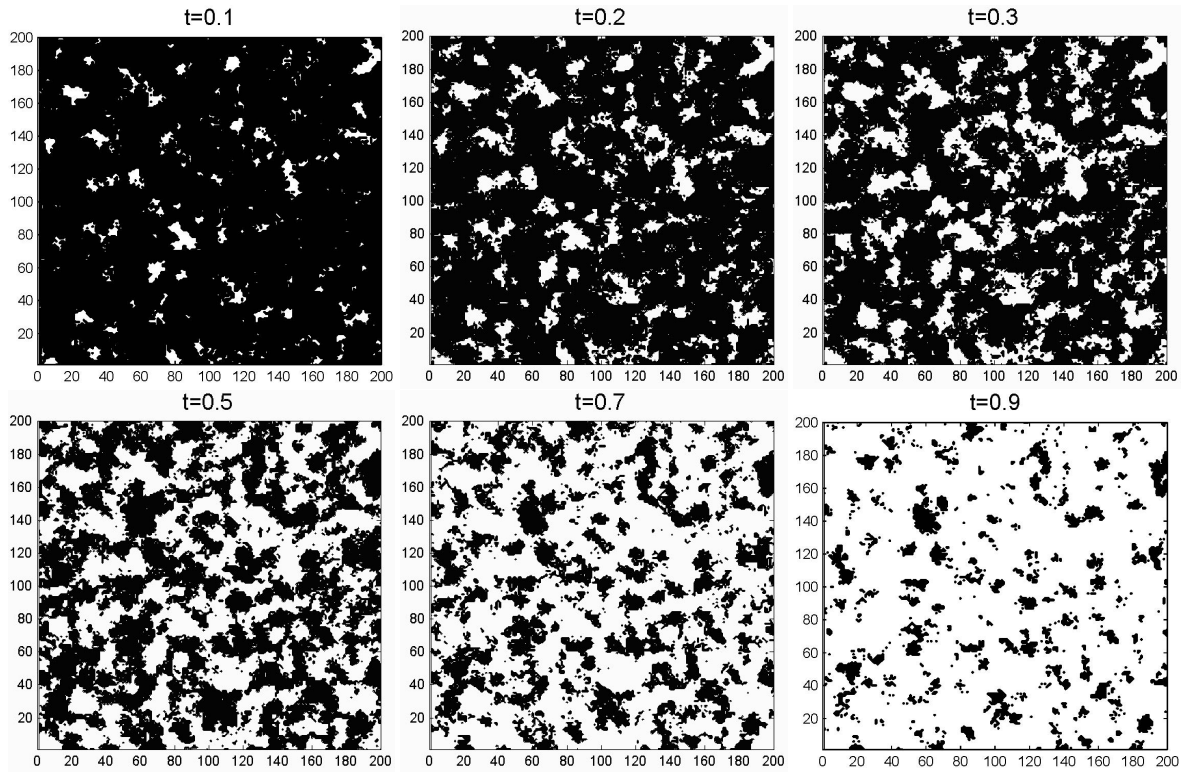


Figure 3.10: Binary images according to different threshold values  $t$  of a spatial field. Black color indicates values above  $t$ , while white color indicates values below  $t$ .

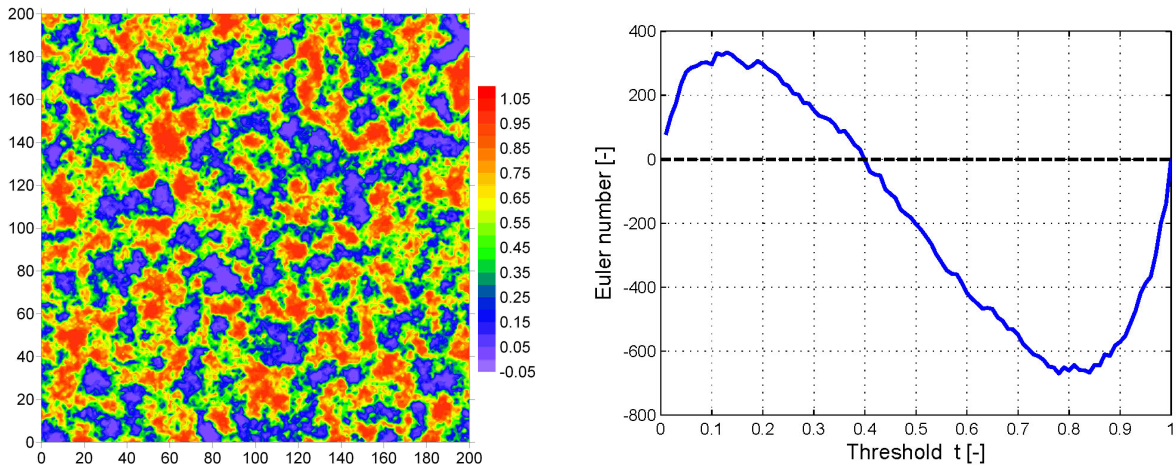


Figure 3.11: The original spatial field (left panel) of the binary images shown in Figure 3.10 and its Euler characteristic against threshold plot (right).

and red regions indicating the extreme low and high quantiles, respectively, are isolated and match the rises of the curve at the two extreme ends.

For most geostatistical simulation approaches, the topological characteristic of the generated spatial pattern is merely an implicit consequence of the simulation process [87]. As will be



addressed in Chapter 7, traditional geostatistical simulation approaches take the univariate distribution and covariance structure of the studied parameter as sole descriptors of the spatial dependence and therefore fail in some cases to reproduce the desired connectivity property. However the copula based simulation methods developed in this thesis work take the multivariate distribution into account and thus are expected to be able to better mimic the disired structures. Studies on the empirical copulas and connectivity functions  $\chi(t)$  of several spatial realizations indicate some relationship between the underlying copulas of the spatial field and its topological characteristics. Figure 3.12 illustrates this point. The calculations start from the high threshold end, which means the values below  $t$  are excluded.

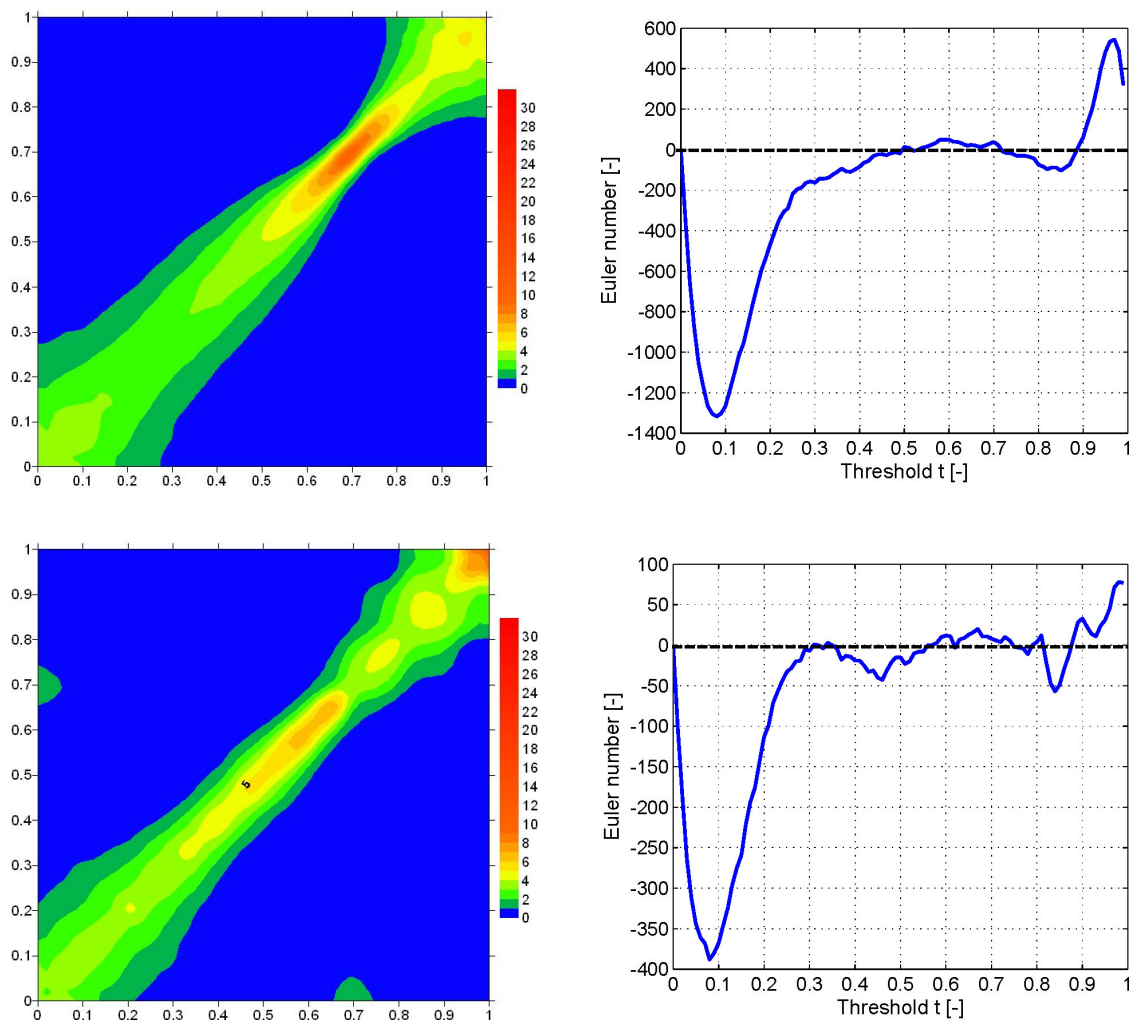


Figure 3.12: Comparison of empirical copulas (right) and Euler characteristic curves (left) of two spatial fields.

From Figure 3.12, it can be found that the meandering of the Euler characteristic curves correspond well to the shape of the copula densities. When there is a local maximum in the copula density plot, there is a peak in the curve. And when there is a local minimum in the copula density, there is also a valley in the Euler characteristic curve. In addition, it

seems that connectivity and copula density follow a kind of inverse relationship, namely, the higher the copula density, the less connected the corresponding quantiles, and vice versa. This indicates that if the underlying copula of a the spatial field of studied parameter is captured and can be well modelled, then it is very likely that the connectivity property of the field can also be well reproduced. However, the quantiles at which the peaks and valleys occur in the copula density plots do not match exactly with the quantiles at which the peaks and valleys occur in the Euler characteristic curves. Therefore the former statement is still a conjecture and needs to be investigated in a more systematic way. This is, however, not within the scope of this thesis and can be included in the outlook of the future work.



## 4 Theoretical Spatial Copulas

The basic hypothesis described in Chapter 3 allows the application of copulas for investigation of spatial dependence. The empirical copulas reflect the bivariate dependence structure of the observed data. However, for interpolation and simulation purposes, a multivariate dependence model is required. The multivariate theoretical copula of the random variables associated with a set of locations should thus have bivariate marginals resembling the estimated empirical copulas.

To define a theoretical copula which is suitable for spatial modeling, the following conditions have to be fulfilled [8]:

- The multivariate copula with any dimension  $n > 2$  can be built up based on its bivariate marginals.
- For a geographically close set of points which have an arbitrarily strong dependence, the spatial copula should allow full dependence (*Fréchet-Hoeffding upper bound* [66]). On the other hand, for distant points, the spatial copula should allow independence.
- The multivariate copula should have a parameterization such that the dependence structure reflects the geometric configuration of the corresponding set of points.

These conditions require that a spatial copula model should allow a wide range of dependence and flexible parameterization. There are numerous existing theoretical copulas, e.g., the Archimedean family [33], the Farlie-Gumbel-Morgenstern copula [66], nevertheless most of them do not fulfill the conditions listed above therefore are not suitable to describe the dependence of random functions in space. It is quite common that most of the known multivariate copulas only depend on a few parameters which causes the problem that the dependence remains the same irrespective of the distances among the points. This leads to the violation of the last condition mentioned before (*Bardossy* [2006] [8]).

The multivariate elliptical copulas (*Fang et al.*, [2002] [27]), including the Gaussian copula, fulfill all the required conditions. Hidden in the correlation matrix  $\Gamma$  of these copulas, the geometric configuration of data points is reflected. They can be used for modelling spatial dependence structures. However, the disadvantage of this family is that they are fully symmetrical in the sense of Equation (2.13) [2], which largely restricts the flexibility of this type of copulas to represent the spatial dependence structure of a natural variable. As will be shown in the sections of dataset investigations in Chapter 6 and 7, it is quite often that the spatial dependence of a natural variable is non-symmetric.

Another possible alternative is the Farlie-Gumbel-Morgenstern copula [66], but it has the problem that the correlation between pairs of variables decreases to zero as the number of points increases. This violates the second condition [8].

In this PhD work, two types of non-Gaussian theoretical spatial copulas are constructed by applying non-monotonic transformations of a Gaussian variable. The resulting copulas inherit the desirable properties of Gaussian copula to fulfill all the required conditions and at the same time they are able to describe non-symmetrical dependence structures. Before introducing the new non-Gaussian theoretical copulas, the Gaussian copulas are first discussed in the next section.

## 4.1 Gaussian Copulas

The Gaussian copula is derived from the multivariate Gaussian distribution which is defined by its correlation matrix. Let  $\Phi$  denote the univariate standard normal distribution function, and  $\Phi_{\Gamma,n}$  the  $n$ -dimensional Gaussian distribution with correlation matrix  $\Gamma$ . Then the corresponding  $n$ -dimensional Gaussian copula can be written as:

$$C_{\Gamma}(u_1, \dots, u_n) = \Phi_{\Gamma,n}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)) \quad (4.1)$$

whose density reads:

$$c_n(u_1, \dots, u_n) = \frac{1}{\sqrt{|\Gamma|}} \left( -\frac{1}{2} \mathbf{x}^T (\Gamma^{-1} - \mathbf{I}) \mathbf{x} \right) \quad (4.2)$$

where  $x_i = \Phi^{-1}(u_i)$  and  $\mathbf{I}$  is the identity matrix.

Hidden in the correlation matrix, Gaussian copulas have the ability to reflect the spatial configuration of the data points. However, its flexibility is largely restricted by its symmetry property which is often in contradiction with natural dependence structures. Figure 4.1 shows the bivariate Gaussian copula densities with different correlation  $\rho$  values. Here the  $\rho$  refers to the Pearson's rho [74] of the underlying normal variables  $X_1$  and  $X_2$  and the correlation matrix  $\Gamma$  also refers to the  $n$ -dimensional matrix of Pearson's rhos among the normal variables  $X_1, \dots, X_n$ .

As can be seen from Figure 4.1 that a Gaussian copula has strongest dependence strength at the two extreme ends. The weakest dependence lies in the middle quantile. This characteristic can be utilized in the construction of desired non-Gaussian nonsymmetrical copulas by applying non-monotonic transformations on Gaussian random variables, which will be illuminated in the following sections.

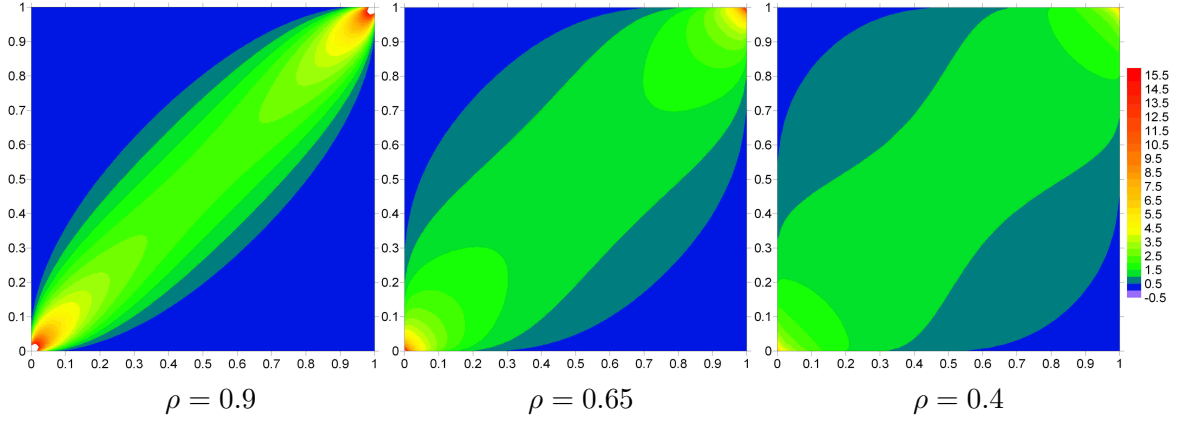


Figure 4.1: Bivariate Gaussian copula densities with different correlation  $\rho$  values

## 4.2 V-transformed Normal Copulas

Since non-monotonic transformation changes the ranks of the variable values, with which copulas are calculated, high dimensional copulas with non-symmetrical dependence of their bivariate marginals can easily be obtained by non-monotonic transformations of the multivariate elliptical distributions. The V-transformed normal copula is for example constructed by a 'V' shaped non-monotonic transformation. Formally, the transformation reads in equation (4.3). Let  $\mathbf{Y}$  be  $N(\mathbf{0}, \mathbf{\Gamma})$ , an  $n$  dimensional normal random variable with  $\mathbf{0}^T = (0, \dots, 0)$  mean and  $\mathbf{\Gamma}$  correlation matrix. All marginals are supposed to have unit variance. Let  $\mathbf{X}$  be defined for each coordinate  $j = 1, \dots, n$  as:

$$X_j = \begin{cases} k(Y_j - m)^\alpha & \text{if } Y_j \geq m \\ m - Y_j & \text{if } Y_j < m \end{cases} \quad (4.3)$$

Where  $k$  is a positive constant,  $m$  and  $\alpha$  are arbitrary real numbers. When  $k = 1$  and  $\alpha = 1$ , this transformation leads to the multivariate non centered  $\chi$ -square distribution [8]. For the case that  $m = 0$ ,  $k = 1$  and  $\alpha = 1$ , the transformation is the same as that applied by Zinn and Harvey in their paper [95] to create fields with connected high or connected low values.

The V-shaped transformations with different parameter sets are illustrated in figure 4.2.

The consequence of this transformation is that the values less than  $m$  which are small values in the old space, turn to high values. The value  $m$  becomes the smallest value in the new space and the values greater than  $m$  become either more concentrated or diviated from the value  $m$  depending on the settings for  $k$  and  $\alpha$ . Then the dependence associated with the values close or equal to  $m$  counts for the dependence of the low values in the new space. The dependence of the previous low values is foldered to the high-value-side. The effect of the transformation can also be explained by the change in the marginal distribution. Figure 4.3 illustrates how the mass in the distribution is folded after the transformation.

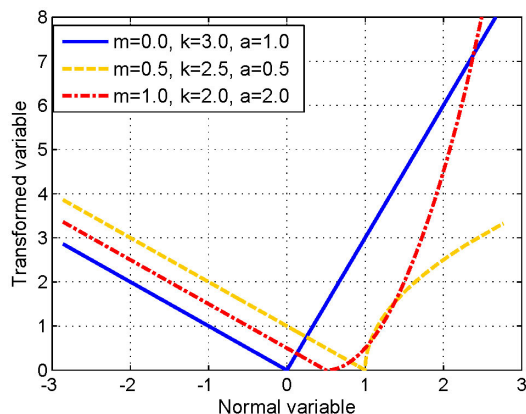
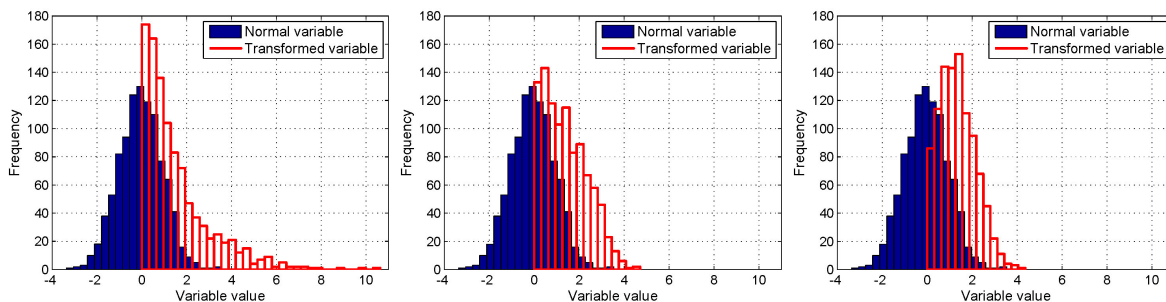
Figure 4.2: *V-shaped transformations.*

Figure 4.3: *Consequence of the V-shaped transformation with parameters:  $m = 0.0$ ,  $k = 3.0$  and  $\alpha = 1.0$  (left chart);  $m = 0.0$ ,  $k = 2.5$  and  $\alpha = 0.5$  (middle),  $m = 1.0$ ,  $k = 2.5$  and  $\alpha = 0.5$  (right)*

Figure 4.3 shows that under the same  $m$  value, a higher  $k$  and  $\alpha$  brings the mass in the distribution more skewed to the low quantiles in the new space and the mass is more widely spread. In addition, if the values of  $k$  and  $\alpha$  remain unchanged, but the value of  $m$  is increased, the mass is more concentrated to the median quantiles in the new space and the distribution looks more symmetrical. The consequence of these differences on the resultant dependence is shown in Figure 4.4 which displays the bivariate copula densities of the random variables obtained by the transformations illustrated in Figure 4.3. The corresponding  $\rho$  values of the underlying bivariate normal distributions are 0.8 for all the three cases.

From Figure 4.4, we can see that nonsymmetrical dependences are created. And the shape of the copulas is dependent on the transformation parameters. Comparing Figure 4.3 and 4.4, it seems that the more skewed the resulting marginal distribution, the more unsymmetric the resulting copula and vice versa. To make it not confused with the relationship between the marginal distribution and the multivariate dependence of the random variable, it must be pointed out that it is not the symmetry of marginal distribution which influences the symmetry of the dependence, but the reshuffle of the ranks of the variates and the dependence strength associated with the quantiles that are shuffled by the non-monotonic transforma-

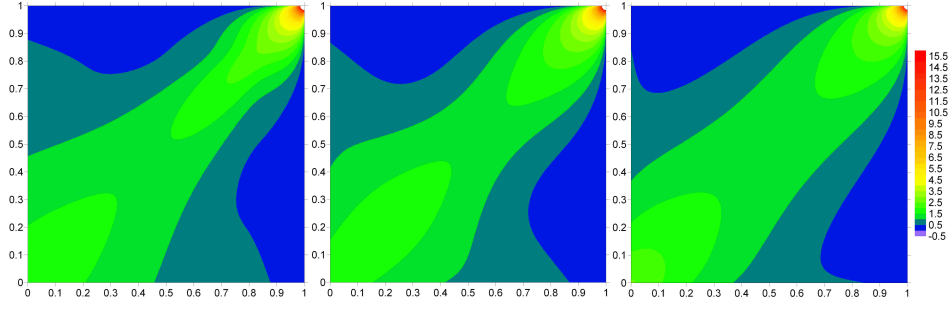


Figure 4.4: *Bivariate copula densities of the random variables after the transformations in Figure 4.3 corresponding to the  $\rho = 0.8$  of the underlying bivariate normal distribution.*

tion. As mentioned before, the old middle quantiles account for the low quantiles in the new space, since originally, the middle quantiles have weak dependence, now the weak dependence dominates the low quantiles. A more concave shape of the right branch of the transformation shown in Figure 4.2 shifts the part of the middle quantiles to the high quantile side and reduces the magnitude of the old high quantiles as well, thus mixes the old middle quantiles and high quantiles in the new space, the frequency for different values are more equalized (a less spreading marginal distribution) and balances the difference in dependence strength. On the other hand, a higher  $m$  cause the left branch of the transformation dominates the right one, the ranks of the variates is less changed, the resulting marginal distribution remains more similar to the original one and the dependence is more symmetric and more Gaussian.

The one dimensional marginals of the resultant variable  $\mathbf{X}$  are identical and have the distribution function:

$$\begin{aligned}
 H_1(x) &= P(X < x) \\
 &= P\left(Y < \left(\frac{x}{k}\right)^{\frac{1}{\alpha}} + m\right) + P(Y > x - m) \\
 &= P\left(Y < \left(\frac{x}{k}\right)^{\frac{1}{\alpha}} + m\right) - P(Y > x - m) \\
 &= \Phi\left(\left(\frac{x}{k}\right)^{\frac{1}{\alpha}} + m\right) - \Phi(-x + m)
 \end{aligned} \tag{4.4}$$

and density:

$$h_1(x) = \frac{1}{k\alpha} \cdot \left(\frac{x}{k}\right)^{\left(\frac{1}{\alpha}-1\right)} \cdot \phi\left(\left(\frac{x}{k}\right)^{\frac{1}{\alpha}} + m\right) + \phi(-x + m) \tag{4.5}$$

$\Phi$  being the distribution function of the standard normal distribution and  $\phi$  the corresponding density function. The corresponding multivariate distribution function is:

$$\begin{aligned}
H_n(x_1, \dots, x_n) &= P(X_1 < x_1, \dots, X_n < x_n) = \\
&= \sum_{i=0}^{2^n-1} (-1)^i \Phi(\zeta_i + \mathbf{m})
\end{aligned} \tag{4.6}$$

where

$$\zeta_i^T = (b((-1)^{i_1}) x_1^a, \dots, b((-1)^{i_n}) x_n^a) \tag{4.7}$$

$i_j = 0$  or  $1$  and

$$i = \sum_{j=0}^{n-1} i_j 2^j \tag{4.8}$$

and

$$b = \begin{cases} -1 & \text{if } (-1)^{i_j} = -1 \\ \frac{1}{k} & \text{if } (-1)^{i_j} = 1 \end{cases} \tag{4.9}$$

and

$$a = \begin{cases} 1 & \text{if } (-1)^{i_j} = -1 \\ \frac{1}{\alpha} & \text{if } (-1)^{i_j} = 1 \end{cases} \tag{4.10}$$

Details about the derivation and explanations of the above equations can be found in the Appendix.

By taking the derivatives of the distribution function, the density  $h_n$  has the form:

$$\begin{aligned}
h_n(x_1, \dots, x_n) &= \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{\Gamma}^{-1}|^{\frac{1}{2}}} \sum_{i=0}^{2^n-1} \frac{1}{(\alpha k)^{n-\sum_{j=0}^{n-1} i_j}} \cdot \left( \left( \frac{x}{k} \right)^{\frac{1}{\alpha}-1} \right)^{n-\sum_{j=0}^{n-1} i_j} \\
&\cdot \exp \left( -\frac{1}{2} (\zeta_i + \mathbf{m})^T \mathbf{\Gamma}^{-1} (\zeta_i + \mathbf{m}) \right)
\end{aligned} \tag{4.11}$$

The density of the  $v$ -transformed multivariate normal copula can be calculated through the joint and marginal distribution densities:

$$c_n(u_1, \dots, u_n) = \frac{h_n(x_1, \dots, x_n)}{h_1(x_1) \cdot h_1(x_2) \dots h_1(x_n)} \tag{4.12}$$

with  $u_i = H_1(x_i)$ . As mentioned before, when  $k = 1$ , the transformation is a symmetrical modulus transformation, therefore the non-centered  $\chi$ -square copula [8] is a special case of the  $v$ -transformed multivariate normal copula for  $k = 1$  and  $\alpha = 1$ . Furthermore, the effect of non-monotonic transformation vanishes with  $m \rightarrow \pm\infty$  and the resulting copula converges to a Gaussian copula. In this sense, Gaussian copula is also a special case of  $v$ -transformed normal copula for  $m = \infty$ . Figure 4.5 shows the bivariate copula densities for different combinations of the values of  $k$ ,  $m$ ,  $\alpha$  and correlations  $\rho$ .

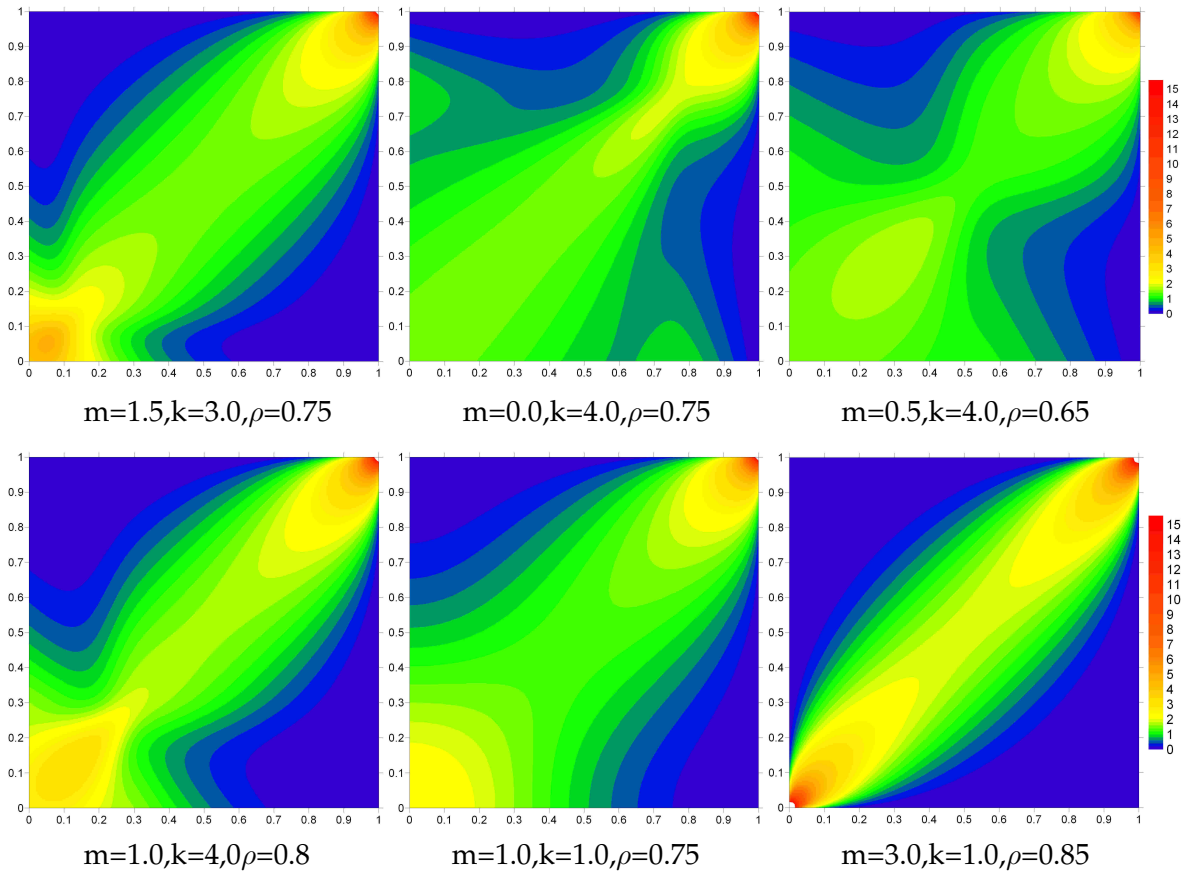


Figure 4.5: Bivariate copula density of  $v$ -transformed normal copula with different parameter values.

Notice that when  $k = 1$  and  $\alpha = 1$ , it reverts to a non centered  $\chi$ -square copula, and when  $k = 1$  and  $m = 3$ , it approaches to a Gaussian copula.

The  $v$ -transformed normal copula has the strongest dependence for the high values. The opposite dependence structure can be obtained by simply taking the copula:

$$\hat{c}'_n(u_1, \dots, u_n) = c_n(1 - u_1, \dots, 1 - u_n) \tag{4.13}$$

Figure 4.6 shows the upper skewed bivariate copula density and its reversed version:

### 4.3 Maximum Normal Copulas

Another way of constructing non-symmetric copula which fulfils the conditions described at the beginning of this chapter is to derive the copula of maximum or minimum of two Gaussian processes. Since it can be proved that the latter one is a special case of the former one, in the following only the copula of maximum of two Gaussian processes is discussed, which is called as 'maximum normal copula' hereafter.

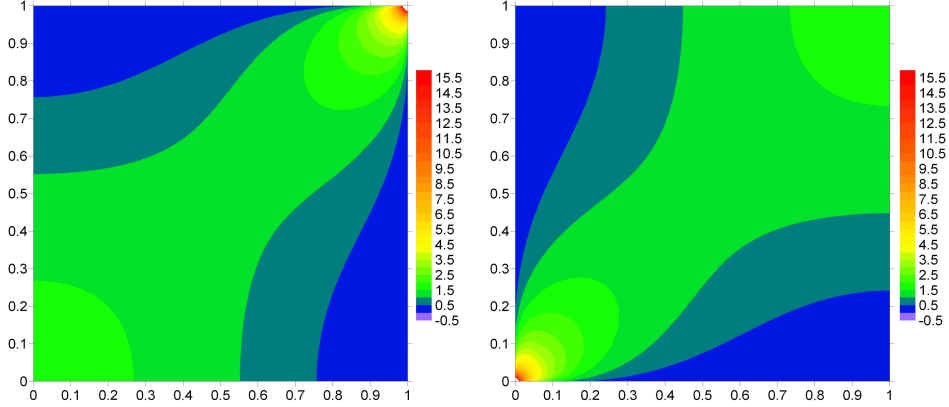


Figure 4.6: *Bivariate copula density of a  $v$ -transformed normal copula (left) and its reversed version (right).*

Let  $\mathbf{Z} = \max(\mathbf{Y}, \mathbf{X})$ , with  $\mathbf{Y}$  distributed as  $N(\mathbf{0}, \Gamma_1)$  being an  $n$  dimensional normal random variable with  $\mathbf{0}^T = (0, \dots, 0)$  mean and  $\Gamma_1$  correlation matrix, and  $\mathbf{X}$  distributed as  $(\mathbf{m}, \Gamma_2)$  also being an  $n$  dimensional normal random variable with  $\mathbf{m}^T = (m, \dots, m)$  mean and  $\Gamma_2$  correlation matrix. All marginals of  $\mathbf{Y}$  are supposed to have unit variance, while the marginals of  $\mathbf{X}$  have  $\sigma^2$  variance. Further,  $\mathbf{Y}$  and  $\mathbf{X}$  are independent of each other.

Then all the one dimensional marginals of  $\mathbf{Z}$  are identical:

$$F(z) = P(\max(Y, X) < z) = P(Y < z, X < z) = P(Y < z) \cdot P(X < z) = \Phi(z) \cdot \Phi\left(\frac{z - m}{\sigma}\right) \quad (4.14)$$

and the density is:

$$f(z) = \phi(z) \cdot \Phi\left(\frac{z - m}{\sigma}\right) + \Phi(z) \cdot \phi\left(\frac{z - m}{\sigma}\right) \cdot \frac{1}{\sigma} \quad (4.15)$$

The  $n$ -dimensional distribution function is derived as following:

$$\begin{aligned} G(z_1, \dots, z_n) &= P(\max(Y_1, X_1) < z_1, \dots, \max(Y_n, X_n) < z_n) \\ &= P(Y_1 < z_1, X_1 < z_1, \dots, Y_n < z_n, X_n < z_n) \\ &= P(Y_1 < z_1, \dots, Y_n < z_n) \cdot P(X_1 < z_1, \dots, X_n < z_n) \\ &= \Phi(z_1, \dots, z_n) \cdot \Phi\left(\frac{z_1 - m}{\sigma}, \dots, \frac{z_n - m}{\sigma}\right) \end{aligned} \quad (4.16)$$

The  $n$ -dimensional density function is not trivial, it is derived as following:



$$\begin{aligned}
g(z_1, \dots, z_n) &= \frac{\partial^n G(z_1, \dots, z_n)}{\partial z_1 \dots \partial z_n} \\
&= \frac{\partial^{n-1}}{\partial z_1 \dots \partial z_{n-1}} \left( \frac{\partial F_1}{\partial z_n} \cdot F_2 + F_1 \cdot \frac{\partial F_2}{\partial z_n} \right) \\
&= \sum_{K=1}^{2^n} \frac{\partial^i F_1}{\partial z_1^{l_1} \cdot \partial z_2^{l_2} \dots \partial z_n^{l_n}} \cdot \frac{\partial^j F_2}{\partial z_1^{m_1} \cdot \partial z_2^{m_2} \dots \partial z_n^{m_n}} \quad (4.17)
\end{aligned}$$

where  $F_1 = \Phi(z_1, \dots, z_n)$ ,  $F_2 = \Phi\left(\frac{z_1 - m}{\sigma}, \dots, \frac{z_n - m}{\sigma}\right)$ ,  $i = 0, 1, \dots, n$  and  $j = n - i$ ,  $l_k = 0, 1$  and  $\sum_{k=1}^n l_k = i$ , if  $l_k = 1$  then  $m_k = 0$ , else  $m_k = 1$ .

The  $2^n$  terms of the products of different partial derivatives and mixed derivatives of the  $n$ -dimensional Gaussian distribution can be very computationally demanding. Therefore a numerical solution is applied to calculate the density from the distribution by the following equation:

$$g(z_1, \dots, z_n) = \frac{\sum_{k=1}^{2^n} (-1)^k G(z_1 + j_1 \Delta z_1, \dots, z_n + j_n \Delta z_n)}{\Delta z_1 \cdot \Delta z_2 \dots \Delta z_n} \quad (4.18)$$

where  $k = n - \sum_{i=1}^n j_i$  and  $j_i = 0, 1$ . Unfortunately, this numerical solution is only applicable for  $n \leq 4$ , else the precision of the approximation by Equation (4.18) is too low and negative density values occur for a discretization length of  $z$  equivalent to 0.1 in the copula space. This is obviously already a very rough discretization.

Figure 4.7 shows bivariate copula densities of the maximum normal copula with different parameter values.

Notice that the plot on the upper left corner is identical with the middle one of the second row, and the middle plot is identical with the lower left one. This shows how the ratio of the standard deviations ( $\sigma$ ) and the ratio of correlations ( $\rho$ ) of the two underlying normal processes compensate each other. The lower right plot of Figure 4.7 shows that if the mean value  $m$  of the second process is increased, while its standard deviation  $\sigma$  and the correlations ( $\rho$ ) of both processes remain unchanged. The copula is more symmetric and closer to the Gaussian copula.

Figure 4.8 shows two examples of realizations of the maximum normal copulas with different parameter settings in the copula space. For the field on the left hand side,  $m = 0.0$  and  $\sigma = 4.0$ . While for the field on the right hand side,  $m = 0.5$  and  $\sigma = 0.25$ . For both fields, the correlation length of the first Gaussian process is 6% of the domain length, while that of the second Gaussian process is two times of the former one.

The main difference between the two realizations shown in Figure 4.8 is the transition between the high values (indicated by the red color) and the low values (indicated by the blue color). There are obviously more abrupt transitions at some locations in the second

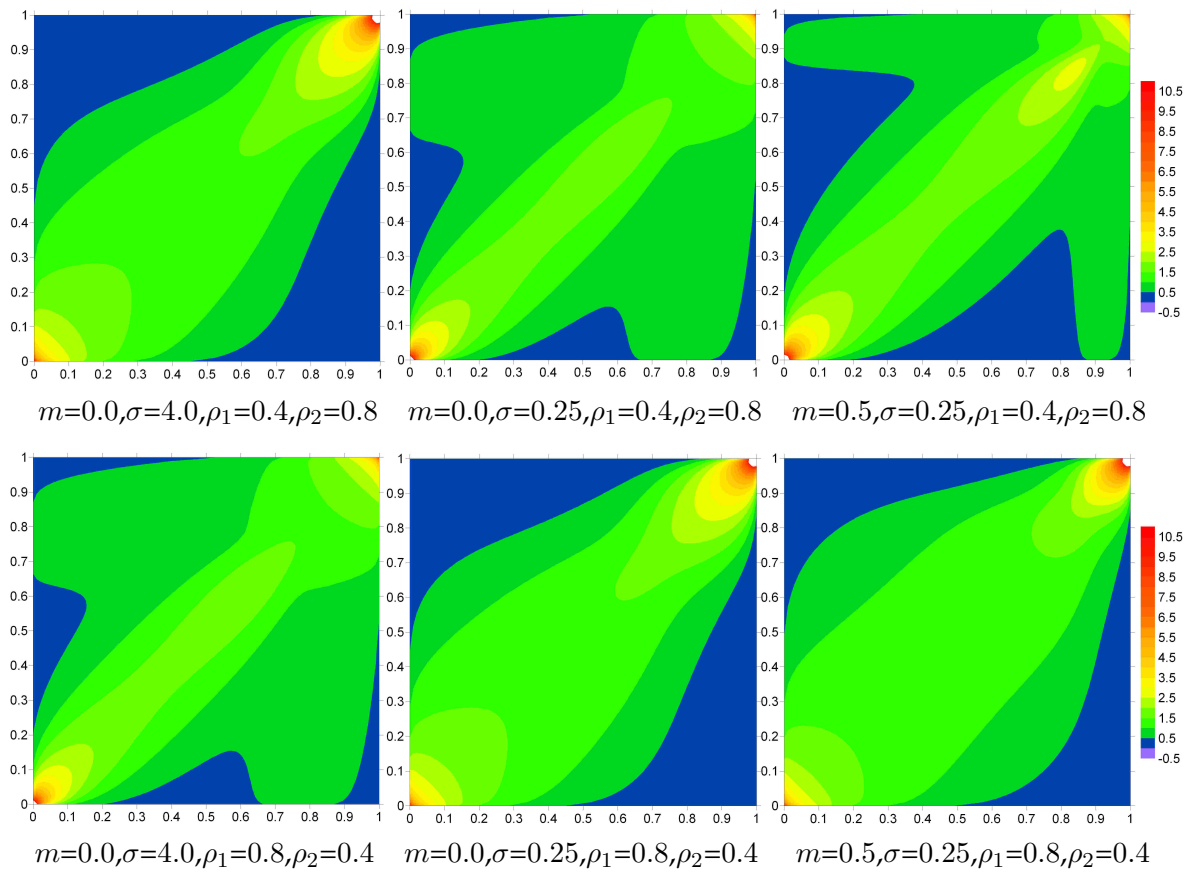


Figure 4.7: Bivariate copula densities of maximum normal copula model with difference parameter settings.

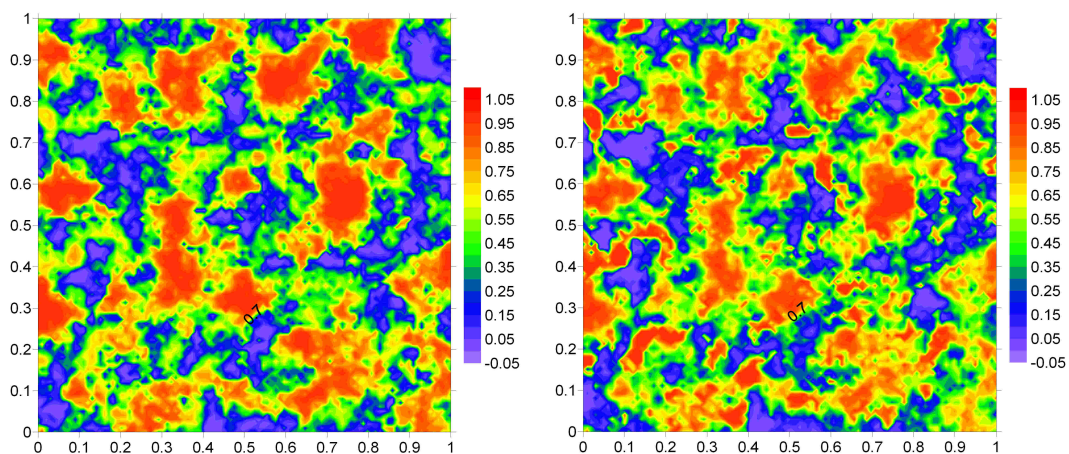


Figure 4.8: Examples of realizations of the maximum normal copulas with different parameters in the copula space.

realization. Moreover the connections among the high values are more tortuous in the second realization. Different behaviour of clusters of the high values can be observed in both realizations. Some of the clusters look very rigid, while the others look quite loose. This phenomenon which is the consequence of the effect of the two different processes is also more prominent in the second realization.

The advantage of the maximum normal copula compared with the v-transformed normal copula is that by tuning the parameters in the correlation matrices of the two underlying Gaussian variables, the asymmetry of the copula can be reversed. In practice, it means that as separation vector  $\mathbf{h}$  increases, the asymmetry of the dependence can be changed from positive to negative, or vice versa, because the correlation is a function of  $\mathbf{h}$ . This is, however, not reflected by the v-transformed normal copula but an inherent property of some natural parameters.

Figure 4.9 shows the contour plot of asymmetry against the correlations of the two underlying bivariate Gaussian variables. The colors indicate the magnitude of the asymmetry values. The parameters used are  $m = 0.5$ ,  $\sigma = 0.5$ . As can be observed that across the red dashed line which indicates for certain combinations of the two correlation values, the sign of the asymmetry is flipped.

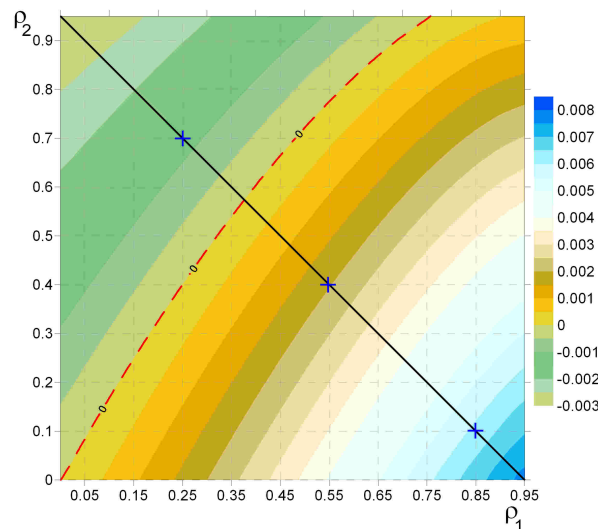


Figure 4.9: Contour plot of asymmetry against the correlations of two underlying bivariate Gaussian variables for maximum normal copula with parameters:  $m = 0.5$  and  $\sigma = 0.5$ .

Figure 4.10 shows the bivariate copula densities of the maximum normal copula with the same parameters used in Figure 4.9. The combinations of the two  $\rho$  values of these bivariate copulas are marked as crosses in Figure 4.10.

From Figure 4.10, one can see how the dependence structure evolves when the weights of the two  $\rho$  values are changed. Some natural variables have such property in their spatial dependence, namely, as the separation length increases, the sign of the measure of asym-

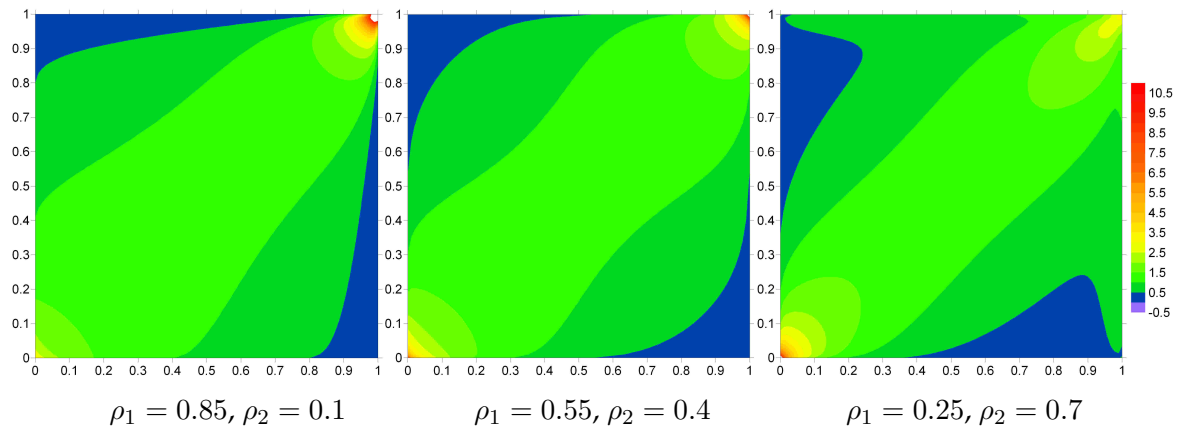


Figure 4.10: Bivariate copula densities of maximum normal copula with parameters:  $m = 0.5$  and  $\sigma = 0.5$  for different  $\rho$  combinations.

metry is reversed. Under such conditions, the maximum normal copula is preferable to the v-transformed normal copula and Gaussian copula. Since the later two are not able to reflect the change in the sign of asymmetry over distance. An example is the saturated hydraulic conductivity of the Las Cruces Trench Site which is discussed in Chapter 7. The goodness-of-fit test also manifests that the maximum normal copula is the best choice as a dependence model for this dataset.

So far, new descriptors, i.e., empirical copulas, rank correlation and asymmetry, to characterize spatial dependence are developed as well as theoretical copulas appropriate for modelling spatial dependence are invented. But in order to interpolate or simulate natural variables with desired spatial dependence, one needs to investigate observations and parameterize the selected theoretical model according to the observations. In the next chapter, the proposed methodology of model inference is presented.

## 5 Model Inference

The study of empirical copulas of datasets provides the insight into how the spatial dependence looks like, aiming at orienting the selection of a model which resembles the dependence structure well. Once the type of the model is selected, the model inference procedure can be carried out to parameterize the selected model. In geostatistics, spatial variabilities are described by empirical variograms or spatial covariance functions. The estimation of the parameters of the variogram/ covariance models can be done by minimizing the squared differences between the empirical and theoretical variograms/covariance functions. The problem of this approach is that it is not based on independent samples, which contradicts the assumption of the least square method, as one observation are considered in a number of pairs used for the calculation of variogram/covariance function. A new parameterization approach based on multiple point statistics and maximum likelihood method is adopted in this work and is discussed in the first section of this chapter. A stability and reliability test of the approach is conducted, result of which is presented in the second section of this chapter. In the following two chapters, this approach is applied to the datasets of groundwater quality parameters of Baden-Württemberg and hydraulic conductivities of Las Cruces Trench Site. The results of the applications show that the parameterized models (Gaussian copula,  $v$ -transformed normal copula and maximum normal copula) are able to reproduce the traditional statistics (variogram/correlation structure), while the non-Gaussian models produce more similar asymmetry structures as the empirical ones.

### 5.1 Model Inference

Empirical copulas provide an insight into how the spatial dependence of the natural variable under interest looks like and sheds a light on what kind of theoretical copulas can be chosen to resemble the spatial dependence. However, alike the calculation of empirical variograms or covariance functions, the calculation of spatial copulas is not based on independent samples (as each observation is accounted for in a number of pairs), hence the parameterization of a copula model on empirical copulas is not appropriate. In *Bardossy and Li* [2008] [11], a more rigorous approach was proposed which is based on multiple point statistics and maximum likelihood method.

In this approach, the observation set is first divided into several subsets of arbitrary sizes (see Figure 5.1). Although for interpolations or conditional simulations, only observations close to the unsampled locations are used as conditioning points, the parameterized theoretical copula should also respect long range spatial patterns, therefore the subsets should not only be restricted to small areas enclosing the adjacent observations, but some of them

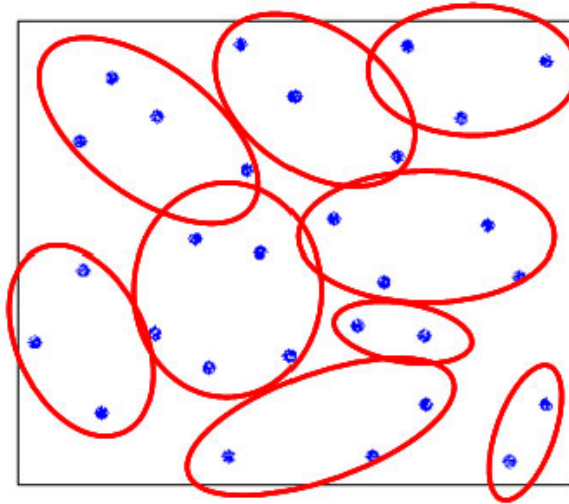


Figure 5.1: *Discretization of the studied domain for parameterization purpose*

should also include distant points. Providing the dependence of closer points is given more emphasis, a linearly decreasing function can be imposed as the relationship between the frequencies of different sized subsets and the sizes of the subsets. The size of the biggest subset should be in the same range of the correlation length of the dataset. Figure 5.2 shows an example of the distribution of the sizes of the delineated subsets.

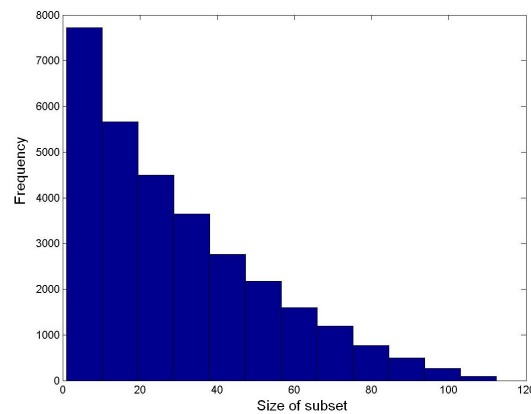


Figure 5.2: *An example of the distribution of the sizes of the delineated subsets for parameterization purpose.*

As the second step, for any of the subsets  $S_k$ , the likelihood of the parameter vector of the chosen theoretical copula can be calculated as the corresponding copula density of the points enclosed in the subset. Then as all the subsets are disjoint, under a weak assumption of independency, the overall likelihood is the product of the individual ones.

Suppose each subset  $S_k$  contains  $n(k) \geq 2$  observations:

$$S_k = \{x_{k,1}, \dots, x_{k,n(k)}\} \quad k = 1, \dots, K \quad (5.1)$$

They are disjoint:

$$S_k \cap S_j = \emptyset \text{ if } k \neq j \quad (5.2)$$

and cover all observation points

$$S = \bigcup_{k=1}^K S_k = \mathcal{S} \quad (5.3)$$

For any of the sets  $S_k$ , the likelihood of the parameter vector can be obtained by calculating the corresponding copula density.

$$c(S_k, \theta) = c\left(F_Z(Z(x_{k,1})), \dots, F_Z(Z(x_{k,n(k)})), \theta\right) \quad (5.4)$$

Then the overall likelihood of the observation points is the product of the copula densities of the individual subsets:

$$L(\theta | Z(x_1), \dots, Z(x_n)) = \prod_{k=1}^K c(S_k, \theta) \quad (5.5)$$

can be maximized with respect to the parameters  $\theta$ .

The multivariate Gaussian copula is identified through its correlation matrix  $\Gamma$ . The  $v$ -transformed normal copula is parametrized by the three transformation parameters  $m$ ,  $k$ ,  $\alpha$  and the correlation matrix  $\Gamma$  of the underlying normal  $\mathbf{Y}$ . The parameters to be estimated for the maximum normal copula includes the mean  $m$  and the standard deviation  $\sigma$  of the second normal variable as well as the parameters in the correlation structures of both underlying normal variables. Therefore the common parameterization targets of all the three theoretical copulas considered in this work are the parameters in the correlation matrix, which is explained in the follows.

Because of the stationarity assumption, the correlations between any two points can be written as a function of the separating vector  $\mathbf{h}$ . Then for any set of observations  $\mathbf{x}_1, \dots, \mathbf{x}_n$  the correlation matrix  $\Gamma$  of the normal variable  $\mathbf{Y}$  can be written as:

$$\Gamma = \left( (\rho_{i,j})_{l,l}^{n,n} \right) \quad (5.6)$$

where  $\rho_{i,j}$  refers to the Pearson's correlation coefficient [74] that only depends on the vector  $\mathbf{h}$  separating the points  $\mathbf{x}_i$  and  $\mathbf{x}_j$ :

$$\rho_{i,j} = R(\mathbf{x}_i - \mathbf{x}_j) = R(\mathbf{h}) \quad (5.7)$$

The correlation function is the positive linear combination of different models:

$$R(\mathbf{h}) = \sum_{j=0}^J D_j r_j(\mathbf{h}, L_j) \quad \sum_{j=0}^J D_j = 1 \quad (5.8)$$

Here  $r_j(\mathbf{h}, L)$  denotes a valid (positive definite) correlation function such as the exponential or spherical model corresponding to the theoretical variogram. The parameter  $L$  stands for the range (range/correlation length) of the correlation function, while  $D$  stands for the maximum contribution of each model (sill). The number of functions  $J$  is usually 2 or 3. One of the functions is usually describing the pure random fluctuations ( $L = 0$ ) while the others describe the variability on different scales.

Table 5.1 lists the parameter contained in  $\theta$  in Equation (5.4) and (5.5) for the three types of theoretical copulas:

Notice that in the derivation of the v-transformed normal copula and maximum normal copula, all the components of the vector  $\mathbf{m}$  and  $\sigma$  are assumed to be equal, which are denoted in scalar format  $m$  and  $\sigma$  in the table 5.1. While  $D_i^1, L_i^1, D_j^2, L_j^2$  stand for the sills and ranges of different correlation models for the two underlying normal variables used in the derivation of the maximum normal copula.

## 5.2 Stability and Reliability Test

It is objective to determine how many points should be included in one subset. However, if the developed parameterization method is stable, similar results from different ways of domain delineation are expected. Moreover, if the method is reliable, the parameterized model should be able to reproduce the spatial structure of the dataset under study. In view of these two aspects, a test is conducted to check the stability and reliability of the method.

In this test, a spatial field is simulated using the v-transformed normal copula with predefined parameter values. This spatial field is used as a synthetic dataset. The domain of the spatial field is divided first into subsets containing 3 points, then into subsets with 6 points and finally with 10 points. The parameterization of a v-transformed normal copula using the proposed approach is done for all the three cases. Figure 5.3 shows the contour plots of

Gaussian copula	$D_i$	$L_i$				
V-transformed normal copula	$m$	$k$	$\alpha$	$D_i$	$L_i$	
Maximum normal copula	$m$	$\sigma$	$D_i^1$	$L_i^1$	$D_j^2$	$L_j^2$

Table 5.1: Parameters to be estimated for the three theoretical copulas



likelihood against the parameters  $m$  and  $k$ . It can be seen that the shape of the contours are similar and the peaks are located at about the same region.

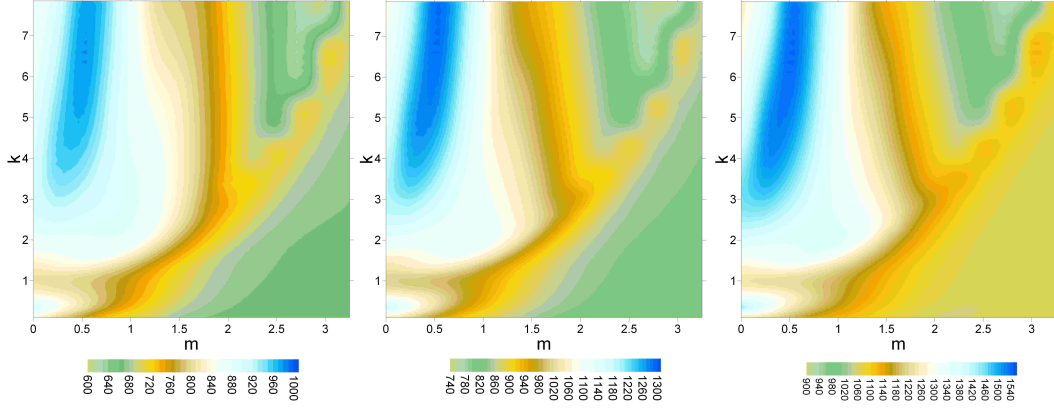


Figure 5.3: Contour plots of likelihoods for parameter estimation of a simulated field. Right panel corresponds to the result when the domain is divided into subsets containing 3 nodes, middle one is for subsets of 6 nodes while the left one is for subsets of 10 nodes.

Table 5.2 shows the predefined parameter values used for the simulation and the parameterization results for all the three cases.

Table 5.2 indicates that the predictions on  $m$  and the sills are quite precise. There are some discrepancies in the  $k$  and the range. But one should notice that it is just one realization of the copula model and therefore fluctuations can occur.

In a word, this test shows that the proposed parameterization method is able to generate similar optimized parameters for different domain delineations indicating its stability. On the other hand, the optimized parameters are close to the predefined parameters of the copula model used for the simulation of the synthetic dataset under study. This demonstrates the reliability of the approach.

	$m$	$k$	Sill	Range
Parameters used in Simulation	0.5	6.0	0.99	930
Result of subsets with 3 points	0.5	6.6	0.97	750
Result of subsets with 6 points	0.5	7.1	0.95	705
Result of subsets with 10 points	0.5	6.6	0.97	725

Table 5.2: Parameterization results of one realization of a  $v$ -transformed normal copula compared with the pre-defined values used in the simulation.

## 6 Interpolation using Copulas

Due to the limitation in observation number, a common problem is to estimate the value of the random function  $Z$  at locations  $\mathbf{s}$  without observations. In the frame work of point kriging, a linear combination of the values of the regionalized variable at sampled locations is to be found as the estimator at the interpolation location:

$$Z^*(\mathbf{s}) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) \quad (6.1)$$

where  $n$  denotes the observations used for interpolation. Because of the 'screening effect' [7], normally only the sampled points which are close to the interpolation points are used. In point kriging, the kriging weights  $\lambda_i$  are searched by minimizing the estimation variance:

$$\begin{aligned} \sigma^2(\mathbf{s}) &= \text{Var} \left[ Z(\mathbf{s}) - Z^*(\mathbf{s}) \right] = E \left[ \left( Z(\mathbf{s}) - \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) \right)^2 \right] \\ &= - \sum_{j=1}^n \sum_{i=1}^n \lambda_j \lambda_i \gamma(\mathbf{s}_i - \mathbf{s}_j) + 2 \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_i - \mathbf{s}) \end{aligned} \quad (6.2)$$

under the unbiasedness condition:

$$\sum_{i=1}^n \lambda_i = 1 \quad (6.3)$$

More details about point kriging and traditional geostatistical tools can be found in *Journal and Huijbregts* [1978] [52], *Kitanidis* [1997] [54] and *Webster and Oliver* [2007] [89].

Equation (6.2) implies that the optimization of the estimation weights is only dependent on the chosen variogram model which is usually a function of the separation vector among locations. However the spatial dependence might also vary for different quantiles of the marginal distribution of the random variable. This aspect is not appropriately considered by the standard ordinary kriging.

Moreover, the kriging estimation variance is only a function of the observation density, data geometry and selected variogram model, but not dependent on the the measured values. However, the latter factor can be of great importance in error estimations of predictions of certain natural phenomena. For instance, the estimation uncertainty of extreme events,

for example one hundred year flood, is often higher than that of the normal events such as medium discharge rate. This property is called heteroskedasticity of the residuals [90] [85], which means the variance of the estimation error is not constant, but varies with the estimated values. Figure 6.1 illustrates one example.

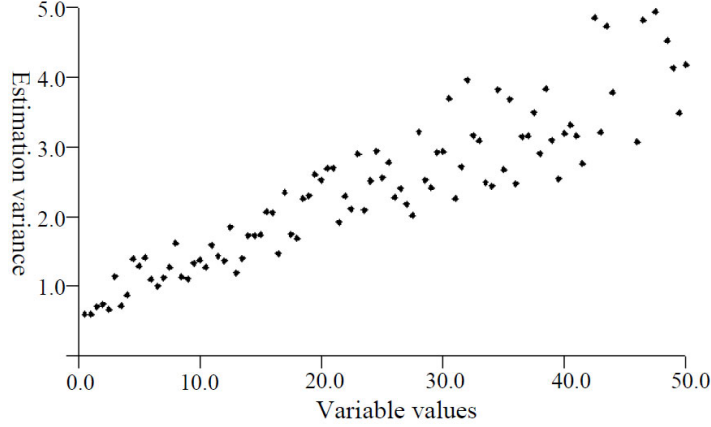


Figure 6.1: *Heteroskedasticity*

The superiority of using copulas over the traditional kriging method is that it describes the spatial dependence not only as a function of spatial configuration of data points, but also as a function of data values. Second, since it allows the estimation of the full conditional distribution of the random variable at the unknown locations, the estimation confidence intervals can be derived which varies with different quantiles of the marginal distribution. These points will be clarified in more details along with the description of interpolation procedure in the following discussions.

## 6.1 Procedure of Interpolation using Copulas

The full conditional distribution of the random variable  $Z$  at the required site  $\mathbf{s}$  conditioned on  $n$  neighbouring observations is:

$$F_n(\mathbf{s}, z) = F_n(Z(\mathbf{s}) < z | Z(\mathbf{s}_1) = z_1, \dots, Z(\mathbf{s}_n) = z_n) \quad (6.4)$$

For a multivariate distribution, this conditional distribution can be written with the help of the corresponding conditional copula  $C_{\mathbf{s},n}$ :

$$F_n(\mathbf{s}, z) = C_{\mathbf{s},n}\left(F_Z(z) | u_1 = F_Z(z(\mathbf{s}_1)), \dots, u_n = F_Z(z(\mathbf{s}_n))\right) \quad (6.5)$$

where  $F_Z$  denotes the marginal distribution and  $z(\mathbf{s}_i)$  refer to the values at observed locations  $\mathbf{s}_i$ .

In order to reduce computational time, the conditional distributions are restricted to local neighborhoods. This restriction does not have a strong influence on the results as demonstrated in Section 6.1.2

The general procedure of interpolation using copulas is as follows:

In the first step, the observed values  $z_i$  are transformed to cumulative distribution values (*cdf* values)  $u_i$ :

$$u_i = F_Z(z_i) \quad (6.6)$$

where  $F_Z$  denotes the empirical marginal distribution of the random function derived from the observations. Since the spatial dependence of the variable under interest is assumed to follow a certain theoretical copula model, in order to calculate the conditional copula density of the unknown point conditioned on the neighbouring measurements in the next step, the *cdf* values have to be transformed into  $z'_i$  by:

$$z'_i = H_{Z'}^{-1}(u_i) = H_{Z'}^{-1}(F_Z(z_i)) \quad (6.7)$$

where  $Z'$  refers to the transformed variable used to generate the applied theoretical copula (see Chapter 4) and  $H_{Z'}$  denotes the corresponding marginal distribution. The theory underpins this step is that the copula among the random variables is invariant under monotonic transformation of the random variables as discussed in Chapter 2. The mapping  $H_{Z'}^{-1}(F_Z(\cdot))$  is strictly increasing, therefore the copula of the transformed random variables  $Z'$  is identical to the copula of the original variables  $Z$ .

In the second step,  $n$  observation points which are closest to the required interpolation point  $\mathbf{x}$  are selected. Then the conditional copula at the interpolation point over the entire range of the values of  $u$  ( $u \in [0, 1]$ ) conditioned on the  $n$  observations will be calculated. The conditional copula is described in Equation (2.11). However, the calculation of the partial derivative in Equation (2.11) is intricate. The finite difference numerical approach is applied here which involves first computing the conditional copula densities at discrete values in the range  $[0, u]$  spaced by equal intervals. Then the conditional copula at the value of  $u$  is approximated by adding up the products of the interval lengths and the corresponding calculated densities over the range  $[0, u]$ . Suppose that the range of  $u$  is discretized into  $k$  elements, the interval length  $\Delta u$  is  $\frac{1}{k}$ , the conditional copula densities are calculated at values of  $u = \frac{1}{2k} + i \cdot \Delta u$ ,  $i = 0, 1, \dots, k - 1$ . The conditional copula density for a certain value of  $u$  can be calculated by:

$$\begin{aligned} & c_{\mathbf{s},n}(u | u_1 = H_{Z'}(z'_1), \dots, u_n = H_{Z'}(z'_n)) \\ &= \frac{h_{n+1}(z', z'_1, \dots, z'_n)}{h_1(z') \cdot h_1(z'_1) \dots h_1(z'_n)} \cdot \frac{1}{c_n(u_1, u_2, \dots, u_n)} \end{aligned} \quad (6.8)$$

where,  $c_{\mathbf{s},n}$  is the conditional copula density at location  $\mathbf{s}$  conditioned on  $n$  observations,  $h_{n+1}$  denotes the  $n+1$  dimensional joint density of the transformed variable  $Z'$  at the selected

observation points and the interpolation point,  $h_1$  denotes the corresponding marginal density while  $c_n$  indicates the multivariate copula density associated with the  $n$  observations. Since  $c_n$  is constant over the range of  $u$  at one target point and does not influence the estimation results calculated in the following step, the second term of Equation (6.8) can be neglected. Thus the following equation is used instead:

$$c_{s,n}(u|u_1 = H_{Z'}(z'_1), \dots, u_n = H_{Z'}(z'_n)) \propto \frac{h_{n+1}(z', z'_1, \dots, z'_n)}{h_1(z') \cdot h_1(z'_1) \dots h_1(z'_n)} \quad (6.9)$$

It is worth mentioning that if the target location overlaps with an observation point, then the conditional copula will be set to 1 at the  $u$  value corresponding to the *cdf* value of the observation at the target location, and set to 0 for the rest range of  $u$  values. In this way, it is ensured that the interpolation is exact, which means for every observation location the interpolation returns the observed value.

For  $v$ -transformed normal copulas, Equation (4.11) and (4.5) will be inserted into the above equation, while for maximum normal copula, Equation (4.17) and (4.15) should be used.

However, for Gaussian copulas, since a closed form of the copula density is available, the calculation is not necessarily through the distribution density  $h$  of  $Z'$ . It is computationally more efficient to use the following equation instead of Equation (6.9).

$$c_{s,n}(u|u_1 = H_{Z'}(z'_1), \dots, u_n = H_{Z'}(z'_n)) = \frac{1}{|\mathbf{\Gamma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{y}^T(\mathbf{\Gamma}^{-1} - \mathbf{I})\mathbf{y}\right) \quad (6.10)$$

where  $\mathbf{y} = (z', z'_1, \dots, z'_n)$  and  $\mathbf{\Gamma}$  is the correlation matrix of the normal variable.

As mentioned before, after obtaining the conditional copula densities over the whole range of  $u$  with a certain discretization, the conditional copula at a certain value of  $u^*$  can be calculated numerically by:

$$C_{s,n}(u^*|u_1, \dots, u_n) = \sum_{u=0}^{u^*} c_{s,n}(u|u_1, \dots, u_n) \cdot \Delta u \quad (6.11)$$

From equation (6.11), the full distribution of the conditional copula at the required location can be obtained. Thus any quantile or statistics can be drawn from that. Typically, the expected value  $z_a^*$  and the median value  $z_m^*$  are selected as estimators for the unknown location.

The former is calculated from the observed values weighted by the corresponding conditional copula using Equation (6.12). The latter is the observed value corresponding to the 50% conditional copula and is calculated by Equation (6.13).

$$z_a^* = \int_0^1 F_Z^{-1}(u) c_{s,n}(u|u_1 = H_{Z'}(z'_1), \dots, u_n = H_{Z'}(z'_n)) du \quad (6.12)$$

$$z_m^* = F_Z^{-1}(u = C_{s,n}^{-1}(0.5)) \quad (6.13)$$

Notice that from the calculated conditional copula, the estimation confidence intervals can also be obtained. For example the 80% confidence interval is calculated as:

$$F_Z^{-1}(u = C_{s,n}^{-1}(0.9)) - F_Z^{-1}(u = C_{s,n}^{-1}(0.1)) \quad (6.14)$$

In this procedure, the spatial correlation among the observation and interpolation points is calculated as pure dependence using the spatial copulas irrespective of the marginal distribution of the observed values. Since the transformation made in the first step is rank preserving, the underlying copula of the variables remains the same. The dependence of the original variables is modeled by the assumed theoretical copula. The calculation should be done via the transformed variables because they are required in the formula to calculate the conditional copula densities. In the last step, the original marginal distribution is reproduced by its substitution into Equation (6.12) or Equation (6.13) when the interpolators are calculated. During this final substitution, the ranks are again preserved and therefore the spatial copulas are untouched.

### 6.1.1 Example of interpolation using different copula models

It must be noted that unlike the traditional Kriging methods, the way in which the dependence varies across the distribution plays a significant role in the estimation using copulas. The following hypothetical examples are used to illuminate this point.

In this example, the value of a target point is going to be estimated by the four surrounding points using two different sets of synthetic observed values at the four corner points. Figure 6.2 shows the configuration of the test field. The input values, i.e. the synthetic observed values, of the conditioning points are given as *cdf* values, which are listed near the conditioning points in Figure 6.2.

Notice that the configuration is fully symmetrical with respect to the interpolation location. In addition, for both cases, It is easy to conclude that Kriging will yield equal weights for all the observation points for both cases. In addition, since the averaged values are the same for both input sets, the estimators are also the same, namely 0.5. But what does the copula method produce? To investigate this problem, a normal copula and a v-transformed normal copula are used to interpolate the value at the target point.

First, the conditional copula density at the target point  $s$ , are estimated by using the two copula models and can be compared from Figure 6.3. The dashed line corresponds to the first input set, while the solid line corresponds to the second one. The expression of the estimated density function  $f_U$  is shown in Equation (6.15).

$$f_U(u) = c_{s,n}(u|U_1 = u_1, \dots, U_n = u_n) \quad (6.15)$$



Figure 6.2: Configuration of the test field. The dots are the observation locations, while the cross is interpolation location. The values listed near the dots are the cdf values of the observations (right panel for the first synthetic observation set and left panel for the second one).

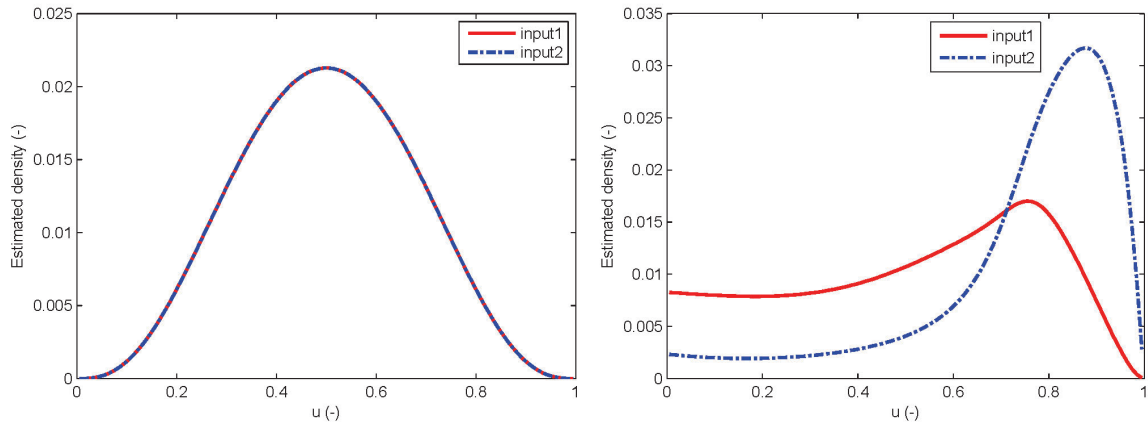


Figure 6.3: Estimation distribution density of the cdf values at the target point using a Gaussian copula (left panel) and a  $v$ -transformed normal copula (right panel).

where  $U$  denotes the uniform variable and  $c_{s,n}$  denote the conditional copula density at location  $s$  conditioned on  $n$  observations. In this example  $n = 4$ .

From the left panel of Figure 6.3, one can see that in case that a Gaussian copula is applied, the resulting curves for both cases are symmetric and overlap with each other. This is due to the following three reasons. Firstly, the configuration of the test field is symmetrical. Secondly, for both input sets, each location with a big value is opposite to the location with a small value as the counterpart (the values of the couple are reflections with respect to the median 0.5). Last but not the least the dependence structure represented by a Gaussian copula is symmetrical. Therefore the influences from the big and the small values balance out and compromise to produce a symmetrical estimation distribution and the median 0.5 as the most probable estimator.

On the contrary, when a  $v$ -transformed normal copula is used, the difference between the estimated conditional density curves of the two cases is obvious. This can be explained by the asymmetrical dependence structure represented by the  $v$ -transformed normal copula, which in this test has stronger dependence strength for extreme high quantiles than for the extreme low quantiles. This means the big values have a stronger influence and therefore the conditional density curves are no more symmetrical in spite of the symmetrical configuration and symmetrical values. Moreover, it can be observed that the distributions are more skewed to the upper tail for the second case. This is because that the two big values given in the second case approach the extreme ends where the dependence strength gets stronger.

Next, the estimated quantile values are transformed to the hypothetical observed values. In other words, the uniform variable  $U$  is transformed to the random variable  $Z$  which is assigned to the synthetic dataset and assumed to follow a certain marginal distribution. Two marginal distributions, namely, a normal distribution ( $N(1,1)$ ) and an exponential distribution ( $E(1)$ ), are used for both input sets in order to elucidate the influence of the marginal distributions on the interpolation results. Figure 6.4 and 6.5 show the conditioning values after imposing the marginal distributions.



Figure 6.4: Conditioning values after imposing a normal marginal distribution ( $N(1,1)$ ).



Figure 6.5: Conditioning values after imposing an exponential marginal distribution ( $E(1)$ ).

After imposing the marginal distributions, the estimation distribution density functions of real values are compared again between the two input sets using the two copula models. Equation (6.16) and (6.17) describe the calculation of the estimation distribution density function after imposing the marginal distribution.



$$\begin{aligned}
F_Z(z) &= P(Z < z) = P(G^{-1}(u) < z) \\
&= P(u < G(z)) \\
&= F_U(G(z))
\end{aligned} \tag{6.16}$$

where  $G$  denote the cumulative marginal distribution, therefore  $u = G(z)$  and  $z = G^{-1}(u)$ . While  $F_Z$  and  $F_U$  denote the cumulative estimation distribution function of  $Z$  and  $U$ .

$$\begin{aligned}
f_Z(z) &= \frac{dF_Z(z)}{dz} = \frac{dF_U(G(z))}{dz} \\
&= f_U(u) \cdot g(z) \\
&= c_{s,n}(u|U_1 = u_1, \dots, U_n = u_n) \cdot g(z)
\end{aligned} \tag{6.17}$$

where  $f_Z$  denotes the estimation density function of  $Z$  and  $g$  denotes the marginal distribution density.

Figure 6.6 shows the estimation distribution densities after imposing the marginal distributions onto the original input sets of *cdf* values. The left panel in Figure 6.6 shows the estimated density curves of  $Z$  at the target location obtained from the Gaussian copula, while the right panel shows those obtained from  $v$ -transformed normal copula. Different colors of the curves indicate different original input sets, while different line styles indicate different marginal distributions.

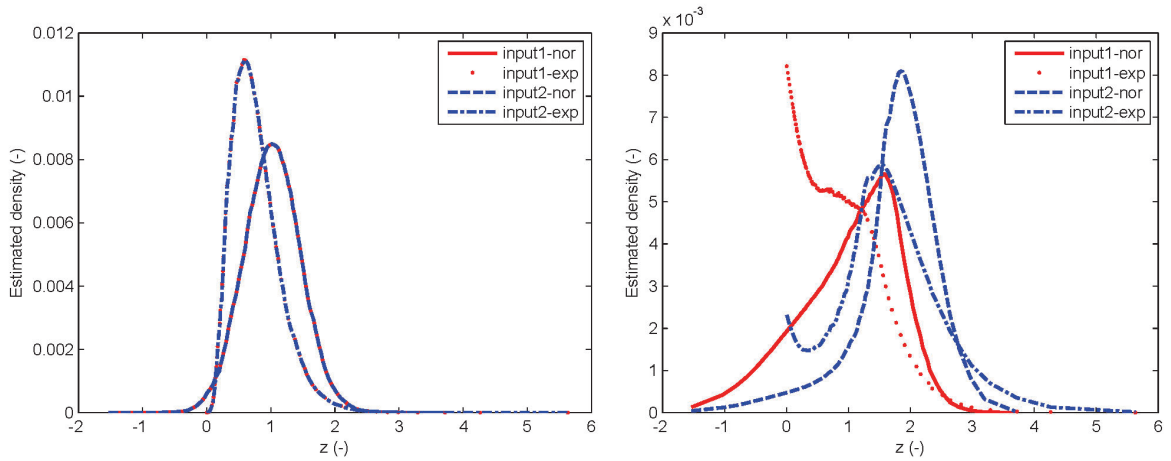


Figure 6.6: Estimation distribution density of the *cdf* values at the target point using a Gaussian copula (left panel) and a  $v$ -transformed normal copula (right panel).

From these estimation distributions, the estimators of any statistics (the median, the mode, etc.) of the variable  $Z$ , can be easily derived. The left panel of Figure 6.6 shows that for

Gaussian copula, the density curves peel away from each other for different marginal distributions which are superimposed onto the same original input *cdf* set, but are glued together for different original input *cdf* sets if the marginal distributions are the same. This is due to the symmetrical dependence structure described by a Gaussian copula. Particularly, when a normal marginal distribution is inserted, the estimation distribution density is symmetrical with respect to the mean. In comparison, for *v*-transformed normal copula, all the resulting curves differ from each other.

Taken together, the interpolation using copulas are influenced not only by the configuration, but also by the marginal distribution and dependence structure which varies across the whole distribution. By considering more facets and providing the full estimation distribution instead of a single estimator and estimation variance, copulas provide a more judicious and informative way of interpolation than traditional Kriging method.

### 6.1.2 Neighborhood selection

As mentioned in Section 6.1, in order to reduce the computational load, the conditioning data points should be restricted to the neighbourhood of the target location. Because the calculation of copula density, especially for the two non-Gaussian models, requires a summation of  $2^n$  terms, if the whole dataset is used for conditioning of a single target location, the value of  $n$  in equation (6.8) can be prohibiting, which makes the calculation very computationally demanding. Moreover, similar as in the Kriging procedure, too many conditioning points can be redundant due to the so-called 'screening effect'. On the other hand, sufficient number of observations should be kept for conditioning in order to provide a reasonable estimator.

The following test is carried out to find the appropriate number of conditioning points which are in the neighborhood of the target location. To this end, the dataset of chloride concentration in the groundwater of Baden-Württemberg is used. An unsampled location inside the measurement network is selected randomly. Copula based interpolations are done at the target point for several times with different number of neighboring conditioning observations. Figure 6.7 shows the resultant estimated distribution density curves for *cdf* values (left panel) and real values (right panel). One can see that as number of observations ( $n$ ) increases, the density curves become more and more similar. For  $n > 9$ , the estimated density curves are almost identical. Therefore in the following application, the value of  $n = 12$  is used for the interpolation of the groundwater quality parameters in Baden-Württemberg.

### 6.1.3 Properties of the copula based interpolation

To summarize the properties of copula based interpolation method, the following terms can be mentioned. The first three are similar to those of Kriging. The last two are particular for the copula based method and also advantageous over the traditional Kriging approach.

1. The interpolation is exact — for every observation the interpolation returns the observed value.

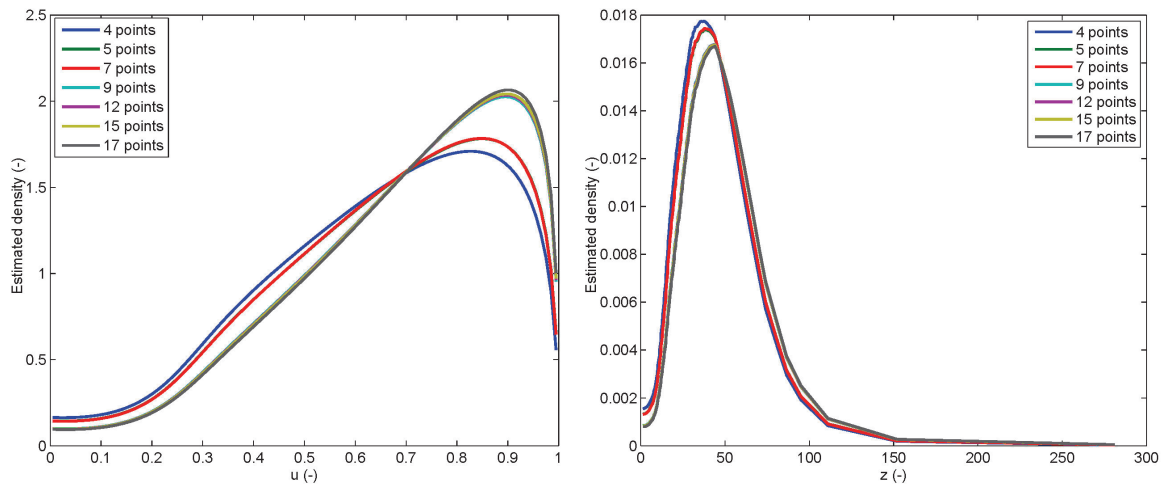


Figure 6.7: Estimated distribution density of the cdf values (left panel) and the real values (right panel) at the selected interpolation point in the studied domain using different number of neighboring observations for interpolation.

2. The interpolation depends on the configuration of the observations and the target point.
3. The conditional distribution depends mainly on the observations near to the target point. Distant observations have minor effect on the calculated conditional distribution.
4. The interpolation delivers full conditional distributions, thus confidence intervals can be identified.
5. The interpolator depends not only on the configuration of the observations, but also on the observed values and the marginal distribution.

## 6.2 Application

The main content of this section is taken from the published paper 'Geostatistical Interpolation using Copulas' (Bárdossy and Li [2008]) [11] on *Water Resources Research*. Some extensions and improvements are done.

An extensive dataset consisting of more than 2500 measurements of groundwater quality parameters of the near surface groundwater layer in Baden-Württemberg were used to illustrate the methodology. Five quality parameters, namely, chloride, nitrate, pH, sulfate and dissolved oxygen were selected for this study. A geostatistical investigation of the data was

carried out and reported in [10] and [9]. Figure 6.8 shows the groundwater quality observation network. Note that the green triangles indicate the locations for a split sampling for the crossvalidation purpose (see Section 6.2.5). Table 6.1 shows the basic statistics for the five selected groundwater quality parameters. Note the high skewness of the parameters of chloride and sulfate.

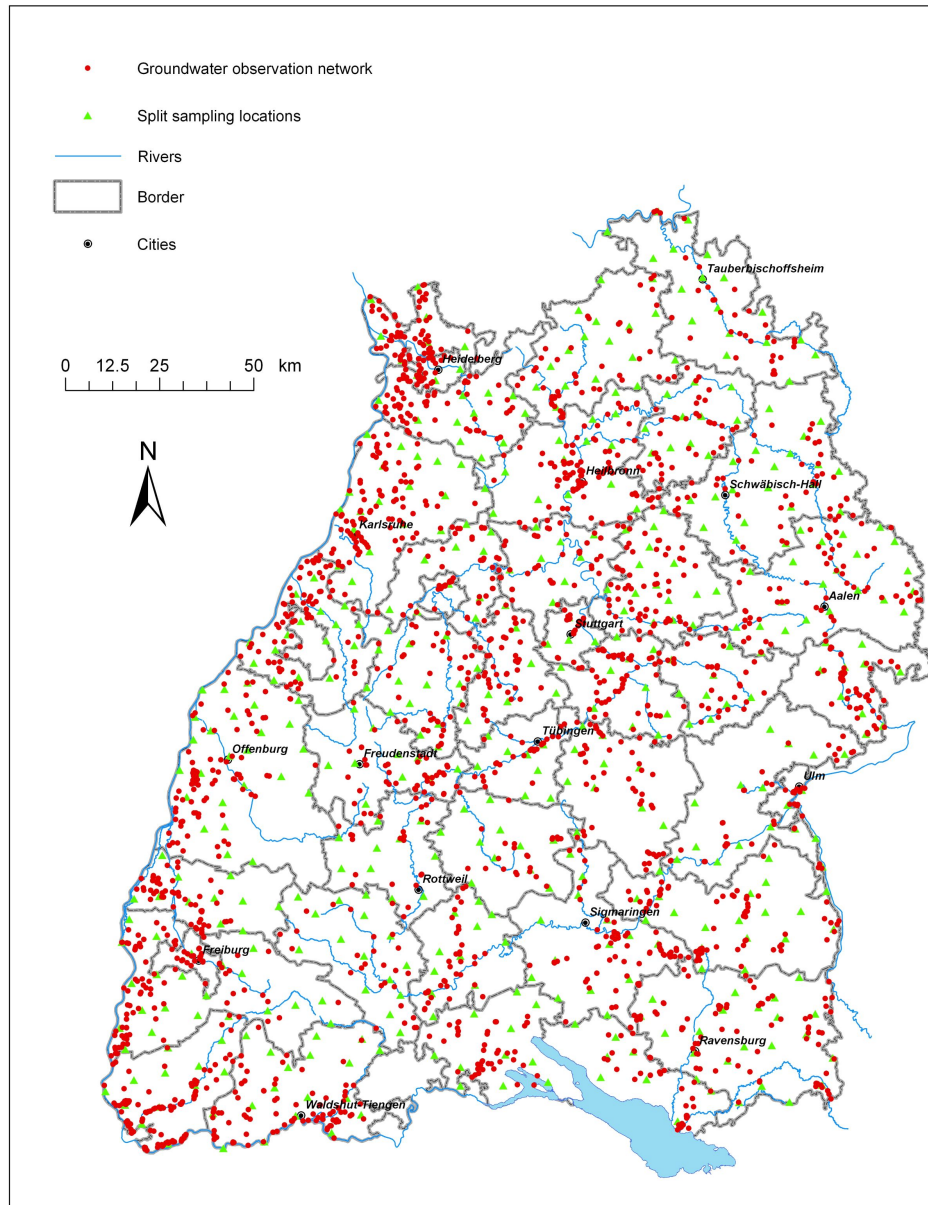


Figure 6.8: Location of the domain under study and the observation wells.

	pH	Sulfate	Nitrate	Chloride	Dissolved oxygen
Mean	7.33	70.11	31.76	33.15	7.98
Median	7.34	36.75	25.00	24.00	8.40
Standard deviation	0.41	126.18	27.79	49.62	3.04
Skewness	2.25	6.73	2.32	13.10	-0.53

Table 6.1: *Statistics of the five groundwater quality parameters*

### 6.2.1 Empirical copulas

Empirical copulas were calculated for all groundwater quality parameters for a set of different separation distances. Since no marked anisotropy was observed, thus only the isotropic case was dealt with, the separation length was used instead of the separation vector. Figure 6.9 shows the empirical copulas of the five parameters for the separation lengths of 3 km (left), 6 km and 9 km. This figure indicates that all five parameters exhibit non-symmetrical, non-Gaussian dependence.

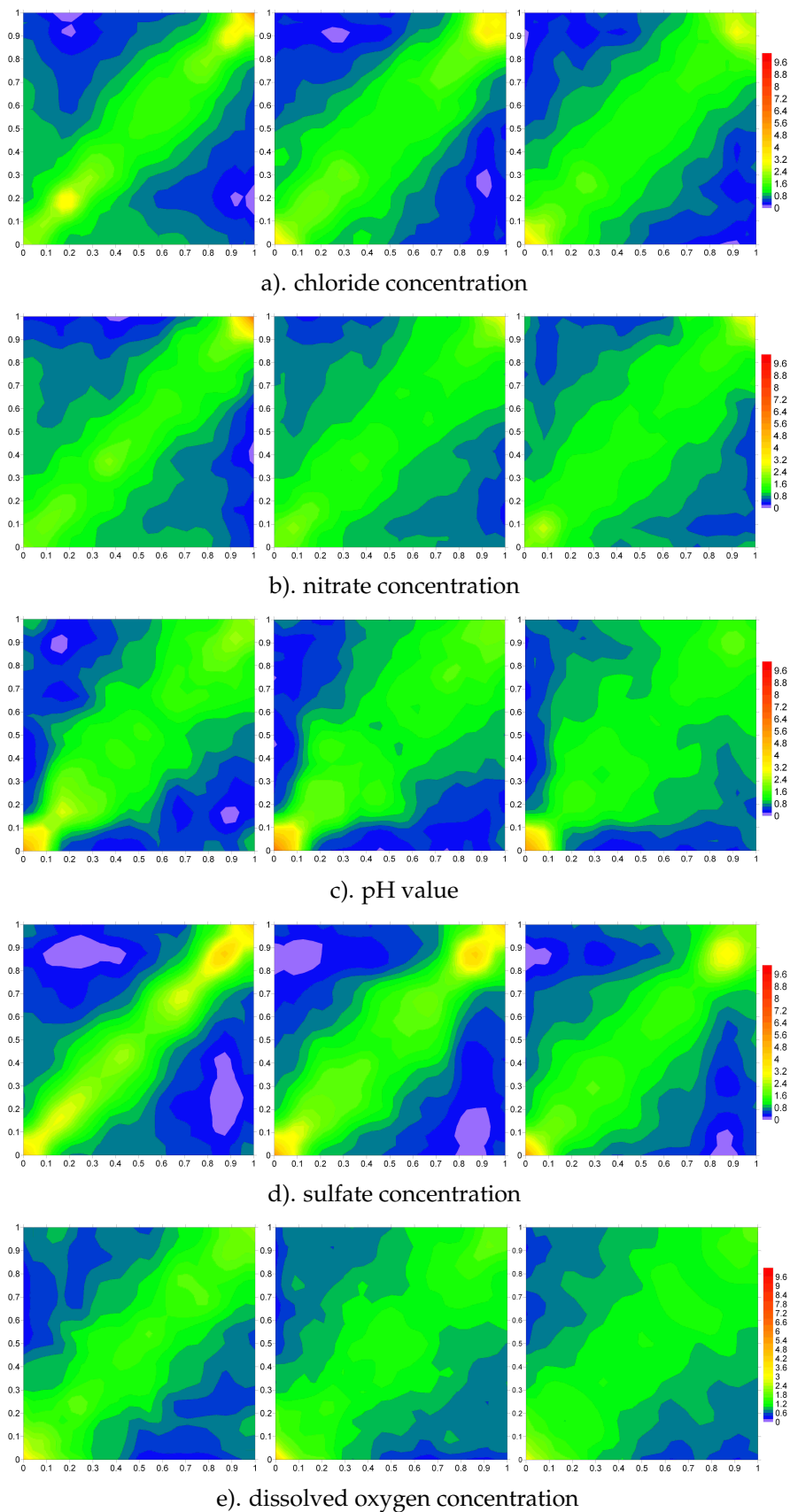


Figure 6.9: Empirical copulas of the five groundwater quality parameters.

### 6.2.2 Parameter estimation and statistical test

Note that the empirical bivariate marginal copulas were not used for the parameter estimation. The empirical and theoretical copulas cannot directly be used to judge the goodness of fit, as the empirical copulas were estimated from a highly dependent sample.

Parameter estimation was carried out using the partition of the observation set into subsets  $S_k$  and a subsequent maximization of the likelihood. Partition sets with a different number of observations were selected. The diameter of the sets was kept small as for interpolation only neighboring observations are considered. The likelihood function described in Equation 5.5 in Section 5.1 corresponding to the selected partition was numerically maximized using the optimization algorithm GRG2 [58] [59].

As indicated by Figure 6.9 that the dependence of all five groundwater quality parameters are non-symmetric and non-Gaussian. The v-transformed normal copula is parameterized for the five groundwater quality parameters.

Seven model parameters  $m, k, D_0, D_1, D_2, L_1, L_2$  were optimized. Here  $m$  and  $k$  are parameters of the v-shaped transformation described in Section 4.2.  $D_1$  and  $D_2$  indicate the contributions of the two nested spherical models, while  $L_1$  and  $L_2$  indicate their ranges. The flipped signed  $m$  combined with the reciprocal of  $k$  can result in the same type of copula with the original  $m$  and  $k$  values, therefore only  $m \geq 0$  is considered in the optimization. The correlation length  $L_0 = 0$ , which means that  $D_0$  represents the so called nugget effect. Results show that the parameters depend only very slightly on the selection of  $S_k$ . Figure 6.10 shows the logarithm of the likelihood function for nitrate as a function of  $m$  and  $k$  (optimized for the remaining five parameters) for two different partitions of the observation domain.

Note the similar structure of the likelihood contours. The maximum is located in both cases in the same region. The lower likelihoods corresponding to values  $m > 3$  indicate that the v-transformed normal copula fits the observed data better than the normal. Similar behavior was observed for all other groundwater quality parameters too.

The model parameters for the v-transformed normal copula are listed in Table 6.2. Notice that  $D_0$  refers to the sill of a nugget effect component, while  $D_1, D_2, L_1$  and  $L_2$  refer to the sills and ranges of two spherical components of the complex correlation model for  $R(h)$ . A general form of the complex correlation model is described in Equation (5.8) in Section 5.1. The asymmetry of the dependence is denoted as negative if the multivariate copula had to be turned upside down as described in Equation (4.13).

Based on the parameterized copula models, statistical tests using Monte Carlo approach are carried out for the rank correlation structure and the asymmetry over distance structure. One hundred unconditional realizations on the observation locations of the groundwater quality parameters are drawn from the parameterized copula models using the sequential conditional simulation approach (see Chapter 7). Rank correlation and measure of asymmetry (see Equation (3.9)) are calculated for several separation lengths of the realizations. Figure 6.11 and 6.12 show the upper and lower 10% boundaries and the average of the rank correlations of the Gaussian realizations and v-transformed normal copula realizations as

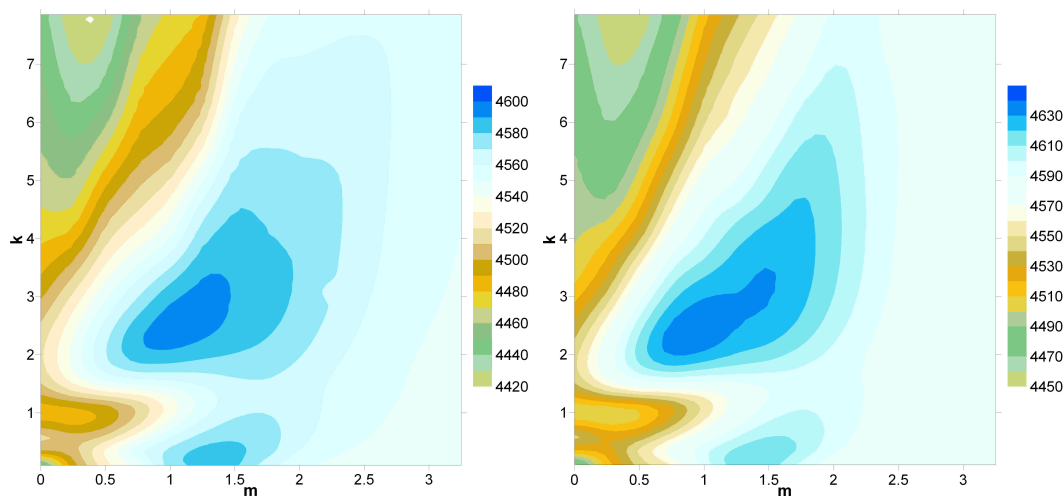


Figure 6.10: Contour plots of log likelihood values against the  $v$ -transformed normal copula parameters  $m$  and  $k$  for nitrate concentration. The left panel is the result corresponding to the partition of the domain into subsets containing 6 points, while the right one corresponds to the partition of the domain into subsets containing 11 points.

	Chloride	Nitrate	pH	Sulfate	dissolved oxygen
$m$	2.7	1.0	0.25	1.0	1.25
$k$	0.10	2.35	2.35	2.1	0.35
$D_0$	0.18	0.05	0.18	0.03	0.3
$D_1$	0.28	0.23	0.11	0.05	0.32
$D_2$	0.54	0.72	0.71	0.092	0.38
$L_1$ [m]	3000	3000	7300	3000	3000
$L_2$ [m]	75000	73000	69500	68200	47500
asymetry	positive	positive	negative	negative	negative

Table 6.2: Estimated model parameters for  $v$ -transformed normal copula

well as the empirical curves from the observations. From the figure, one can see that both models reproduces the rank correlation structure reasonably, although it is not expressed explicitly in the objective function of the parameterization.

Figure 6.13 and 6.14 shows the upper and lower 10% boundaries and average of asymmetry of the Gaussian realizations and  $v$ -transformed normal copula realizations as well as the empirical curves from the observations.

From Figure 6.13 and 6.14, the statistical boundaries and expected behaviour of the 100 realizations show that the  $v$ -copula is the more suitable model for the spatial dependence of most parameters. This is, however, not true for the chloride concentration, for which the performances of Gaussian and the  $v$ -copula are almost the same. That is because the



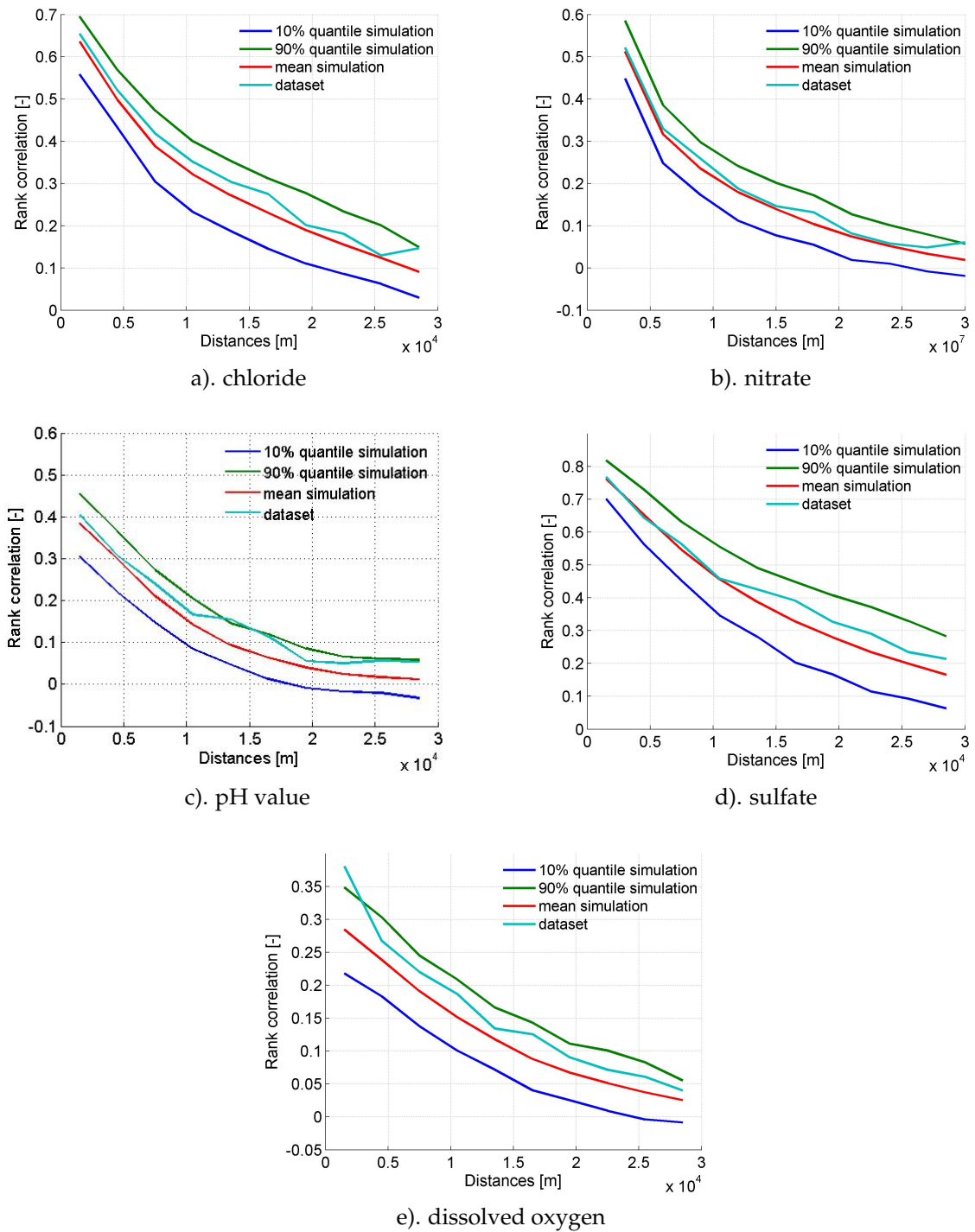


Figure 6.11: Statistical boundaries of the rank correlation over distance structures of the Gaussian realizations compared with the empirical ones.

parameterized  $v$ -copula for the chloride concentration gets a relatively high  $m$  value, namely 2.7. As mentioned in Section 4.2 that if the value of  $m$  is increased to 3, then the resultant

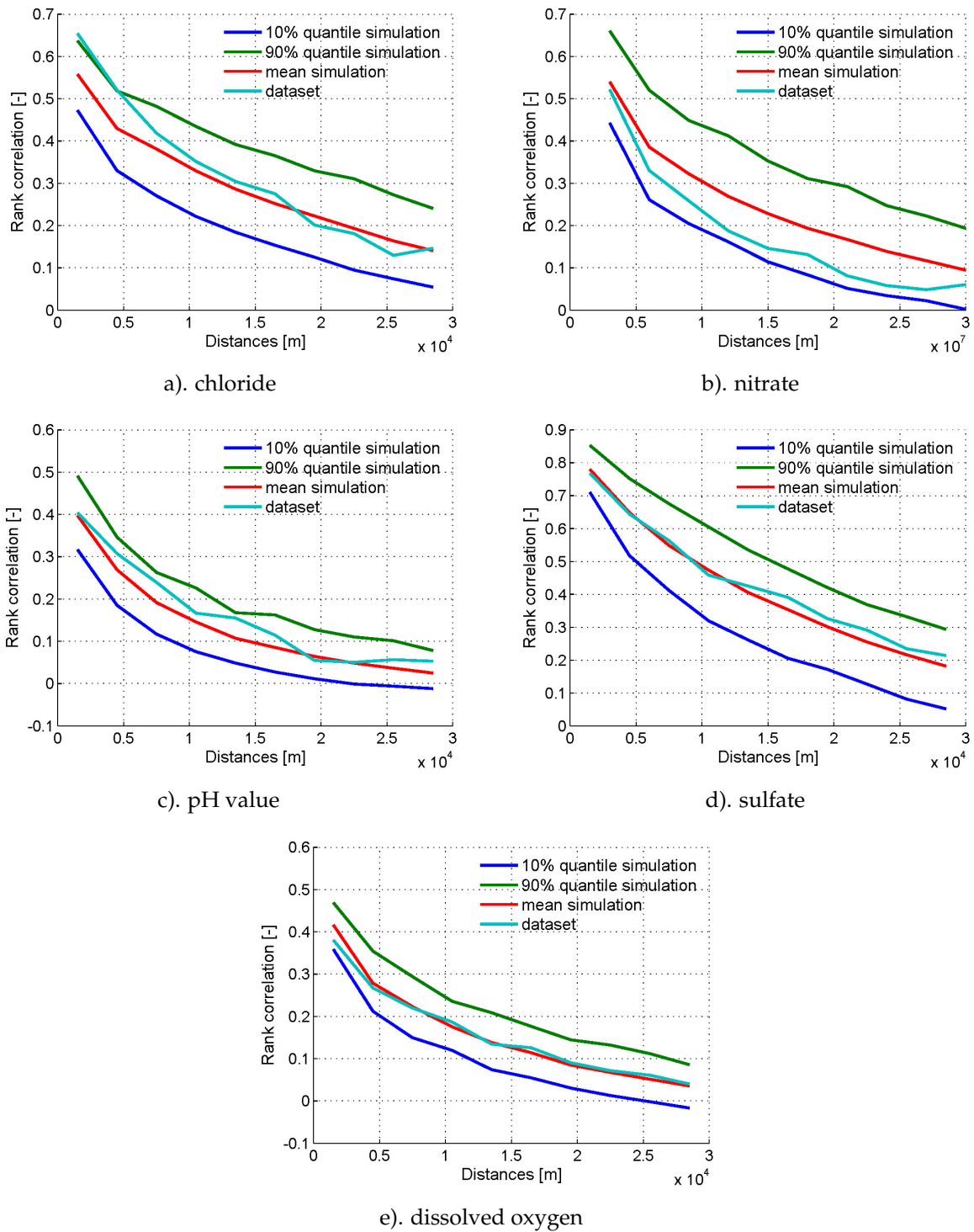


Figure 6.12: Statistical boundaries of the rank correlation over distance structures of the  $v$ -transformed normal copula realizations compared with the empirical ones.

$v$ -copula is already approaching a Gaussian copula.

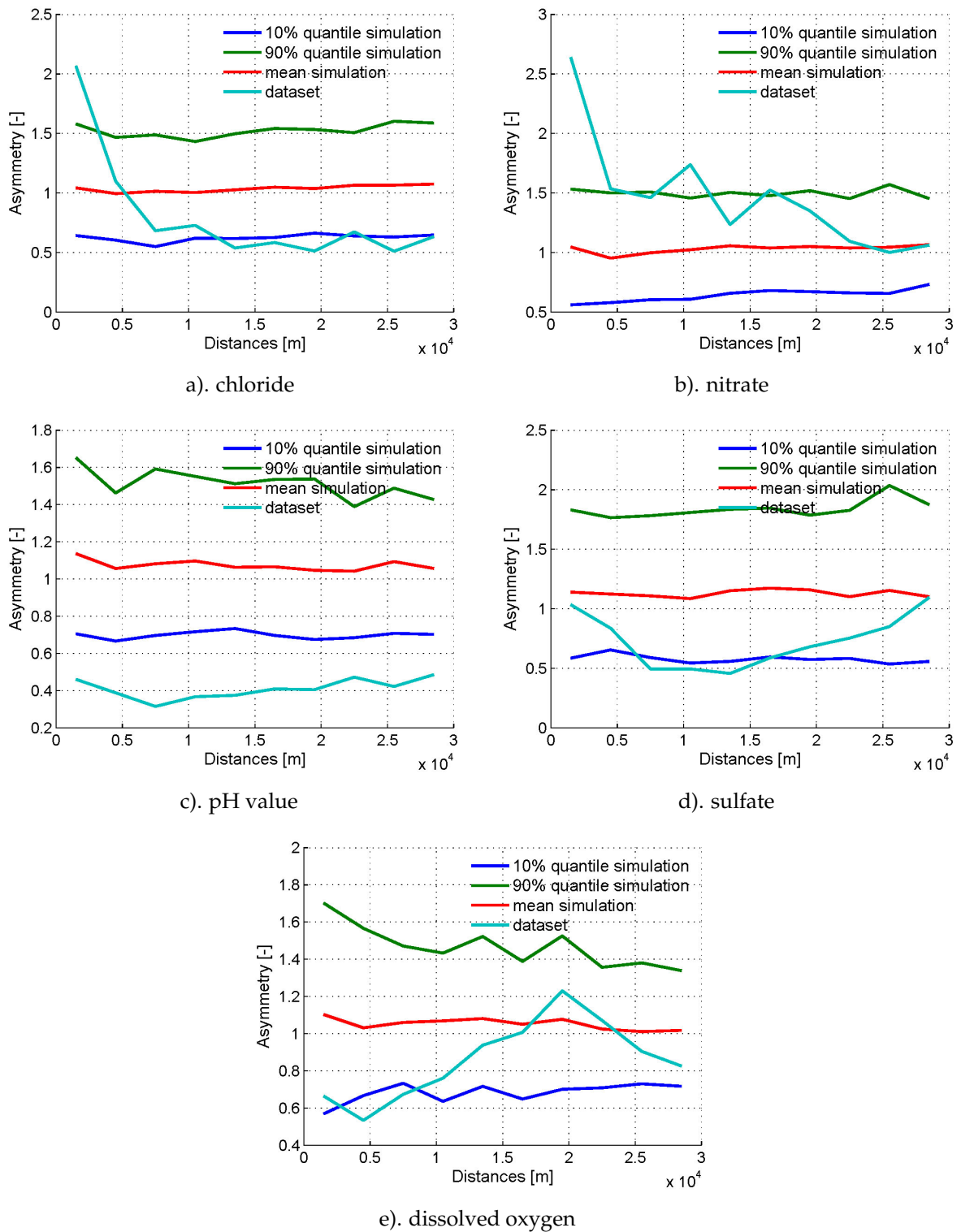


Figure 6.13: Statistical boundaries of asymmetry over distance structures of the Gaussian realizations compared with the empirical ones.

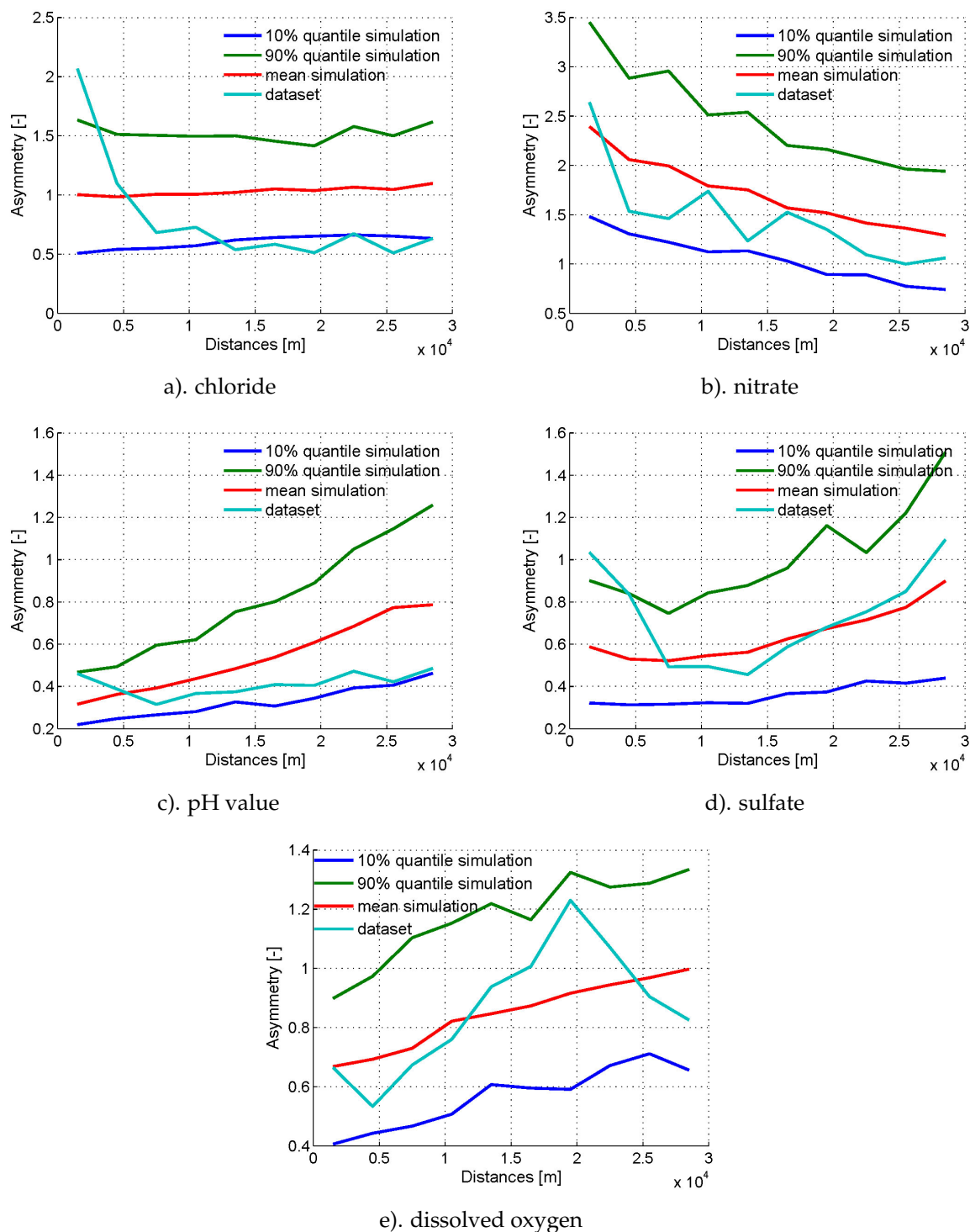


Figure 6.14: Statistical boundaries of asymmetry over distance structures of the  $v$ -transformed normal copula realizations compared with the empirical ones.

### 6.2.3 Interpolation

Interpolation was carried out on a regular grid of  $300 \text{ m} \times 300 \text{ m}$  resolution for the whole state of Baden-Württemberg. Four different interpolation methods were applied:

1. Interpolation with multivariate  $v$ -transformed normal copula function
2. Interpolation with multivariate Gaussian Copula function
3. Ordinary-Kriging
4. Indicator-Kriging: Note that different variograms were used at the different thresholds  $\beta$ . The relationship between indicator variograms corresponding to a cutoff  $\beta$  and the bivariate marginal copulas

$$\gamma_{\beta}(\mathbf{h}) = F_Z(\beta) - C_S(\mathbf{h}, F_Z(\beta), F_Z(\beta)) \quad (6.18)$$

derived in [8] is used for this purpose. Here  $C_S(\mathbf{h}, u, v)$  denotes the theoretical bivariate marginal copula corresponding to the vector  $\mathbf{h}$ .

Figure 6.15 shows the interpolation maps obtained from Gaussian copula and  $v$ -transformed normal copula for nitrate.

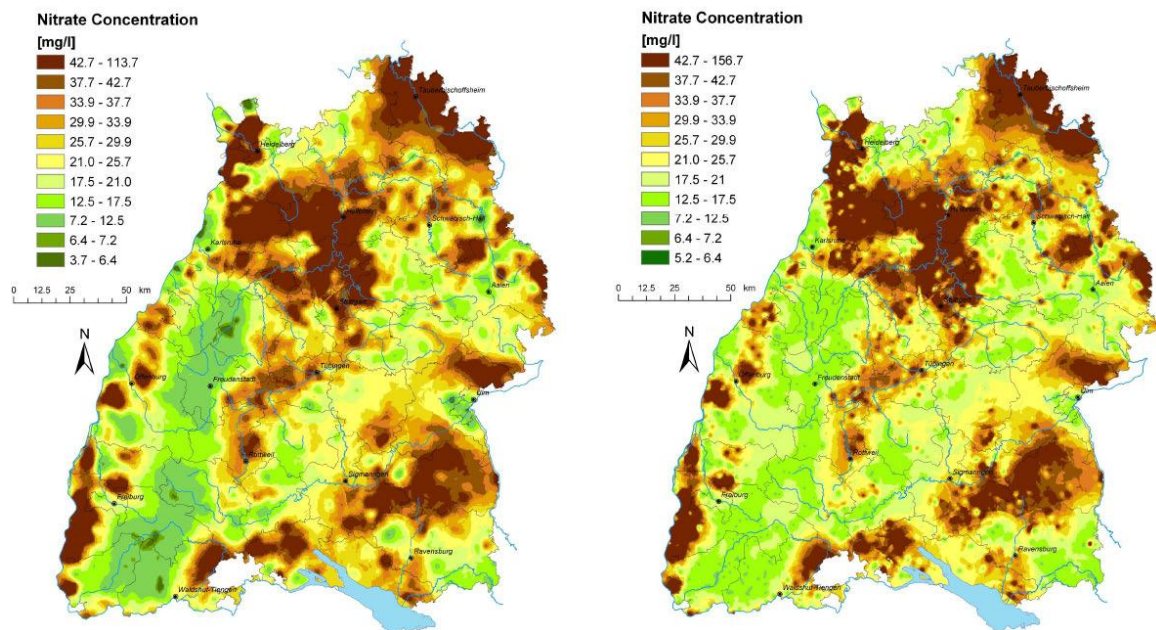


Figure 6.15: Interpolated mean values of nitrate concentration using a Gaussian copula (left) and a  $v$ -transformed normal copula (right).

The interpolated maps are different, most strikingly in the Black Forest region (south-western part) where the  $v$ -copula yields more conservative estimations.



Figure 6.17 and 6.16 show the interpolated mean maps obtained from the  $v$ -transformed normal copula and Ordinary Kriging for sulfate concentration and pH value, respectively.

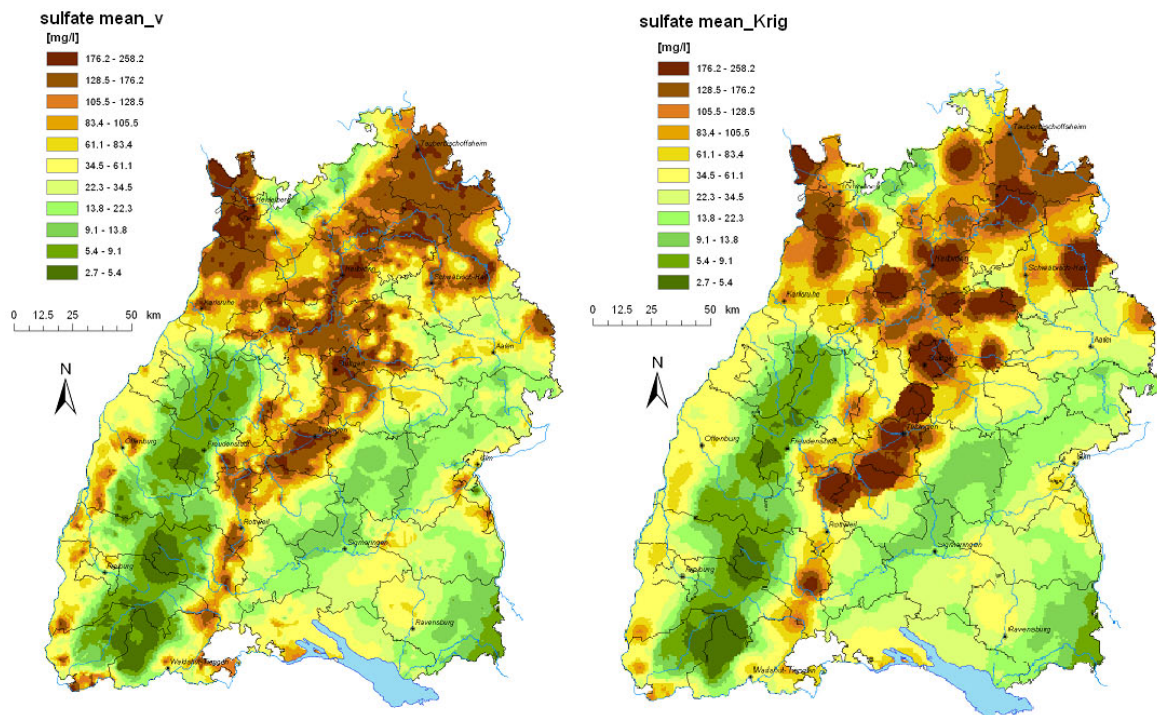


Figure 6.16: *Interpolated means of sulfate concentration obtained from the  $v$ -transformed normal copula (left) and Ordinary Kriging (right).*

Figure 6.16 shows that the high values are more clustered in the kriging map of sulfate, which indicates a stronger dependence than the high values in the  $v$ -copula interpolated map. This finding is in agreement with the parameterization result which is shown in table 6.2. Table 6.2 shows that the parameterized  $v$ -copula for sulfate has a negative asymmetry. That means low values are more dependent than the high values. That is why the  $v$ -copula interpolated map exhibits more scattered high value regions. For the case of pH, from Figure 6.17, the differences can not be easily detected. In order to have a better visible comparison, the maps of difference between the results of the applied interpolation methods were also produced. Figure 6.18 shows the maps of the differences between the  $v$ -transformed normal copula results and Ordinary Kriging for pH and sulfate, respectively.

Comparing the difference maps and the corresponding interpolation maps, it can be seen that the differences are concentrated in areas with high and low values. For example, the zoomed maps for pH shown in Figure 6.19 offer a close examination of the difference in the results of the area around Tübingen. The reason for this is that the  $v$ -transformed copula interpolation treats high and low values differently while in Ordinary Kriging all values are treated the same way (kriging weights do not depend on the values but the locations and the applied variogram model).

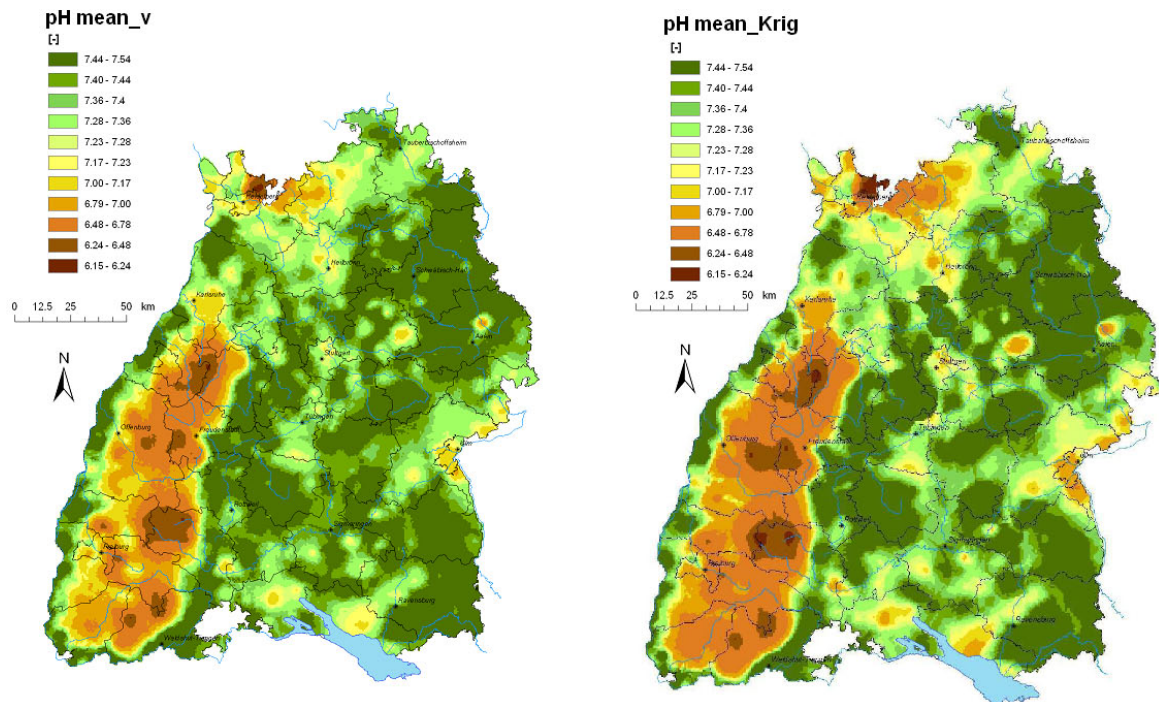


Figure 6.17: Interpolated means of pH obtained from the  $v$ -transformed normal copula (left) and Ordinary Kriging (right).

As mentioned in Section 6.1 that any quantile values can be drawn from the estimated conditional copula, the mean and median are normally considered as appropriated estimators. Figure 6.20 shows the interpolation maps obtained from the  $v$ -transformed normal copula for chloride using the mean and the median estimators.

#### 6.2.4 Confidence intervals

Since interpolation using copulas provides the full estimation distribution of the parameters under interest, confidence intervals of the estimators can easily be assessed from the estimation distribution. For example, in Figure 6.21 the 90% confidence intervals obtained by using the  $v$ -transformed normal copula are shown for the parameters dissolved oxygen concentration and pH value. They were calculated by subtracting the 5%-quantile from the 95%-quantile. It can be observed that near the observations the confidence intervals are narrow while in other regions they become much wider.

The structures of the two maps are very different even though the observation locations are the same. The highest uncertainty of the pH estimates is in the Black Forest area (Western side of the state Baden-Württemberg) where acidification of the groundwater body is a serious problem (see Figure 6.17). In contrast the uncertainty corresponding to dissolved oxygen is low in the same area as it is unproblematic for this parameter, as it is shown in



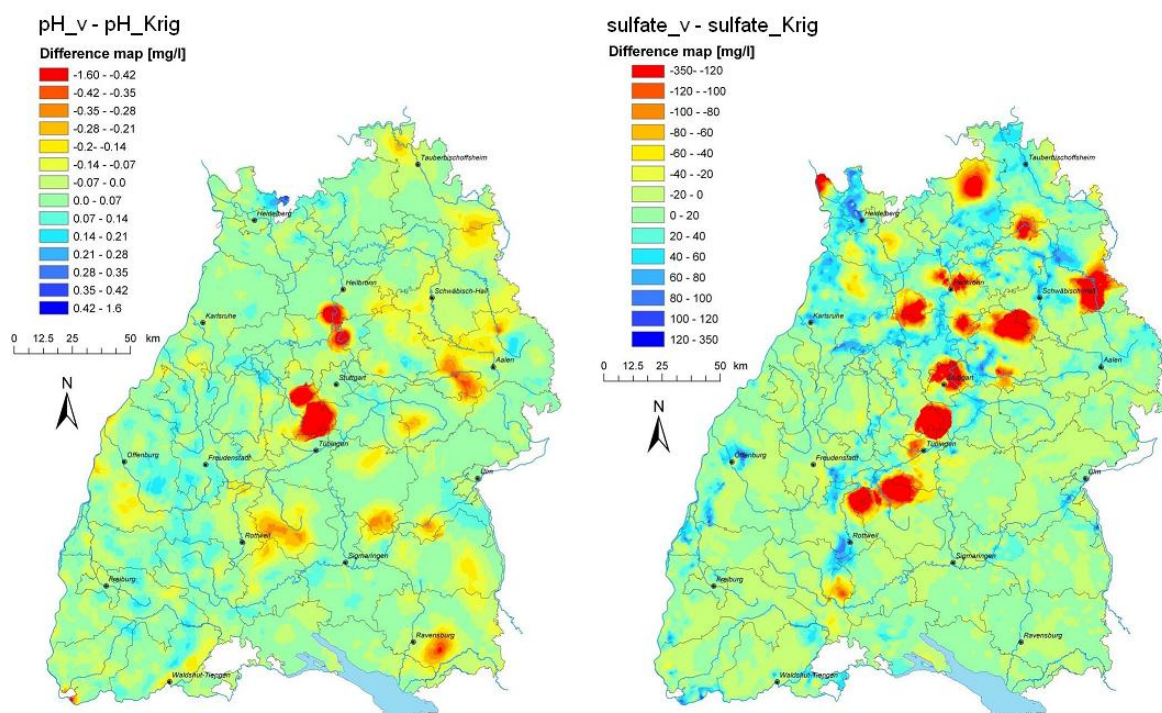


Figure 6.18: Difference maps between the interpolated means obtained from  $v$ -transformed normal copula and Ordinary Kriging for pH (left) and sulfate concentration (right).

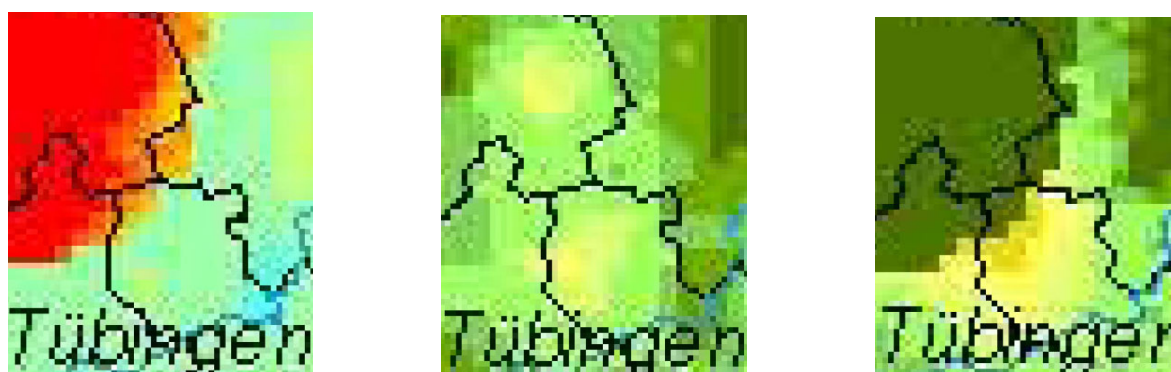


Figure 6.19: Zoomed maps of pH around Tübingen. Left one is the difference map, middle one is the  $v$ -copula interpolated map and right one is the kriging map.

Figure 6.20 that the concentration of dissolved oxygen is quite high in the Black Forest. Figure 6.21 also indicates that along the western boundary, the estimation is uncertain in spite of the dense observation points shown in Figure 6.8. An examination of Figure 6.20 reveals the possible reason behind it, i.e. the concentration of dissolved oxygen is very low which indicates pollution problem and detection limit problem. These facts can cause the high uncertainties in the estimation. In addition, the difference in the confidence interval maps is also a consequence of the very different dependence structures of the two parameters. The



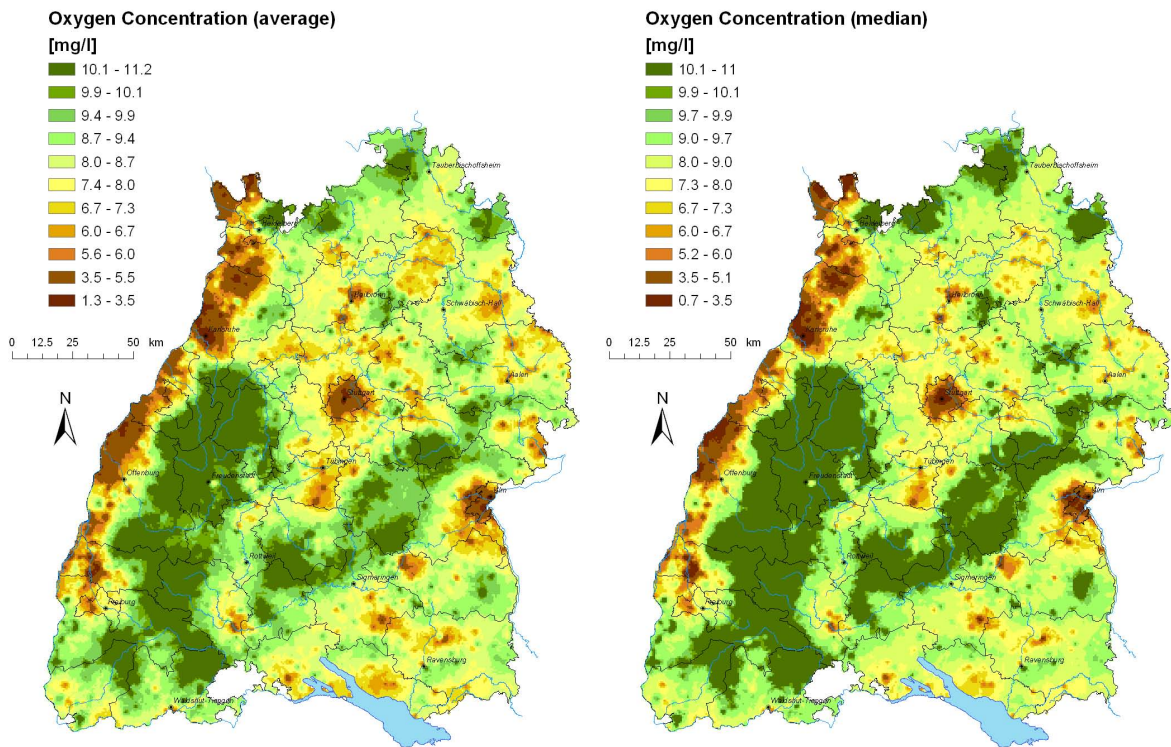


Figure 6.20: Interpolated mean values (left) and median values (right) of dissolved oxygen concentration using a  $v$ -transformed normal copula.

study of confidence intervals demonstrates that the value-dependent dependence structures are well considered in the uncertainty estimation delivered by the copula approach.

On the contrary, the estimation variances calculated using ordinary kriging would lead to very similar patterns, since it only reflects the observation density and applied variogram model.

### 6.2.5 Cross validation and split sampling results

In order to test how well the model fits the observed data, crossvalidation was carried out for the four interpolation methods. For the multivariate  $v$ -transformed normal copula, two estimators, the mean and the median were considered. While for the other methods, the estimated means are considered.

Further a dense uniformly distributed set was selected for a split sampling validation. Figure 6.8 shows the locations of the control set (indicated by the green triangles). Parameter values were estimated at each point of the control set using the rest of the set. The results are analyzed with different criteria. Besides the conventional mean squared and absolute error, two criteria in the probability space were considered:

- MAE (mean absolute error)

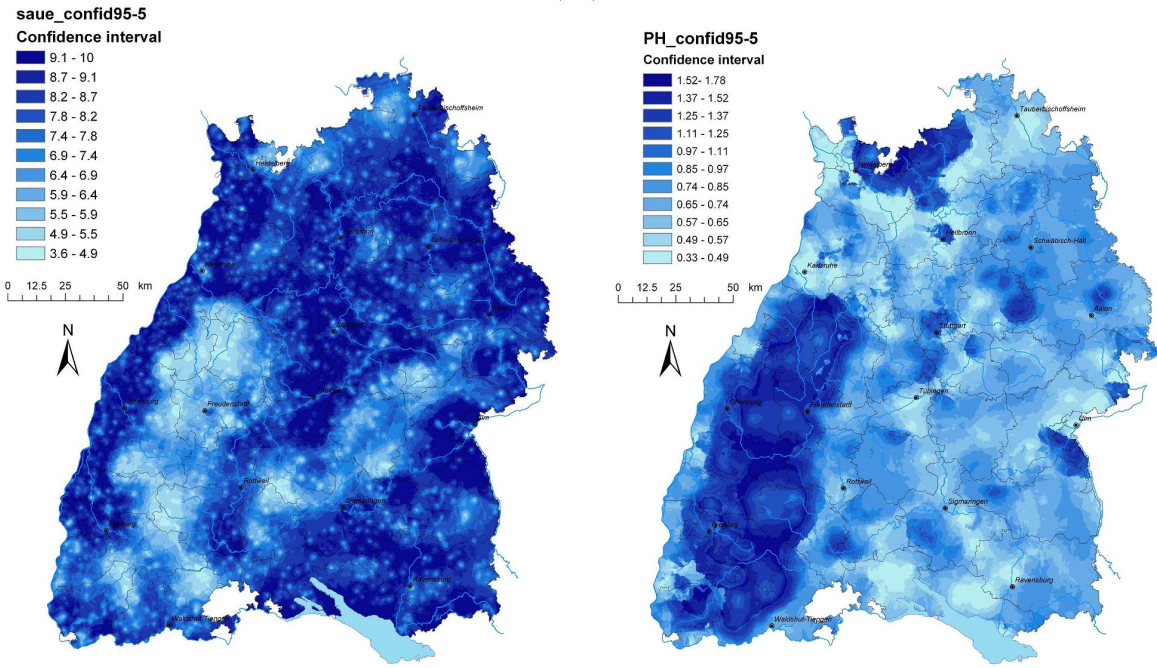


Figure 6.21: 90% confidence intervals of dissolved oxygen concentration (left) and pH value (right) obtained from the  $v$ -transformed normal copula.

- RMSE (square root of the mean squared error)
- LEPS (Linear error in probability space) The quantity  $L$  is the difference between the estimated and the observed values in the probability space [88]:

$$L = \frac{1}{n} \sum_{i=1}^n |F_z(z(\mathbf{x}_i)) - F_z(z^*(\mathbf{x}_i))| \quad (6.19)$$

An advantage of this quantity is that it does not depend on the magnitude of variable values. The possible values of  $L$  range from 0 (perfect) to 1 (improper).

- Missclassification rate: For a selected percentile (85 % for our case) the portion of misclassified points with regard to the exceedence of the threshold was calculated. The reason for taking this measure is that for groundwater quality parameters often the exceedence of a certain threshold is of great interest. For the sake of compareability the same quantile is selected for all parameters.

As one can see in table 6.3 the copula based approaches give better results than ordinary or indicator kriging. The mean of the  $v$ -transformed normal copula is the best estimator in the sense of RMSE while the median of the  $v$ -transformed normal copula is the best for MAE. Results of the cross validation and the split sampling are very similar.

Parameter	Method	Cross Validation				Split Sampling			
		MAE	RMSE	LEPS	CLE	MAE	RMSE	LEPS	CLE
Chloride	V mean	15.182	42.718	0.154	0.093	16.146	40.765	0.182	0.096
	V median	14.861	49.245	0.145	0.086	15.184	41.362	0.167	0.094
	Gauss	15.380	43.232	0.156	0.096	15.960	40.071	0.181	0.098
	O. Kriging	16.817	44.845	0.164	0.107	18.658	45.508	0.187	0.118
	I. Kriging	16.561	44.095	0.169	0.114	17.580	41.362	0.205	0.124
Nitrate	V mean	13.852	22.165	0.170	0.085	14.911	22.869	0.210	0.070
	V median	13.689	23.512	0.163	0.084	13.765	23.140	0.187	0.068
	Gauss	13.938	22.079	0.171	0.088	14.899	22.969	0.208	0.072
	O. Kriging	13.853	22.486	0.166	0.090	15.507	24.625	0.200	0.076
	I. Kriging	15.501	23.742	0.189	0.102	15.634	24.040	0.215	0.086
pH	V mean	0.192	0.337	0.185	0.098	0.218	0.358	0.209	0.098
	V median	0.192	0.339	0.184	0.098	0.218	0.361	0.206	0.098
	Gauss	0.194	0.342	0.185	0.101	0.223	0.373	0.203	0.098
	O. Kriging	0.198	0.343	0.187	0.115	0.221	0.362	0.208	0.106
	I. Kriging	0.200	0.345	0.194	0.117	0.226	0.373	0.214	0.106
Dissolved Oxygen	V mean	1.903	2.477	0.189	0.095	2.069	2.630	0.218	0.122
	V median	1.876	2.499	0.185	0.095	1.969	2.599	0.205	0.122
	Gauss	2.049	2.612	0.205	0.095	2.095	2.646	0.221	0.122
	O. Kriging	1.911	2.498	0.189	0.102	2.255	2.998	0.228	0.176
	I. Kriging	1.989	2.569	0.198	0.105	2.214	2.777	0.235	0.174
Sulfate	V mean	40.903	113.028	0.131	0.139	42.768	127.054	0.150	0.092
	V median	34.992	115.860	0.112	0.095	39.433	128.980	0.131	0.074
	Gauss	38.128	113.740	0.124	0.101	42.546	127.451	0.149	0.094
	O. Kriging	42.365	115.104	0.135	0.125	50.238	139.489	0.155	0.132
	I. Kriging	43.979	114.601	0.148	0.139	46.900	127.399	0.182	0.144

Table 6.3: Results of crossvalidation and split sampling for different interpolation methods

Parameter	v-Copula		N-Copula		O-kriging		I-kriging	
	75 %	90 %	75 %	90 %	75 %	90 %	75 %	90 %
Chloride	73.8	87.2	77.8	88.3	94.8	96.8	70.2	83.9
Nitrate	75.8	89.0	79.8	89.5	85.8	92.5	68.9	82.0
pH	74.8	90.5	80.2	91.1	88.2	94.1	70.5	84.0
Dissolved oxygen	78.8	91.0	63.2	79.0	76.7	89.5	69.0	83.6
Sulfat	79.4	89.3	78.6	88.5	93.3	95.9	71.5	84.6

Table 6.4: Results of crossvalidation for the confidence intervals of the different interpolation methods

As pointed out above the confidence intervals corresponding to the estimators are very different. The cross validation approach was also used to validate the confidence intervals. For this purpose the number of observations falling into the 75 % and the 90 % confidence intervals were calculated. The results are shown in table 6.4. As one can see the most reliable confidence intervals are obtained by the v-copulas followed by the normal copulas. The kriging estimation errors are nearly useless, they are a mere index of the spatial distribution of the observation points ([49]). The only reasonable confidence intervals for kriging are obtained for dissolved oxygen. The reason for this is the nearly Gaussian multivariate distribution of the parameter. The Indicator kriging intervals are far too optimistic.

### 6.3 Summary of Chapter 6

In this chapter an interpolation method based on copulas was introduced.

- The proposed method is based on the separation of the marginal distributions from the dependence structure. The dependence structure is described using copulas.
- An interpolation method for two different parametric copulas the Gaussian and the v-transformed normal copulas was developed. The conditional distribution (conditioned on the observed values at the observation points) for unobserved locations was derived from the conditional copula and the marginal distribution.
- Conditional multivariate copulas corresponding to any finite set of points conditioned on  $n$  observations can be written explicitly. Equation (6.10) as well as Equation (6.9) and (4.12) allow the specification of the copulas.
- The copula based interpolator is non linear and the estimated value depends both on the configuration of the observations and target points as well as the observed values.
- For the application, the copula based interpolators give better cross validation results than ordinary and indicator kriging.
- The copula based interpolation allows the calculation of confidence intervals. Cross validation results indicate that these intervals are more realistic than those based on the estimation variance obtained by ordinary or indicator kriging.

---

The copula-based estimation can be incorporated into the sequential conditional simulation procedure (see Chapter 7). At each simulation point, an estimator of the *cdf* value can be drawn from the conditional copula of that point conditioned on the neighbouring observations.

## 7 Simulation of Random Fields using non-Gaussian Dependence

In hydrogeological engineering designs, such as groundwater management, aquifer remediation, underground waste disposal, etc., groundwater flow and mass transport equations are solved to provide the output flow rate and transport variables serving as input to the optimization process that leads to the final decision (*Gómez and Wen [1998] [34]*). Since the heterogeneity organization of the sub-soils or aquifers considerably influence flow and transport behaviour, it is of vital importance to characterize the spatial field of relevant parameters.

However, due to the spatial variability of the parameters, the full characterization of the heterogeneous fields is very costly and even impossible. In practice, only limited information on the spatial variation of the parameters under interest is available. As a consequence, one must work with highly uncertain spatial parameters which are considered as realizations of certain prior random functions. The uncertainties in the field characterization of hydrogeological parameters propagate through the flow and transport equations to the output variables. In most cases, statistical bounds of required output variables (e.g., effective permeability, breakthrough time, etc.) are estimated using Monte Carlo analyses by running flow and transport simulations through equiprobable spatial realizations of important hydrogeological parameters (e.g., hydraulic conductivity, parameters in the water retention curves, etc). Thus geostatistical simulation which conditions all the subsequent flow and transport simulations becomes an indispensable component of decision making process.

Because of the lack of measurements in adequate resolutions, the heterogeneity of the spatially distributed parameters is often characterized only by a histogram, which represents the univariate distribution, and a covariance/variogram function of separation distance, which reflects the bivariate spatial dependence (*Zinn and C. F. Harvey [2003] [95]*). The majority of geostatistical simulation algorithms also rely on these two statistics to model the spatial variability of the parameter under study. This fact entails the assumption of a multivariate Gaussian dependence for either the parameter itself or for its normal score transform (*Journel and Deutsch [1993] [51]*). The multivariate Gaussian model is attractive mainly due to two reasons. First of all, it is analytically simple, i.e., it can be fully characterized by its mean and covariance. Therefore it allows easy model inference from the available observations. Second, among all the random functions sharing the same covariance function, the Gaussian model is the maximum entropy model (*Journel and Deutsch [1993] [51]*). This means it maximizes the spatial disorder beyond the imposed covariance and minimizes unwarranted structural properties (*Journel and Deutsch [1993] [51]*). However, it is not justified to choose a congenial model just because data shortage prevents refuting it (*Gómez and Wen [1998] [34]*).

There are several arguments against an arbitrary choice of a multivariate Gaussian model. First, although most field data have been reported to follow a Gaussian univariate distribution for log hydraulic conductivity (*De Rooij* [2004] [75]). It does not ensure a multivariate Gaussian dependence. In fact, many multivariate spatial random functions can share the same univariate Gaussian distribution and covariance function, yet differ by their patterns of spatial continuity at different threshold values. In other words, a univariate histogram and a covariance function estimated from the field data may not be sufficient to characterize the underlying spatial dependence. Second, a multivariate Gaussian model exhibits a symmetrical dependence structure with respect to the median and always implies the minimal spatial correlation length or connectivity for the extreme values. *Journel and Alabert* [1989] [50] studied the indicator correlogram of a Gaussian model and showed that the indicator covariance has its maximum at the median threshold and diminishes symmetrically as the threshold departs from the median toward the low and high extreme ends. This feature contradicts with many geological formations. For instance, aquifers containing fractures, channels, aquitards or cavities, etc., can have connected high conductivity values or low conductivity values, and the pattern of spatial continuity or dependence strength is not necessarily symmetrical with respect to the median. Thus the multivariate Gaussian model lacks the flexibility to model the spatial connectivity/dependence of specific classes of relevant hydrogeological parameter values, hence considerably restricts the flow and transport simulations and may produce inaccurate prediction of required output variables. Third, the multivariate Gaussian model is a maximum entropy model that maximizes the uncertainty beyond the known information. However, a maximum entropy model does not entail maximum entropy of the response distributions obtained from processing realizations of that model (*Journel and Deutsch* [1993] [51]). *Journel and Deutsch* showed that flow simulations performed on multiple realizations of a multivariate Gaussian distribution model are very similar. Thus the space of response uncertainty could be too narrow causing a misleading sense of precision.

In the last decades, the limitations of the multi-Gaussian assumption have been received more and more attentions. A few researches have been carried out to study the effect of spatial dependence on flow and transport behavior. As mentioned before, *Journel and Deutsch* [1993] [51] compared the results of flow simulations performed on realizations of a Gaussian model, a multiple indicator model and a mosaic model. They found significant differences in the response distributions. *Gómez and Wen* [1998] [34] demonstrated the misleading overstatement of safety of transport problem by selecting a multi-Gaussian model under the condition that no higher-order moments are estimated from the field data to validate the multi-Gaussian dependence. *Zinn and Harvey* [2003] [95] showed that spatial fields sharing the same marginal distributions and covariance structure, yet differ in the way whether extreme low, median or extreme high values are most connected, exhibit very different flow and transport characteristics. *Neuweiler and Cirpka* [2005] [67] showed that the relative permeability of soils with well-connected coarse materials differ largely from that of the soils with well-connected fine materials or those without particular connectedness. In *Neuweiler and Vogel* [2007] [68], the information of connected paths of different parameter ranges are taken into account in estimating effective unsaturated parameters. All the arguments and studies call for new measures of spatial dependence and simulation methods which are able

to characterize and generate spatial fields with non-Gaussian dependence, hence new types of random functions.

The most widely used algorithm of generating multivariate non-Gaussian realizations is the sequential indicator simulation [4] [24] [35] [47] which requires the specification of a variogram model for each cutoff. A similar approach called sequential isofactorial simulation was developed by *Emery* [2002] [26]. It is an improved version of the former in terms of applying a complete bivariate law and mitigating the inference effort. The simulated annealing algorithm was also suggested as a possible alternative ([1] [23] [28]), where the generation of a realization is obtained by perturbing iteratively an initial pure random field using a stochastic relaxation technique. The perturbation process is controlled by optimizing an objective function which measures the difference between the desired spatial characteristics and those of the realizations [34]. When dealing with categorical variables and random sets, other simulation techniques are available such as truncated plurigaussian method [62] and Boolean models [18] [57]. Recently, a simulation approach of categorical variables using multi-point statistics [36] [84] is developed and widely used. It tackles the problem by incorporating higher-order-moment-statistics inferred from a training image which is either taken directly or synthesized from the field data.

In this thesis work, the concept of copulas is adopted to both investigate and simulate different spatial dependence. Empirical copulas and measures of asymmetry (see Chapter 3) are used as supplements to the traditional geostatistic tools to investigate the spatial dependence of flow and transport relevant parameters. Simulation approaches of multivariate non-Gaussian spatial fields for both the unconditional and conditional cases are developed. The non-Gaussian copula models described in Chapter 4 are applied as prior random functions for the simulation purposes. The methodology proposed in this thesis differs from the above mentioned approaches in that it does not only consider the multiple-point statistics but also imposes an explicit prior non-Gaussian random function model in the simulation procedure.

In the following two sections, the procedures of unconditional simulation and conditional simulation of multivariate non-Gaussian fields are described, respectively. In the third section, the way of generating realizations of different multivariate dependence structures with the same second moment statistics, namely the correlation structure, is discussed. In the last section of this chapter, the spatial dependence of a dataset of hydraulic conductivity related parameters are studied using the concept of copulas. Both Gaussian and non-Gaussian copula models are parameterized using the parameterization approach described in Chapter 5 and the corresponding simulations are carried out. Statistical tests are done to compare the resulting images of the simulations with the original image of the dataset.

## 7.1 Unconditional simulation

For a spatial realization of a multi-Gaussian model, the study of the empirical copulas conveys two important informations about the dependence. First, the highest copula density is located in the extremes, while the median values have the lowest density along the main



diagonal. Second, the copula density is fully symmetrical with respect to the secondary diagonal. Figure 7.1 shows the empirical copulas of a multi-Gaussian realization for three separation lengths. Figure 7.2 shows the corresponding spatial realization, values of which are in the copula space.

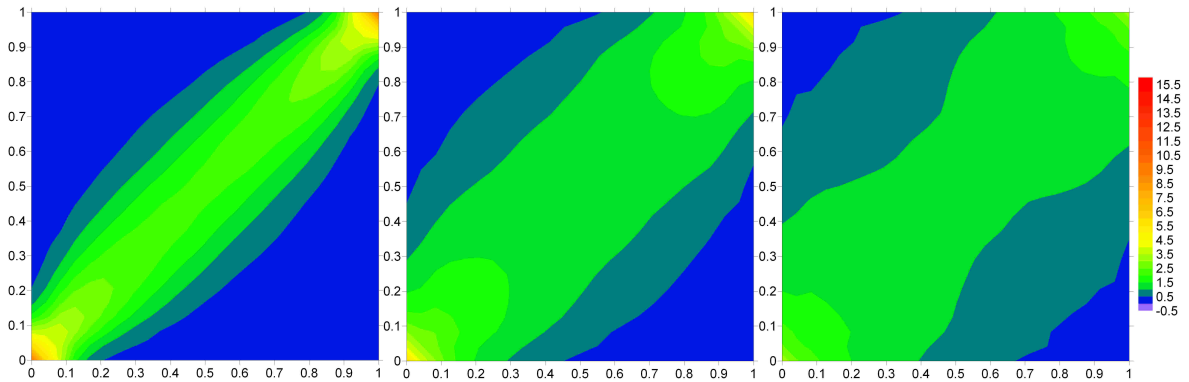


Figure 7.1: Empirical copulas of a multi-Gaussian realization for the separation lengths of 11%, 33% and 56% correlation length (from left to right).

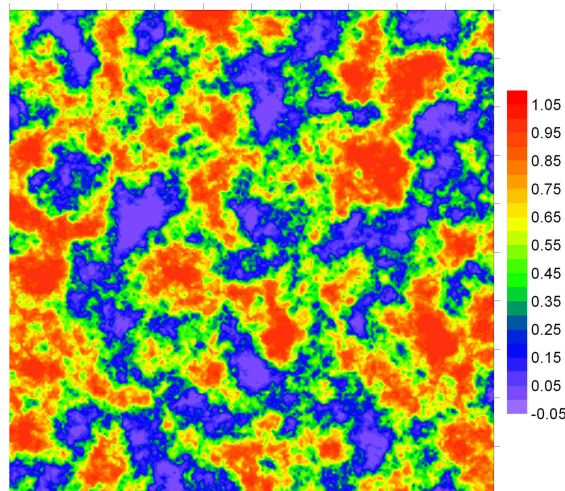


Figure 7.2: A spatial multi-Gaussian realization in copula space.

The typical spatial patterns of the multi-Gaussian realizations can be seen from Figure 7.2. In Figure 7.2, both the high quantiles (indicated by the red color) and low quantiles (indicated by the blue color) tend to form cluster in isolated blobs, while the middle quantiles (indicated by the green color) take shape in connected channels that span the entire length of the field. In agreement with the findings in the relationship between Euler characteristics and empirical copulas of the spatial field (see Chapter 3), Figure 7.1 and 7.2 show that the

less connected quantiles have higher copula densities, while the more connected quantiles have lower copula densities.

As mentioned before, the extreme values from a multi-Gaussian realization are least connected, a feature critical for flow and transport and a feature that contradicts with many geological settings. *Zinn and Harvey* [2003] [95] described in their paper how to generate spatial fields with connected extreme values from multi-Gaussian realizations by applying a certain transformation of the Gaussian random variable. Actually any non-monotonic transformation of a Gaussian process can create spatial fields with non-Gaussian dependence, since non-monotonic transformations change the ranks of the values, and the dependence strength associated with the values are shifted together with the values to a different quantile in the new space. The v-shaped transformation applied to construct the v-transformed normal copula is a special case of non-monotonic transformations (see Chapter 4).

Since a large number of algorithms exist for generating realizations from a multi-Gaussian model, e.g. the turning bands method [22] [46] [63], the spectral decomposition method [65] [79] [14], the covariance matrix decomposition method [5] [21] and so on. Utilizing the effect of non-monotonic transformation on the dependence structure, non-Gaussian realizations can be obtained easily by performing the following procedure:

1. Generate a multi-Gaussian realization with the existing approaches.
2. Apply a non-monotonic transformation to the generated Gaussian process.
3. Transform the univariate marginal distribution of the simulated field to uniform distribution by calculating the cdf values of the simulated values.
4. Put on top a desired marginal distribution onto the realization by mapping the cdf value at each point to the desired distribution cdf.

Figure 7.3 and Figure 7.4 show two 'wilder' non-monotonic transformations and the resulting non-Gaussian realizations as well as the corresponding empirical copulas. In the first transformation (shown in the upper left corner of Figure 7.3), the originally extreme high and low quantiles are shifted to the middle quantiles, while the originally middle quantiles are shifted to the low quantiles in the new space, which results in a new copula with middle values being most correlated (shown in the upper right corner of Figure 7.3). Therefore in the final realization (shown in the bottom row of Figure 7.3) in copula space, the middle quantiles become most clustered, while the extreme low quantiles are most connected. Since the extreme high quantiles in the new space come from the original middle to high quantiles, the dependence of the extreme high quantiles in the new space is thus stronger than the extreme low quantiles and weaker than the middle quantiles in the new space, leading to a small scale connected spatial pattern which does not span the entire field as the channels formed by the extreme low quantiles.

In the second transformation (shown in the upper left corner of Figure 7.4), the original middle quantiles are also folded to the extreme low quantiles in the new space. The originally extreme low quantiles are shifted to the middle to high quantiles in the new space which

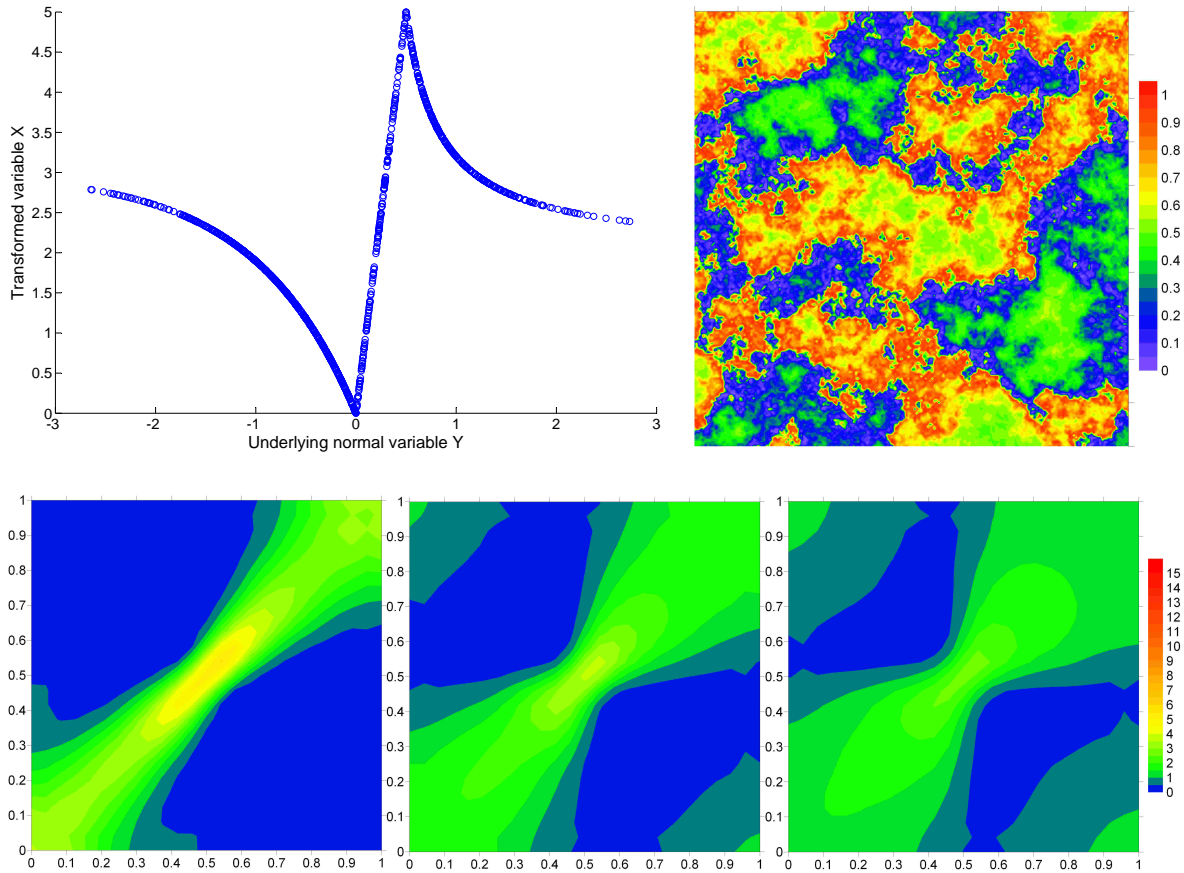


Figure 7.3: An example of non-monotonic transformation (right upper corner) and resultant spatial realization (left upper corner) and its corresponding empirical copulas (bottom row) for the separation lengths of 12.5%, 18.75% and 31.25% correlation length.

brings a peak in this part in the empirical copula plot. A sinusoidal function is applied to the original middle to high quantiles and this part becomes the extreme high quantiles in the new space. The dependence strengths associated with the original middle and high quantiles blend together. That is why the peak in the extreme high quantiles in the empirical copula plot is blurred and weakened (shown in the bottom row of Figure 7.4). In the resulting realizations (shown in the upper right corner of Figure 7.4) in copula space, the low quantiles are still most connected, the middle to high quantiles are most clustered. Whereas a 'salt-and-pepper' pattern appears in the clusters of extreme high quantiles which is due to the randomization effect of the sinusoidal function.

In this way, realizations of various dependence structures can be obtained. More interesting structures can be generated by combining different processes which are independent of each other and have different spatial continuity characteristics. One example is the combination of two independent random processes which have orthogonal anisotropic directions:

$$\mathbf{X} = f(\mathbf{Y}_1, \mathbf{Y}_2) \quad (7.1)$$

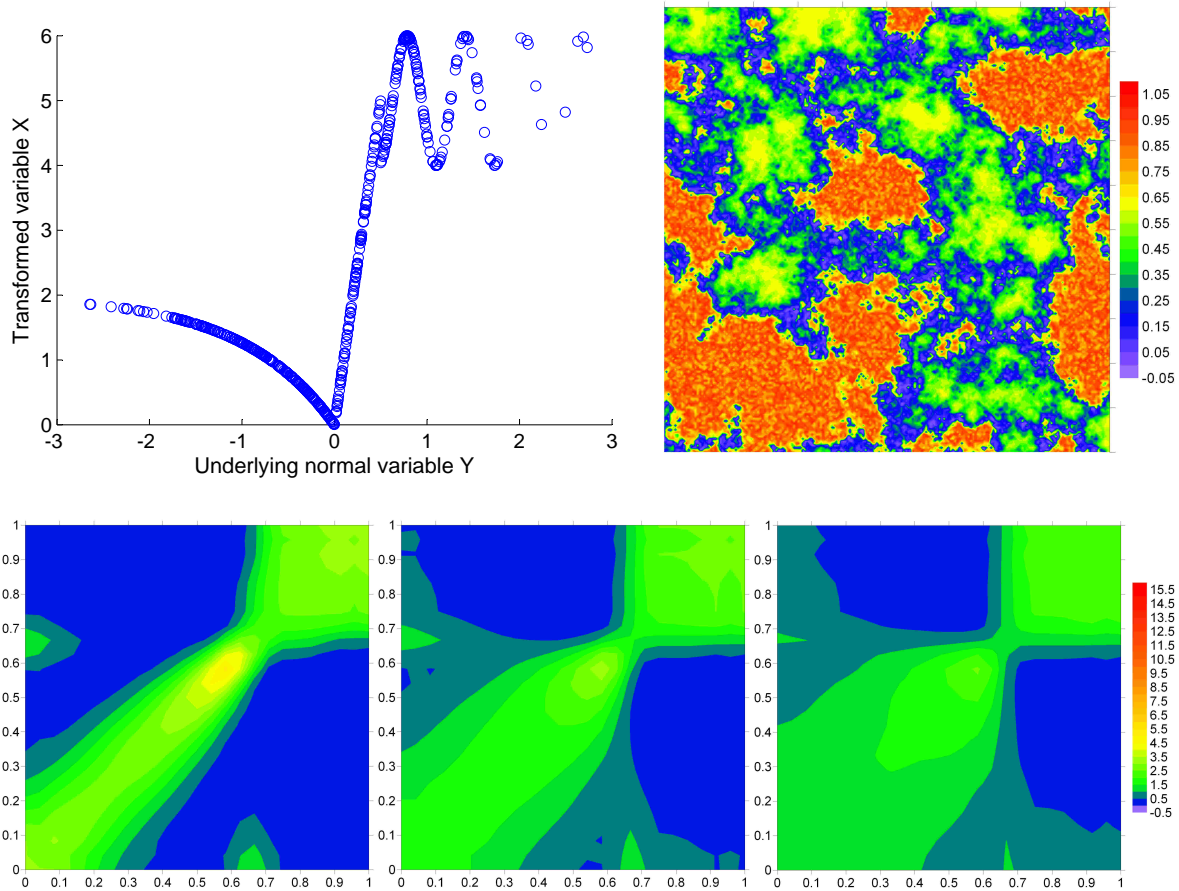


Figure 7.4: Another example of non-monotonic transformation (right upper corner) and resultant spatial realization (left upper corner) and its corresponding empirical copulas (bottom row) for the separation lengths of 13%, 19% and 32% correlation length.

where  $X$  is the resulting non-Gaussian process,  $Y_1$  and  $Y_2$  are independent random processes, which can be Gaussian or non-Gaussian. If a non-Gaussian process is used, then a non-monotonic transformation has to be applied to the corresponding underlying Gaussian process. One of  $Y_1, Y_2$  has a vertical to horizontal anisotropy (highest continuity along the horizontal direction) and the other has a horizontal to vertical anisotropy (highest continuity along the vertical direction).  $f(\cdot)$  denotes the combination function, which can be any kind of function. For instance, taking the maximum is a special case of the combination function. The left chart of Figure 7.5 shows a realizations of combining two independent Gaussian processes with orthogonal anisotropies using the maximum function. It can be used to resemble some root structures. The middle and the right charts of Figure 7.5 are realizations generated with more complicated combination functions. The former one is one realization of the process described in Equation (7.2) and mimics leakage of contamination from layered pools. The latter one is the realization of the process described in Equation (7.3) and looks like a karstic formation. All the realizations are transformed to copula space.

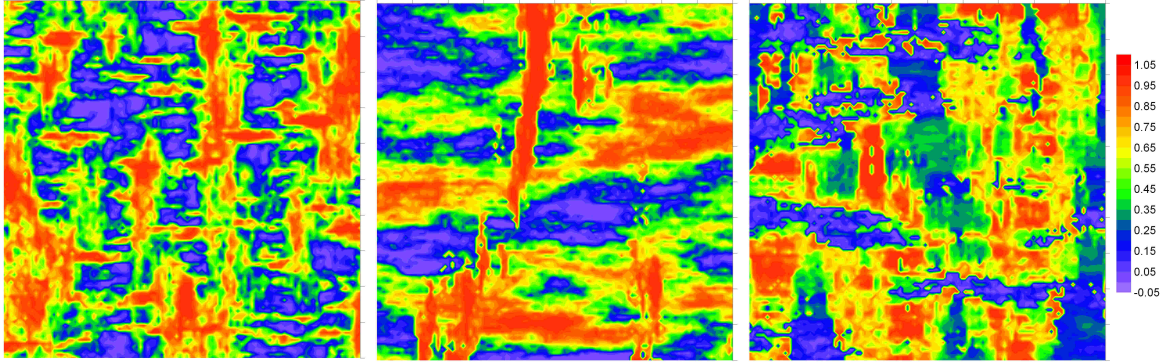


Figure 7.5: Example of realizations from combined processes.

$$X = \begin{cases} (Y_2 - 2.0)^2 & \text{if } Y_1 < 1.3 \\ 5.0Y_1^2 & \text{else} \end{cases} \quad (7.2)$$

$$X = \begin{cases} \exp(1 - Y_1 \cdot Y_2) & \text{if } 1.0 < Y_1 < 1.5 \text{ and } Y_2 \leq 0.8 \\ Y_2^{-1} - Y_1 \cdot Y_2 & \text{elseif } Y_2 > 1.5 \\ \log|Y_2 - 1.0| & \text{else} \end{cases} \quad (7.3)$$

with  $Y_1$  and  $Y_2$  being independent normal processes with orthogonal anisotropic directions.

This method offers tremendous degrees of freedom to generate interesting and fantastic spatial structures. Nonetheless, the challenge of this approach is to find out the corresponding transformation which generates the desired spatial structures or variabilities.

## 7.2 Sequential conditional simulation

Conditional simulations aim to generate realizations reproducing the prescribed variability honoring the observation data at the sampled locations. In this case if realizations of multi-non-Gaussian functions are required, the procedure proposed for the unconditional simulation is not applicable anymore. Because non-monotonic transformations also change the dependence among the sampling points, thus the information on the spatial variability delivered by the observations will be lost after the transformation. In order to solve this problem, the concept of sequential simulation is adopted for the conditional case. In the following paragraphs, the principle of sequential conditional simulation is introduced. The reference of this part is taken from *King* [2000] [53]



The foundation of sequential conditional simulation is *Bayes' Theorem* and Monte Carlo simulation. Let  $Z(\mathbf{s}'_j), j = 1, \dots, N$  be a set of random variables, also called as a random function at the  $N$  grid cells with locations denoted by  $\mathbf{s}'_j$ . Suppose there are  $n$  observations  $z^o(\mathbf{s}_k), k = 1, \dots, n$  in the study domain. The objective of a sequential conditional simulation is to generate several realizations  $z^l(\mathbf{s}'_j, j = 1, \dots, N), l = 1, \dots, L$  of the random function, conditioned on the data set  $z^o(\mathbf{s}_k), k = 1, \dots, n$ . This gives an  $N$ -point conditional cumulative distribution function (ccdf) that models the joint probability at the  $N$  locations  $\mathbf{s}'_j$ :

$$\begin{aligned} & F\left(z_1(\mathbf{s}'_1), \dots, z_N(\mathbf{s}'_N) | z_1^o(\mathbf{s}_1), \dots, z_n^o(\mathbf{s}_n)\right) \\ &= P\left(Z(\mathbf{s}'_1) \leq z_1, \dots, Z(\mathbf{s}'_N) \leq z_N | Z(\mathbf{s}_1) = z_1^o, \dots, Z(\mathbf{s}_n) = z_n^o\right) \end{aligned} \quad (7.4)$$

By recursively applying *Bayes' Theorem*, the  $N$ -dimensional ccdf can be decomposed as a product of  $N$  one-dimensional ccdfs:

$$\begin{aligned} & F\left(z_1(\mathbf{s}'_1), \dots, z_N(\mathbf{s}'_N) | (n)\right) \\ &= F\left(z_N(\mathbf{s}'_N) | (n + N - 1)\right) \times F\left(z_{N-1}(\mathbf{s}'_{N-1}) | (n + N - 2)\right) \times \dots \times F\left(z_1(\mathbf{s}'_1) | (n)\right) \end{aligned} \quad (7.5)$$

Eq. (7.5) enables the simulation to be made in  $N$  successive or sequential steps. At the first location,  $\mathbf{s}'_1$ , the ccdf is calculated conditioned on the  $n$  original observations:

$$F\left(z(\mathbf{s}'_1) | (n)\right) = P\left(Z(\mathbf{s}'_1) \leq z | (n)\right) \quad (7.6)$$

Then an estimate  $z^l(\mathbf{s}'_1)$  is randomly drawn from the above ccdf and added to the conditioning data set for all the subsequent simulations of the rest nodes. This process is repeated until all of the  $N$  nodes have simulated values. Then one realization is done. More realizations can be obtained by starting with the original data set and visiting the  $N$  nodes in different sequences.

The sequential simulation algorithm requires the determination of the ccdf at each simulation node. When the sequential Gaussian simulation (SGS) [24] is performed, the ccdf is determined by a mean and a variance estimated from either simple kriging or ordinary kriging with the neighboring data points, given the ccdf follows a normal distribution. Since the kriging procedure uses covariance/variogram function as the sole descriptor to determine the spatial dependence, the resulting realizations implicitly follow a multivariate Gaussian distribution. The SGS requires the normal score transformation of the observation series before the simulation. After the simulation, a back transform of the normal scores is necessary in order to reproduce the histogram of the attribute under study.

For the Gaussian process, the analytical expression of relationship between the indicator correlogram and  $p$ -quantile used to define the corresponding indicator group has been found as following [6] [94] [3]:

$$\rho_l(\mathbf{h}; p) = \frac{1}{2\pi p(1-p)} \int_0^{\arcsin \rho_Y(\mathbf{h})} \exp - \frac{y_p^2}{1 + \sin \theta} \quad (7.7)$$

with  $y_p$  being the standard normal  $p$ -quantile, and  $\rho_Y(\mathbf{h}) = EY(\mathbf{s}) \cdot Y(\mathbf{s} + \mathbf{h})$  being the  $Y$ -correlogram.

The relationship reveals that as the  $p$ -quantile tends toward low extreme end or high extreme end, the correlogram of that indicator group also tends toward 0. In addition, the correlograms exhibit a symmetric behaviour about the median quantile. As mentioned before, these properties are in contradiction with many natural processes. Combining the sequential simulation principle and the indicator kriging procedure (SIS) [24], it is possible to generate multivariate non-Gaussian realizations if the set of specified indicator variogram/covariance functions depart from those properties. The cost is the additional inference effort to fit a variogram/covariance model to each cutoff.

The indicator random function models are best suited for simulating categorical variables controlled by two-point statistics [24]. However, many geological structures to be reproduced are curvilinear, e.g., sand channels in a clastic reservoir [84]. Modelling of such spatial structures requires multiple-point statistics that are beyond the traditional two-point variogram/covariance statistics. In the nineties, simulation approaches accounting for multiple-point statistics [29] [23] [48] [81] [82] arised and are mainly applied for reservoir characterization and modelling. In these approaches, no explicit multivariate random function model is defined, instead the multiple-point information is inferred from a training image which is assumed to be representative of the geological heterogeneity.

In this thesis, another way is explored, i.e., a random function is sought to capture the geometrical features of the spatial fields. The principle of sequential simulation is combined with the concept of copula to perform the conditional simulation of non-Gaussian realizations. The detailed procedure of this approach is as follows:

1. Transform the data values into the copula space by calculating their empirical cumulative distribution values.
2. Select a theoretical copula model which can describe the spatial dependence of the dataset and parameterize that model.
3. Define a random path through all the unknown nodes to be simulated.
4. Calculate the conditional copula using the parameterized copula model at an unknown node conditioned on the neighbouring data:

$$C_{\mathbf{s},n} \left( u | U_1 = F_Z(z_1^o), \dots, U_n = F_Z(z_n^o) \right) \quad (7.8)$$

where  $F_Z$  denotes the empirical marginal distribution of the random variable,  $z_i^o$  denotes the conditioning data.

5. Draw randomly from the conditional copula a cdf value using Monte Carlo simulation:

$$u^*(\mathbf{s}) = C_{\mathbf{s},n}^{-1}(p|(n)) \quad (7.9)$$

where  $p$  is a random number drawn from the uniform distribution on  $[0, 1]$ .

6. Assign the retrieved  $u^*$  to the corresponding node and add it to the conditioning data.
7. Repeat steps 4 to 6 for all the simulation nodes.
8. Superimpose the marginal distribution estimated from the observations onto the realization.

Like in the interpolation procedure, the conditional distribution is calculated for each unknown node. The spatial dependence is modeled by the multivariate copula which is able to reflect the variation of dependence strength with the change of data values. The  $v$ -transformed normal copula and maximum normal copula (see Chapter 4) offer some flexibilities to model the non-Gaussian spatial dependence. Figure 7.6 shows the examples of realizations of the sequential conditional simulation using the  $v$ -transformed normal copula with different model parameters.

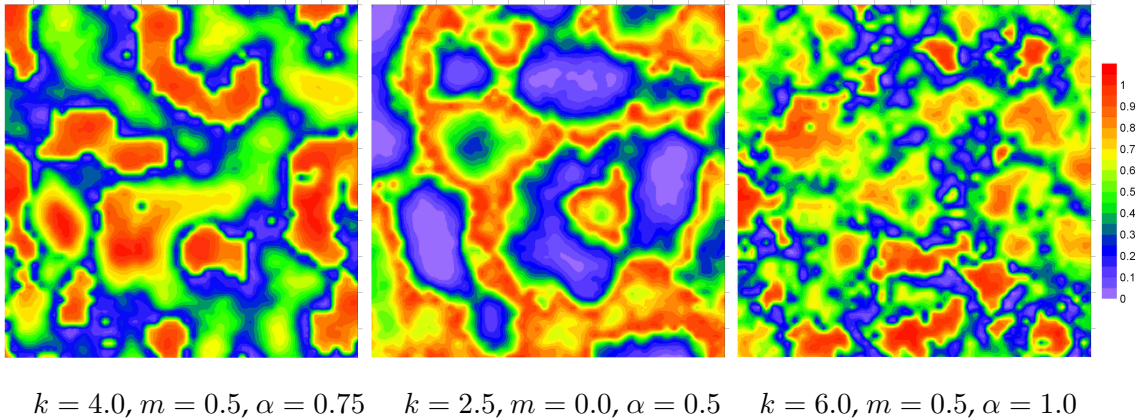


Figure 7.6: Realizations of conditional simulation using  $v$ -transformed normal copula.

From Figure 7.6, one can see that realizations with curvilinear structures can also be generated. The limitation of this approach is that it requires the analytical solution of the selected theoretical copula model. The non-Gaussian theoretical models described in Chapter 4 are derived by applying non-monotonic transformations of the Gaussian process. But if the transformation is getting complicated, it is extremely difficult and even impossible to derive the analytical solution of the resultant non-Gaussian copula. Compared with the method used for the unconditional simulation, the flexibility of generating various spatial structures is reduced.

For the maximum normal copula model, the sequential conditional simulation is not feasible at the moment. Since normally, ten to fifteen or even more conditioning points are required



in order to reflect the long range spatial pattern, the  $2^n$  terms of the product of different partial derivatives and mixed derivatives of the  $n$ -dimensional Gaussian distribution (see Equation (4.17) in Chapter 4) makes it a prohibiting computation task. However, though it is not in the framework of this thesis work, if multi-processor parallel computation is accessible, it is still doable. In this thesis, the maximum normal copula model is parameterized based on the datasets using the proposed parameterization approach described in Chapter 5 where the subsets are defined to enclose three to four points, because three to four dimensional conditional densities can be handled using the numerical approximation described in Equation (4.18) with acceptable precision. After that, unconditional simulations of the parameterized maximum normal copula can be carried out without any difficulty.

### 7.3 Parameter Optimization for Correlation Structure

In order to illuminate the effect of different multivariate dependence structures, the realizations from different multivariate models for comparison should share near-identical marginal distribution and correlation/covariance structure so that any differences observed in the response variables cannot be interpreted as the consequence of the traditional statistics but of the multivariate dependence. To ensure that two realizations have the same marginal distribution, one has to just put on top the same univariate distribution onto both realizations by a one by one mapping of the *cdf* value of the desired distribution to the *cdf* value simulated at each target point. Equation 7.10 describes the details:

$$z_i = F_m^{-1}(u_i) \quad (7.10)$$

where  $u_i$  denotes the simulated *cdf* value at location  $i$ ,  $F_m^{-1}$  denotes the inverse of the desired marginal distribution and  $z_i$  refers to the resultant value at location  $i$  after superposition of the marginal distribution.

As mentioned in Chapter 2 that the copula of a set of random variables is not influenced by the marginal distribution, no matter what kind of marginal distribution is required as the shared property of the realizations of different copula models, the copula parameters are not influenced either.

In this study, instead of the covariance/correlation function we consider the rank correlation function, because as mentioned in Chapter 3, we would like to use scale-invariant measures. In order to ensure that the realizations from different models to share the same rank correlation structure, the model parameters of the applied theoretical copulas have to be optimized. The reason is that the bivariate copula of the spatial realization determines the rank correlation structure (see Equation (3.6)). Under the condition that the main parameters of a theoretical copula model are fixed, i.e., the  $m$ ,  $k$  and  $\alpha$  for the  $v$ -transformed normal copula and  $m$  and  $\sigma$  (the mean and the standard deviation of the second normal process) for the maximum normal copula (Chapter 4), the parameters in the Pearson's correlation model (see Equation (5.8)) should be parameterized.

Note that the Gaussian copula is fully determined by the correlation/covariance structure. Therefore to generate Gaussian realizations with correlation structure corresponding to that of the non-Gaussian realizations can be achieved with less effort. The easiest way is to first generate a large number of non-Gaussian realizations and calculate the power spectrum of the mean field. Using the fact that the power spectrum is the *Fourier* transform of the auto-covariance [15] [19], the backward *Fourier* transform provides the covariance structure which is the single input (the marginal can be superimposed later) required in the spectral simulation method [79] [65] [14] of Gaussian realizations. However, the reverse procedure is not valid, because non-Gaussian dependence cannot be fully characterized by the first and second moment statistics. Due to the same reason, this approach is not applicable to generate realizations of two different non-Gaussian dependence with identical correlation/covariance structure. To overcome this obstacle, a general method towards the optimization of theoretical copula parameters according to a desired rank correlation structure is proposed. The optimization procedure is as following:

1. Specify the desired rank correlation structure for the spatial fields.
2. Fix the main parameters of the selected theoretical copula models. Discretize the range of the Pearson's  $\rho$  [74], which determines the bivariate theoretical copula, into  $n$  (e.g.,  $n = 100$ ) number of intervals.
3. Calculate the bivariate theoretical copula corresponding to each value of the discretized Pearson's  $\rho$ .
4. Once the bivariate copula is obtained, the Spearman's rank correlations  $r$  corresponding to the discretized Pearson's  $\rho$  can also be computed using Equation (3.6).
5. Fit a polynomial to the relationship  $\rho = f(r)$  using the resultant pairs of  $\rho$  and  $r$  (see Figure 7.7).
6. Select out several separation lengths within the correlation length of the spatial fields. And using the relationship derived in step 5 to calculate the  $\rho$  values corresponding to the rank correlations of the selected separation lengths.
7. Select a correlation model to describe the  $\rho$  over distance structure and set tentative values for the correlation model parameters (sills and ranges). Using the tentative parameter values to calculate the  $\rho$  values at the selected separation lengths.
8. Calculate the differences between the  $\rho$  value obtained from step 6 and the  $\rho$  value obtained from step 7 at each separation length (see Figure 7.8). Minimize the sum of the squared differences for all the selected separation lengths with respect to the correlation model parameters.

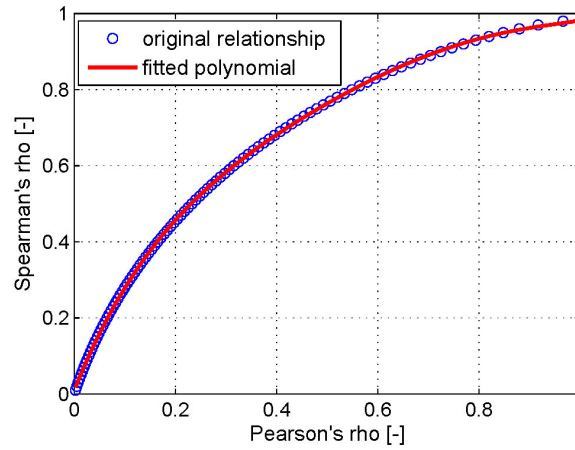


Figure 7.7: Example of fitting a polynomial to the relationship  $\rho = f(r)$  for a  $v$ -transformed normal copula under  $m = 0.0$ ,  $k = 2.5$  and  $\alpha = 0.5$ .

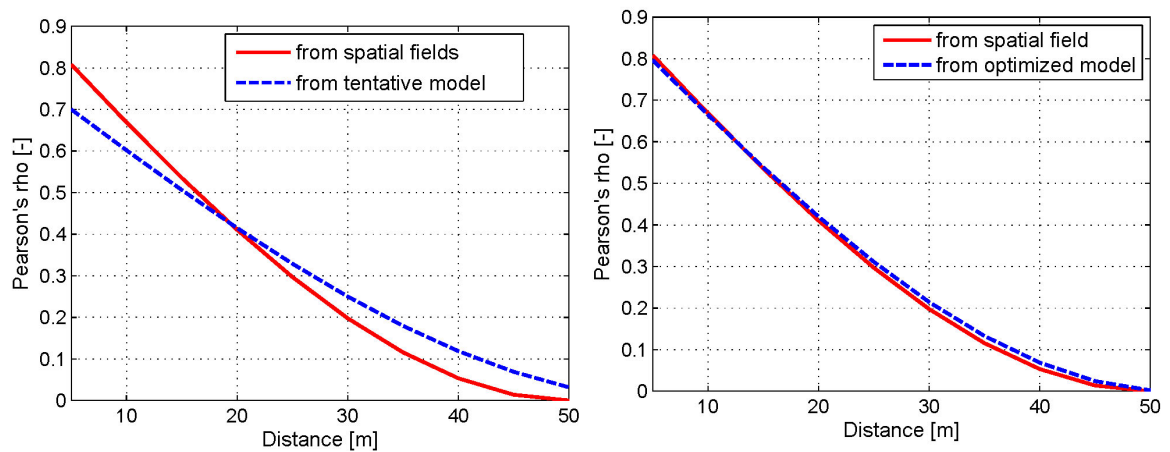


Figure 7.8: Comparison of Pearson's rho over distance structures calculated from step 6 and from step 7 before (left) and after several iterations for optimization (right).

## 7.4 Dataset Investigation

The field-scale test at the Las Cruces Trench Site which is located in the northeast of Las Cruces, New Mexico, provides a database for comparing and validating both stochastic and deterministic models in highly variable, unsaturated soils [40]. The samples used for characterization of hydraulic properties were taken from nine soil layers with about 50 samples per layer spaced 50 cm apart in a 6-m deep by 25-m wide vertical plane in a trench. Water retention data and saturated hydraulic conductivities were measured in the laboratory. Van-Genuchten's relationship was fitted to the water retention data for each sample core. The saturated hydraulic conductivities at each sampling location were also measured in situ using a bore hole permeameter [40]. More details about the dataset can be found in *Wierenga et al.* [1988] [92], [1991] [93] and *Hills et al.* [1991] [41].

In this study, the in-situ measured hydraulic conductivities were used to investigate the spatial dependence. Figure 7.9 shows the classified post plot of the saturated hydraulic conductivity field indicating significant spatial variability with a vertical to horizontal anisotropy.

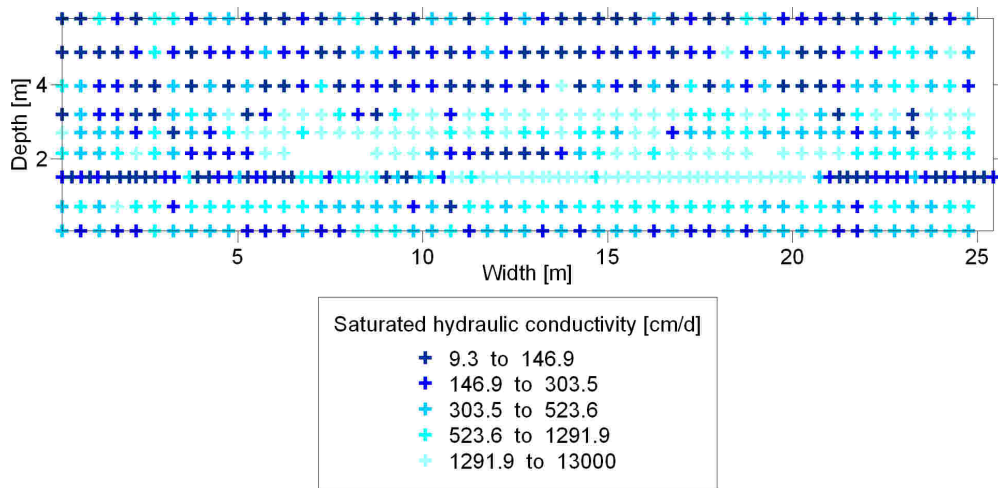


Figure 7.9: *Classified post plot of the field saturated hydraulic conductivity.*

The histogram of the logarithm of the the saturated hydraulic conductivity indicates Gaussianity of the marginal distribution (see Figure 7.10). However, the study of the empirical copulas reveals that the spatial dependence of the variable exhibits a non-Gaussian behaviour.

Empirical copulas along the horizontal direction and for the omnidirection are calculated for several separation lengths with an interval of 0.5 meters. Figure 7.11 shows the corresponding contour plot. The contours of the copula density show the nonsymmetrical dependence with high quantiles being more correlated for the calculated separation lengths. Compared with the omnidirection, the structure of the copula density along the horizontal direction preserves for longer separation lengths. This is consistent with the findings of anisotropic condition from the post plot in Figure 7.9.

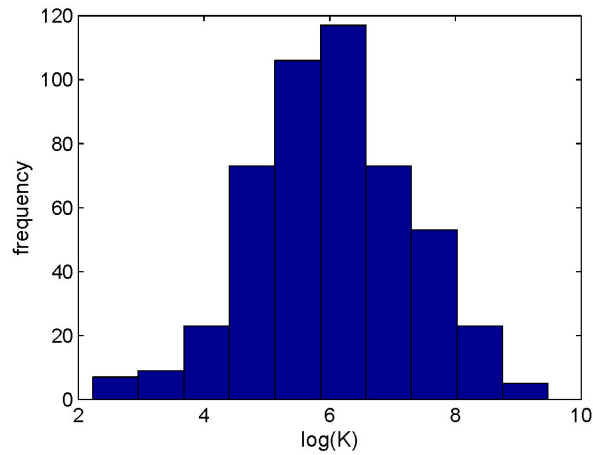


Figure 7.10: Histogram of the logarithm of the saturated hydraulic conductivity.

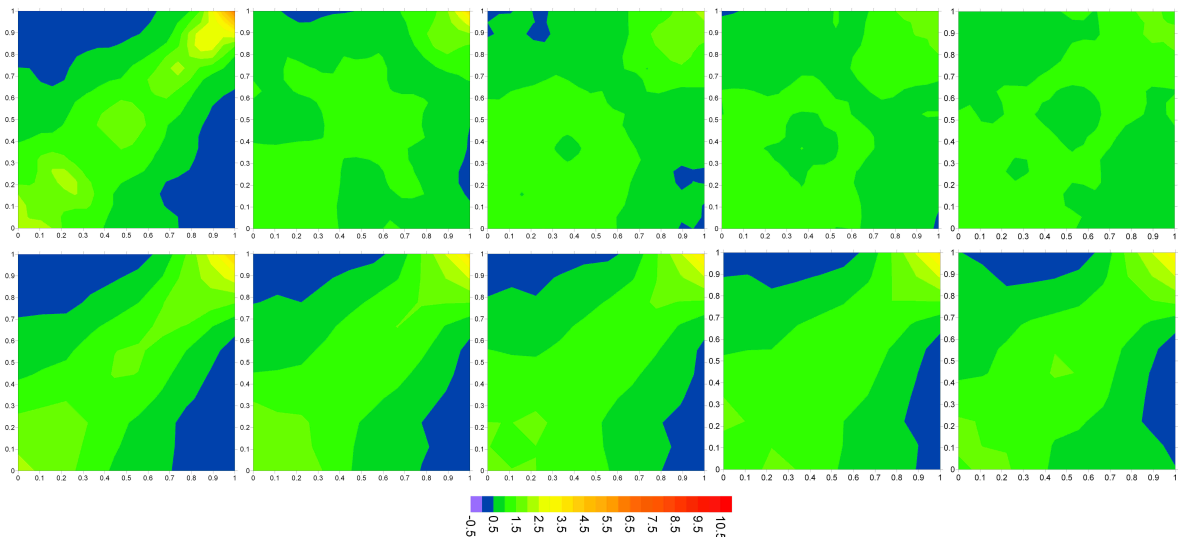


Figure 7.11: Empirical copulas of saturated hydraulic conductivity for omnidirectional (upper row) and horizontal direction (bottom row). The separation lengths from left to right are 0.5 m, 1.0 m, 1.5 m, 2.0 m and 2.5 m.

The rank correlations and measure of asymmetry are also calculated for the horizontal and omnidirectional for several lags spaced by 0.5 meters. Figure 7.12 shows the charts of rank correlation over distance (left) and asymmetry over distance curves (right).

The asymmetry over distance curves also reveals the nonsymmetrical property of the spatial dependence. For shorter separation lengths, the asymmetry values are positive indicating that the high quantiles are more dependent. When distance is beyond 5 meters, the sign

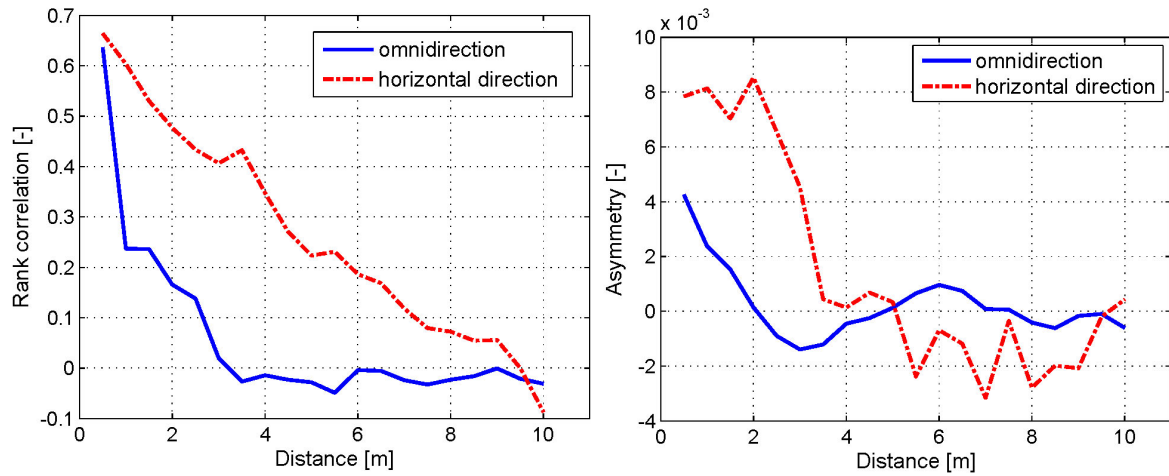


Figure 7.12: Rank correlation over distance curves (left) and asymmetry over distance (right) curves of the saturated hydraulic conductivity field.

of the asymmetry is reversed, which means the low quantiles start to be more dependent. Comparing both the asymmetry and rank correlation over distance curves along the horizontal direction and those of the omnidirection, the anisotropic property of the spatial field is also evident.

For the vertical direction, the resolution is relatively low and there is not enough data pairs to clearly reveal the spatial variation, therefore the calculations are only performed for the horizontal direction and the omnidirection.

Three copula models, a Gaussian copula, a  $v$ -transformed normal copula and a maximum normal copula, are parameterized for this dataset with the proposed parameterization approach described in Chapter 5. The test field is divided into 160 subsets with each subset containing 3 sample points. The radii of the circumcircle of the subsets range from 0.5 m to 24 m. Figure 7.13 shows the distribution of the radii.

More emphasis has been put on the short-distance-dependence. A few subsets with bigger radii has been chosen in order to consider the long-range pattern.

Table 7.1, Table 7.2 and Table 7.3 list the optimized model parameters of Gaussian copula,  $v$ -transformed normal copula and maximum normal copula, respectively.

In Table 7.1,  $C_0$  and  $C_1$  denote the contribution of a nugget effect and spherical model to the correlation structure model,  $R_1$  denotes the range of the spherical component along the horizontal direction and  $\lambda$  refers to the ratio between the horizontal correlation length and the vertical correlation length. In Table 7.2,  $m$ ,  $k$  and  $\alpha$  denote the parameters in the  $v$ -shaped transformation (see Chapter 4).  $C_1$ ,  $R_1$  and  $\lambda$  have the same meaning as in those in Table 7.1. In Table 7.3,  $Y_1$  and  $Y_2$  indicate the underlying two independent normal processes,  $m$  and  $\sigma$  indicate the mean and standard deviation of the underlying normal processes.  $C_1$ ,  $R_1$  and  $\lambda$  have the same meaning as in those in Table 7.1 and Table 7.2, while  $C_2$  and  $R_2$  denote the contribution of a second spherical model and its range along the horizontal direction, respectively.

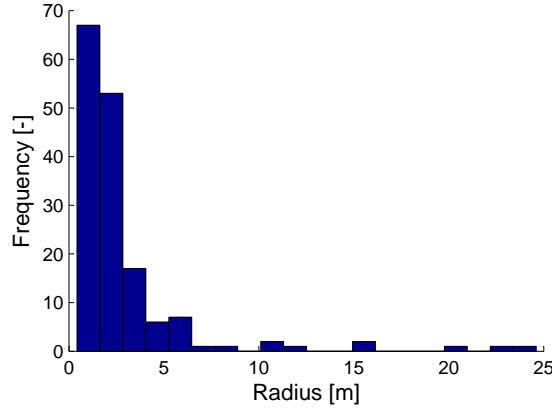


Figure 7.13: Histogram of the radii of the circumcircles of the subsets.

$C_0$	$C_1$	$R_1$ [m]	$\lambda$
0.28	0.72	11.32	7.54

Table 7.1: Optimized parameters of a Gaussian copula from the saturated hydraulic conductivity measurements.

$m$	$k$	$\alpha$	$C_0$	$C_1$	$R_1$ [m]	$\lambda$
0.57	2.43	0.21	0.09	0.91	18.58	4.85

Table 7.2: Optimized parameters of a  $v$ -transformed normal copula from the saturated hydraulic conductivity measurements.

In order to check the affinity between the empirical copulas and the theoretical copulas, the analytical bivariate copula densities of the three parameterized models are calculated given the  $\rho$  values of the underlying normal process corresponding to the separation lengths used for the empirical copula calculations. The  $\rho$  values are computed with the optimized parameters of the covariance function. Figure 7.14 shows the bivariate theoretical copula densities of the parameterized Gaussian copula (first row),  $v$ -transformed normal copula (second row) and maximum normal copula (third row), for the  $\rho$  values corresponding to the separation lengths of 0.5 m, 2.5 m, 3.5 m, 4.4 m and 6.5 m from left to right. The last row of Figure 7.14 are the empirical copulas of the corresponding separation lengths.

The  $v$ -transformed normal copula and the maximum normal copula resembles better the empirical copula than the Gaussian copula model. But it is hard to tell which one of the  $v$ -transformed normal copula and the maximum normal copula fits better to the dataset. It seems that for the shorter distances, the former one fits better. While for the medium to long distances, the latter one is more suitable.

Once the theoretical bivariate copulas are obtained, the corresponding rank correlations and

	$m$	$\sigma$	$C_0$	$C_1$	$C_2$	$R_1$ [m]	$R_2$ [m]	$\lambda$
$\mathbf{Y}_1$	0.0	1.0	0.32	0.24	0.44	1.0	33.0	7.5
$\mathbf{Y}_1$	0.0	5.0	0.0	1.0	-	8.0	-	5.9

Table 7.3: Optimized parameters of a maximum normal copula from the saturated hydraulic conductivity measurements.

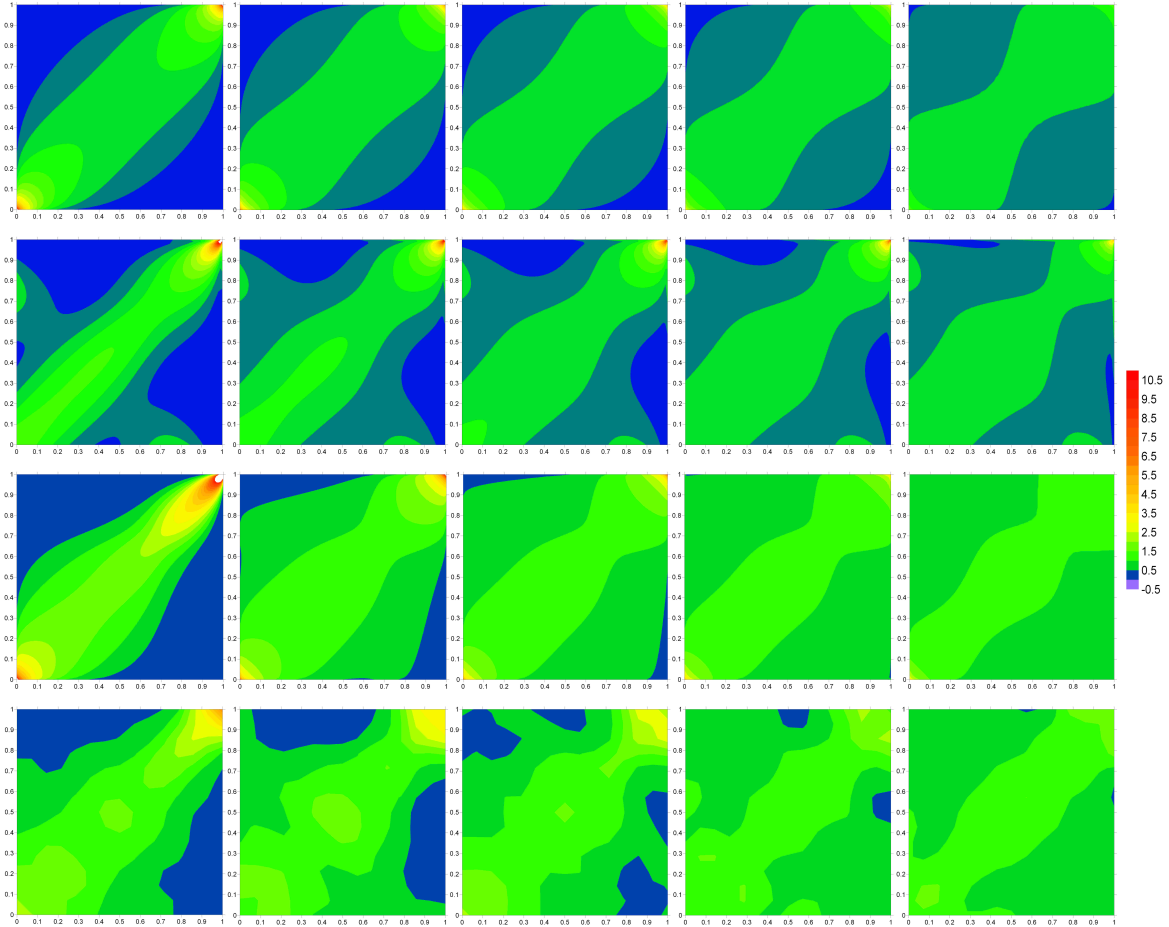


Figure 7.14: Comparison between the theoretical bivariate copulas of the three parameterized models: a Gaussian copula model (first row), a  $v$ -transformed normal copula model (second row), a maximum normal copula model (third row), and the empirical copulas along the horizontal direction of the saturated hydraulic conductivity field.

asymmetry values can be calculated by Equation (3.6) and Equation (3.10), respectively. The left panel of Figure 7.15 shows the rank correlation over distance curves calculated from the three models and compared with that from the dataset along the horizontal direction. While the right panel of Figure 7.15 shows the asymmetry over distance curves calculated from the three models and compared with that from the dataset along the horizontal direction.



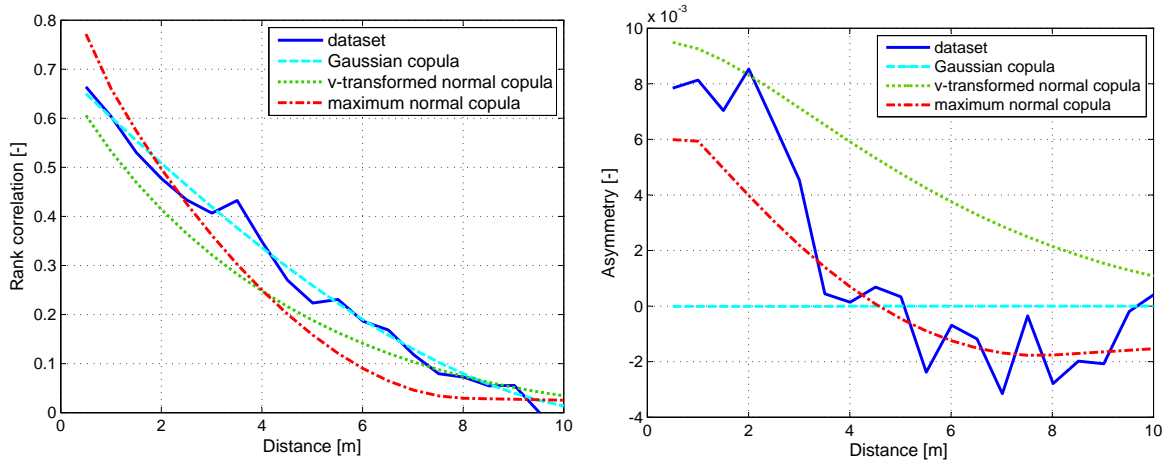


Figure 7.15: Comparison between the theoretical rank correlation (left) and asymmetry (right) over distance structures of the three parameterized copula models and the empirical ones of the saturated hydraulic conductivity field along the horizontal direction.

From the rank correlation curve fittings, it can be observed that the Gaussian copula matches the empirical rank correlation structure the best. While the v-transformed normal copula underestimates slightly for most distances. The maximum normal copula overestimates for the short distance and underestimates for the long distances, but still performs quite well in the sense of the integral scale.

The comparison of the asymmetry over distance structure shows that the Gaussian model underestimates the asymmetry dramatically. While the v-transformed normal copula overestimates the asymmetry, especially for the long distances. The maximum normal copula underestimates for short distances. However from the shapes of curves, the maximum normal copula resembles the dataset the best. In addition, like the empirical asymmetry over distance structure, the sign of the asymmetry of the maximum normal copula model is also reversed after a certain distance is exceeded (a property mentioned in Chapter 4).

The vertical statistical measures can also be calculated for the theoretical model, given the ratio ( $\lambda$ ) of the horizontal range to the vertical range. But as mentioned before, the vertical resolution of the measured field is too low to derive reasonable spatial statistics, instead the omnidirectional empirical statistics are calculated. Therefore it makes more sense to calculate the theoretical omnidirectional statistical measures for the purpose of comparison. However, the omnidirectional statistics cannot be calculated directly for the fitted theoretical models, because the ratio of horizontal range to the omnidirectional range is unknown. The following algorithm is applied to approximate the theoretical omnidirectional statistics.

Since a geometric anisotropy is assumed for the variable  $Z(s)$  under study, a linear transformation  $T$  of coordinates which is a combination of axis rotations and stretchings, can be applied such that in the new coordinate system,  $Z(s') = Z(Ts)$  is isotropic [54]. The existence of such a transformation implies that the values of the sills are the same for all directions and ranges corresponding to different directions can be plotted on an ellipse [7]. The transformation involves rotating the axes to align them with the directions of maxi-

mum and minimum range/correlation length, i.e., the main and minor axes of anisotropy, and stretching the axis along direction of the minimum range/correlation length to bring the range/correlation lengths the same in all directions (See Figure 7.16 and Equation (7.11)).

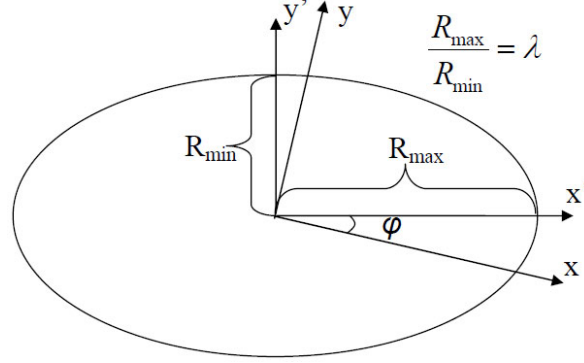


Figure 7.16: Rotation of coordinate system (new system is denoted by ' ).

$$\begin{aligned} x &= \cos(\varphi)x + \sin(\varphi)y \\ y &= -\lambda\sin(\varphi)x + \lambda\cos(\varphi)y \end{aligned} \quad (7.11)$$

where, as Figure 7.16 indicates,  $\varphi$  is the angle between the  $x$  coordinate and the main axis of the anisotropy.  $\lambda$  is the ratio of the orthogonal ranges representing the highest and the lowest correlation structure. Since the  $x$  coordinate of this dataset overlaps with the main axis of the anisotropy, i.e., the horizontal direction,  $\varphi$  equals to 0, only a stretching of the  $y$  axis is necessary. The factor of stretching is equivalent to the parameterized  $\lambda$ . The omnidirectional statistics, i.e, the rank correlations and measure of asymmetry are approximated as the average of those statistics along all the possible directions. For this purpose, the quadrant of the ellipse is divided into  $n$  sectors with equal central angles (see Figure 7.17). The representative length of the vectors falling into each sector which corresponds to a certain separation length used for rank correlation or asymmetry calculation along the horizontal direction is calculated as:

$$h' = \sqrt{(h\cos(\theta))^2 + (\lambda h\sin(\theta))^2} \quad (7.12)$$

where  $h$  refers to the separation length along the horizontal direction and  $h'$  denotes the corresponding omnidirectional lengths after coordinate transformation. The  $\rho$  values of the representative vectors of all the sectors can be calculated using the parameterized variogram model parameters along the horizontal direction. Those  $\rho$  values are inserted into the theoretical bivariate copula functions. Then using Equation ((3.6) and Equation (3.10), the theoretical rank correlations and asymmetries can be calculated. The average of resultant rank

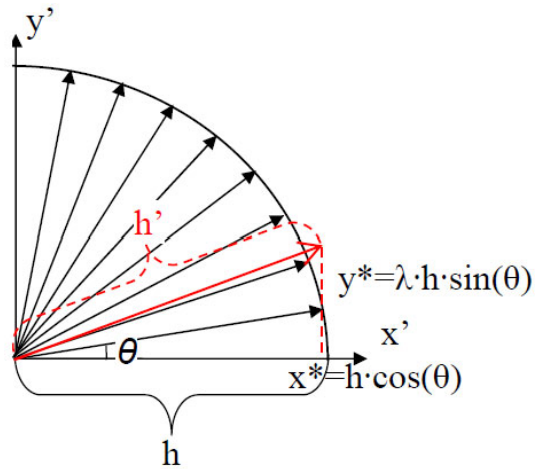


Figure 7.17: Discretization of the quadrant to calculate the omnidirectional separation lengths in the new coordinate system.

correlations and asymmetries over all the sectors are computed as the approximation for the omnidirectional rank correlation and asymmetry of a certain separation length  $h$ . The left panel of Figure 7.18 shows the theoretical omnidirectional rank correlation over distance curves of the three copula models compared with the empirical one. The right panel of Figure 7.18 shows the theoretical omnidirectional asymmetry over distance curves of the three copula models compared with the empirical one.

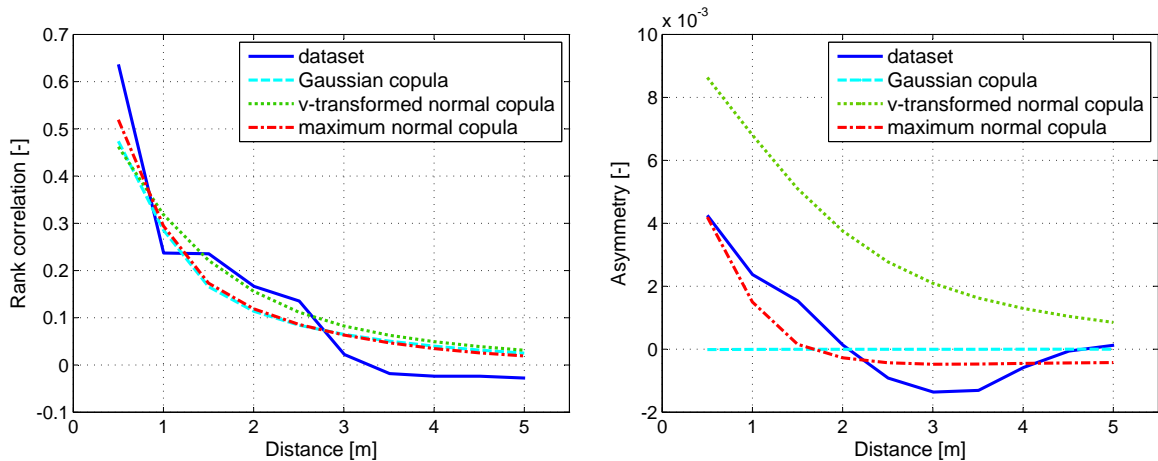


Figure 7.18: Theoretical omnidirectional rank correlation over distance curves (left) and theoretical omnidirectional asymmetry over distance curves (right) compared with the empirical ones.

Figure 7.18 shows that the maximum normal copula model fits the omnidirectional rank correlation structure slightly better than the other two models. However, a underestimation for the very short distance and an overestimation for the long distances can be observed

for all the three models. For the omnidirectional asymmetry over distance structure, the Gaussian model significantly underestimates, while the  $v$ -transformed normal copula significantly overestimates. The maximum normal copula fits the omnidirectional asymmetry over distance structure the best among all the three.

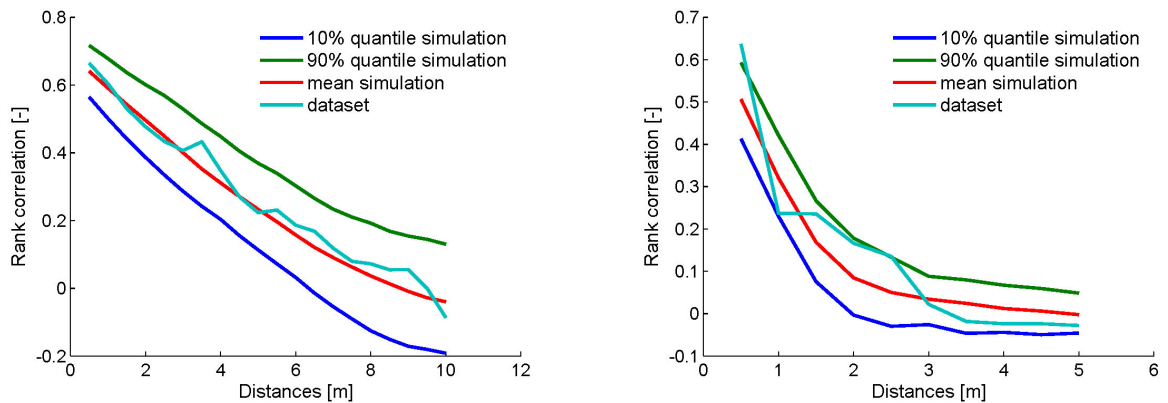
As another way of goodness-of-fit test of the applied models, i.e. statistical hypothesis tests are done based on unconditional simulations of the three parameterized copula models. The reason of using unconditional simulations but not conditional simulations is that the observation is quite dense which can cause the realizations of different random functions look very similar to each other, irrespective of the copula or other characteristic of the random function model. The point of this thesis, however, is the comparison of results obtained from random functions with different spatial copulas. This point is better demonstrated if not blurred by the effects of conditioning [51].

For this purpose, 100 unconditional simulations are run for each copula model and the upper and lower 10 percent rank correlations and asymmetry values corresponding to the required separation lengths are calculated from the 100 realizations for both the horizontal direction and omnidirection. Figure 7.19 shows the upper and lower 10 percent rank correlation over distance curves of the horizontal direction (left column) and omnidirection (right column) for the three models and compared with the empirical ones.

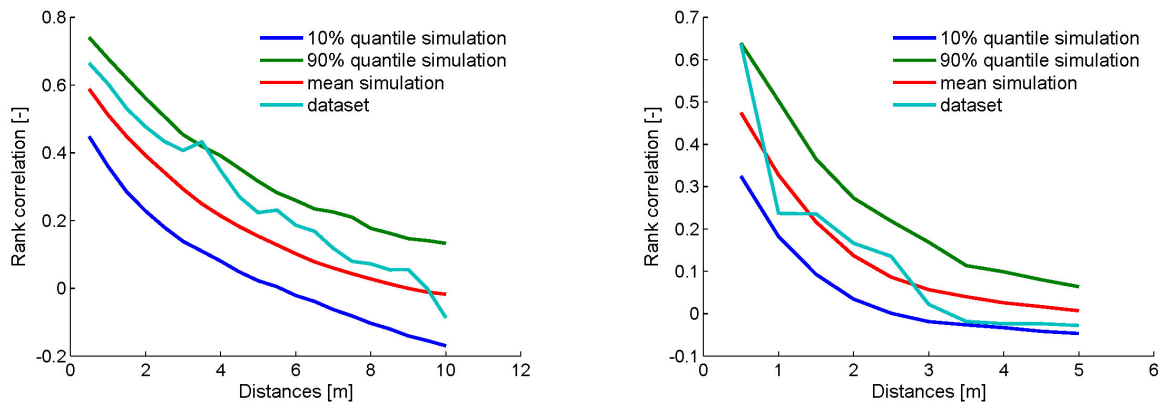
Figure 7.20 shows the upper and lower 10 percent asymmetry over distance curves of the horizontal direction (left column) and omnidirection (right column) for the three models and compared with the empirical ones.

From the averaged behaviour of the 100 realizations (red lines), one can see that the prediction from the calculations of theoretical rank correlation and asymmetry is accurate. The empirical rank correlation over distance curve is almost encompassed in the 80% confidence intervals which are constructed by the lower and upper 10% values for all the three models. For the very short distance, the upper bound of the 80% confidence interval from the Gaussian copula is smaller than the empirical one for the omnidirection. While the lower bound of the 80% confidence interval from the maximum normal copula is bigger than the empirical value along the horizontal direction. The  $v$ -transformed normal copula yields a wider confidence interval of the rank correlation structure than the other two models and thus encloses the empirical correlation curve for both cases, however, at the cost of loss of precision.

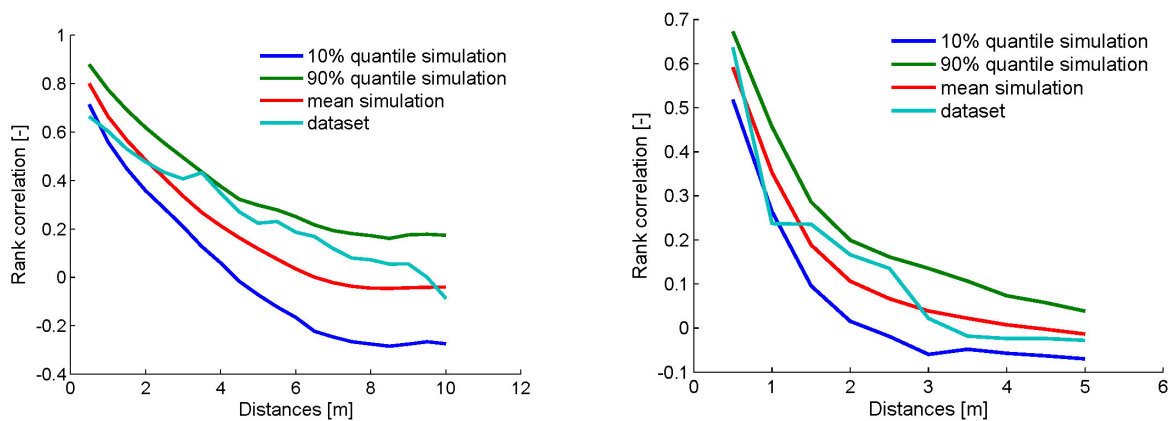
For the asymmetry over distance curves, the 80% confidence intervals of Gaussian copula fails to enclose the empirical values. Significant underestimations for the short distances can be observed. The  $v$ -transformed normal copula is able to enclose the empirical asymmetry curve for most distance classes. However, from the trend of the mean behaviour and those of the two boundaries, the difference from the empirical one is striking. Moreover, the empirical asymmetry curve is very close to the lower boundaries and even below them for some long distance classes. The maximum normal copula is the only one which fully encloses the empirical asymmetry structure within its 80% confidence intervals. The trend of the mean behaviour and of the two boundaries fits the empirical structure the best among the three models, though the shape is still flatter than the empirical one.



a). from Gaussian copula



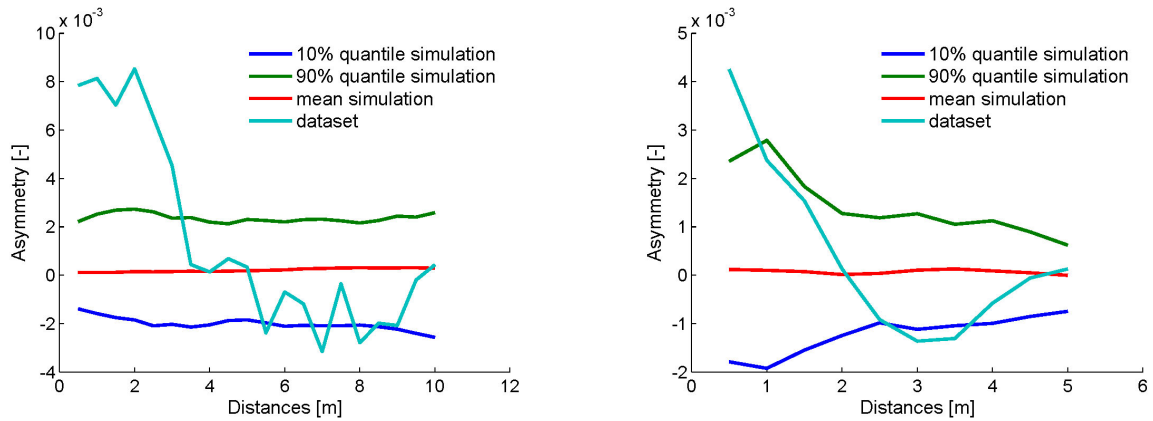
b). from v-transformed normal copula



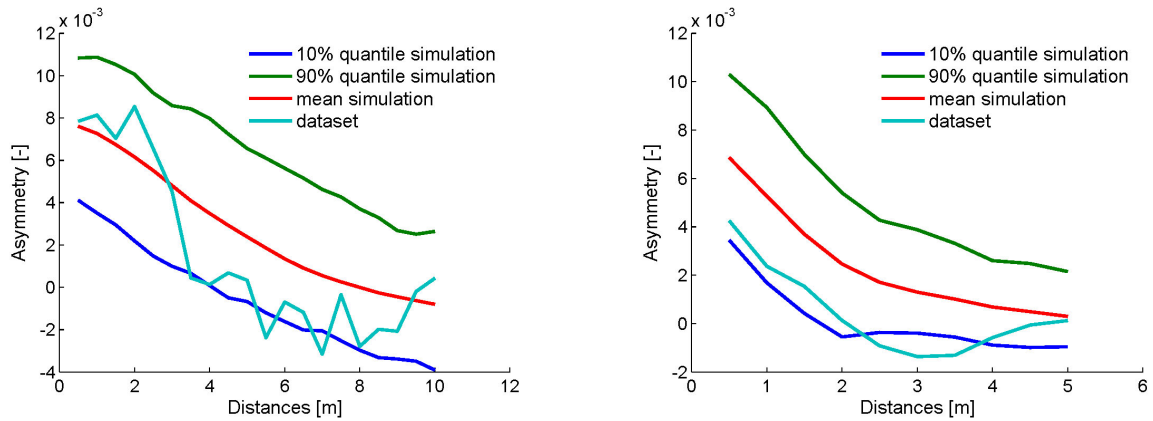
c). from maximum normal copula

Figure 7.19: Statistical hypothesis test of rank correlation structure for the horizontal direction (left column) and the omnidirection (right column).

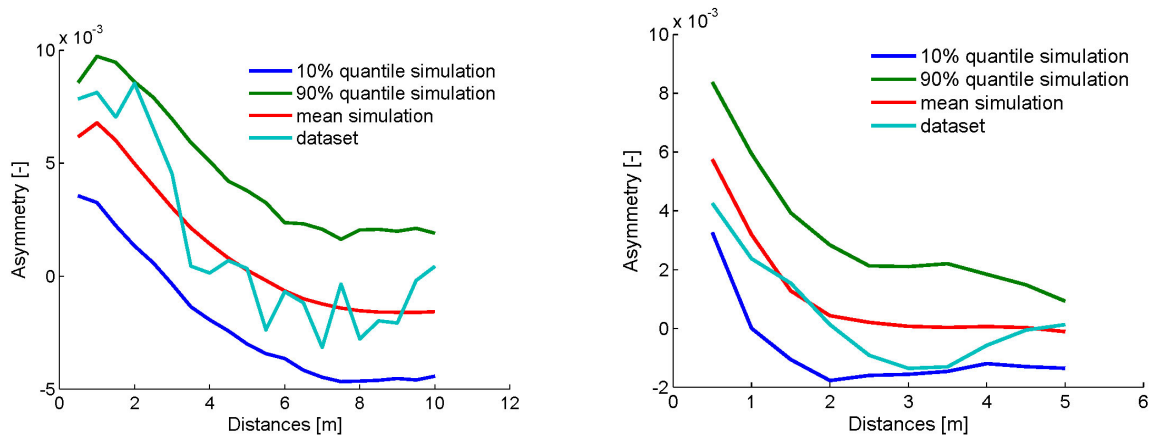
It can be concluded from the test that if a significance level of 20% is selected, based on the chosen statistical measures, the Gaussian copula model is grounded to be rejected, while the



a). from Gaussian copula



b). from  $v$ -transformed normal copula



c). from maximum normal copula

Figure 7.20: Statistical hypothesis test of asymmetry over distance structure for the horizontal direction (left column) and the omnidirection (right column).

maximum normal copula is most unlikely to be rejected.

In addition, the agreement between the conclusions drawn from the statistical tests and



those drawn from the comparison between statistical measures calculated from the theoretical copula models and the empirical ones demonstrates that the developed theoretical spatial copulas and the simulation procedure are well established.

Figure 7.21 show the realizations from the three copula models (top one is from the Gaussian model, middle one from the  $v$ -transformed normal copula and bottom one from the maximum normal copula) with similar omnidirectional rank correlation structures (see Figure 7.22).

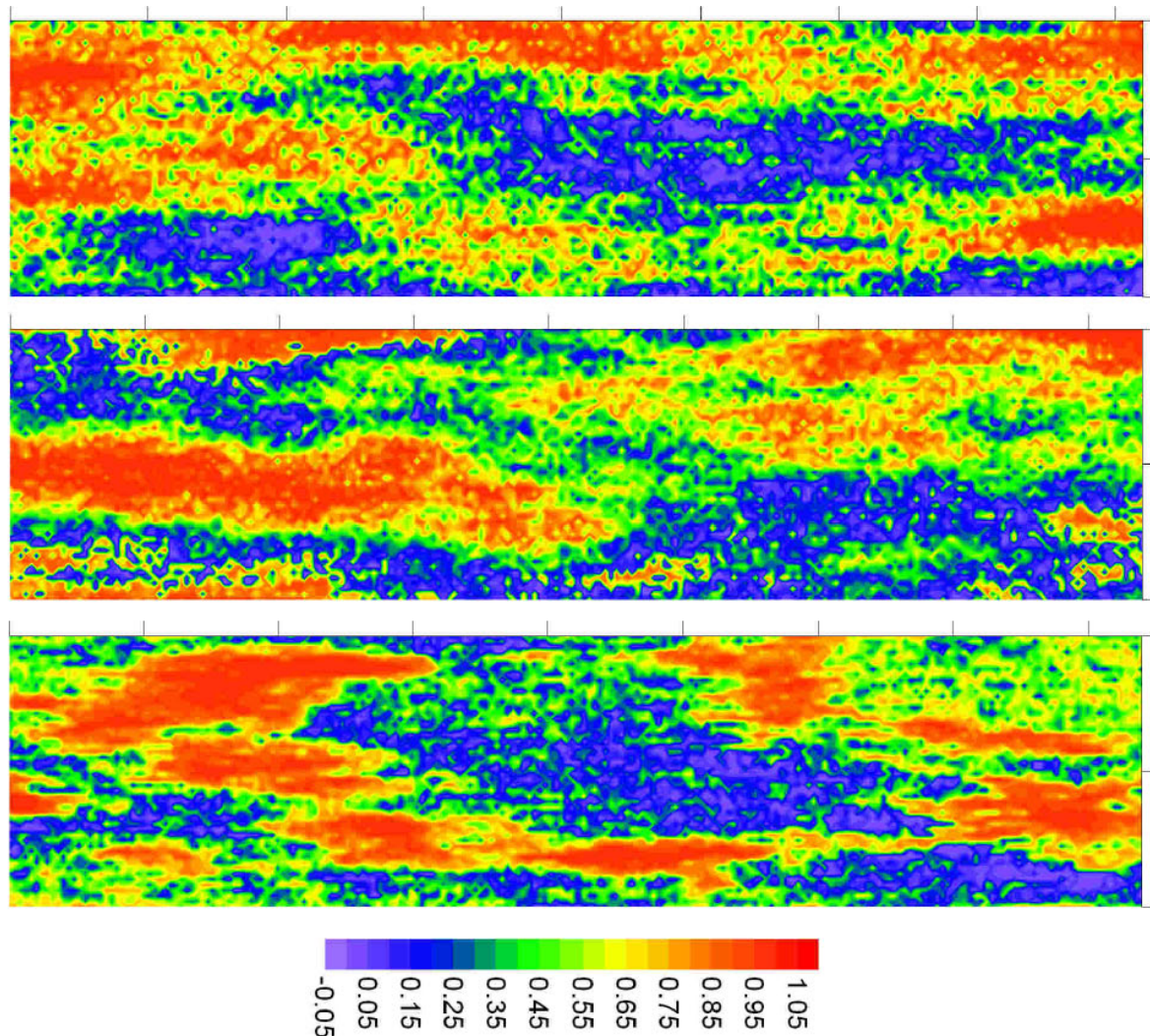


Figure 7.21: Comparison of realizations from the three copula models. Top one is from the Gaussian model, the middle one is from the  $v$ -transformed normal copula and bottom one is from the maximum normal copula.).

The differences of the spatial structures shown in Figure 7.21 can be detected. For the Gaussian realization, similar big structures are observed for the high quantiles (red) and low quantiles (blue). The overall structure looks very tortuous with a lot of redundant connec-

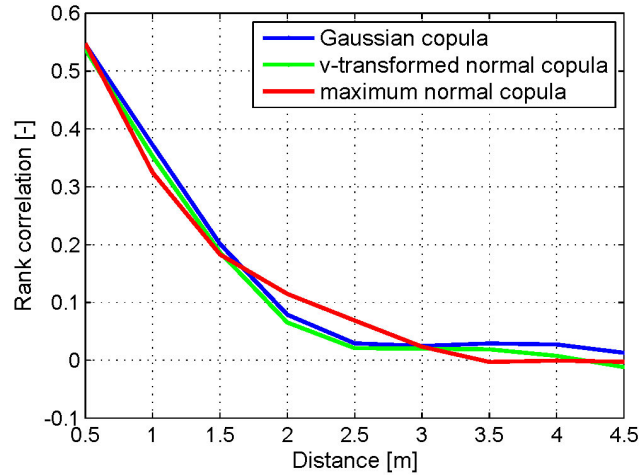


Figure 7.22: Rank correlation over distance curves of the realizations shown in the above figure.).

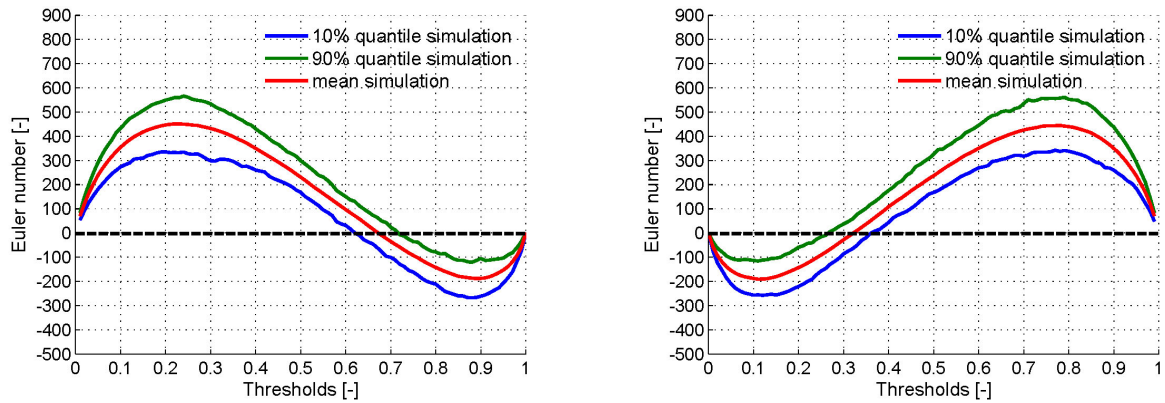
tions. Connected paths spanning the whole domain can be found for the middle quantiles (green). For the v-transformed normal copula realization, the high quantiles are more clustered than the low quantiles. Compared with the Gaussian realization, the pattern of middle quantiles look less tortuous and are discontinuous in some region. In addition, sharp transitions between the high quantiles and low quantiles can be observed in some parts of the field. Contrast to the other two realizations, the realization from the maximum normal copula exhibit most smoothed spatial structures of high quantiles. Between the bulks of the high quantiles, some fingering like connections are present. The low quantiles also tend to form big structures. But they are less smooth and have more tortuous patterns than the bulks of the high quantiles.

The investigation of the dependence measures and the statistical tests manifested clear differences in the spatial dependence of the realizations from the three models. But whether these differences are significant for flow and transport is still a question. To answer this question, in fact one should carry out the flow and transport simulations on the generated spatial realizations and the subsequent Monte Carlo analysis. However, this is not within the scope of this thesis. Therefore another easier and indirect test is performed aiming to shed a light on the question. As *Hans-Jörg Vogel* demonstrated in his paper [87] that hydraulic conductivity fields sharing the same covariance structure and histogram but differing in their topological/connectivity property lead to distinguish effective permeabilities and flow behaviour. The topological/ connectivity property of the spatial fields are described by the Euler characteristics of the fields.

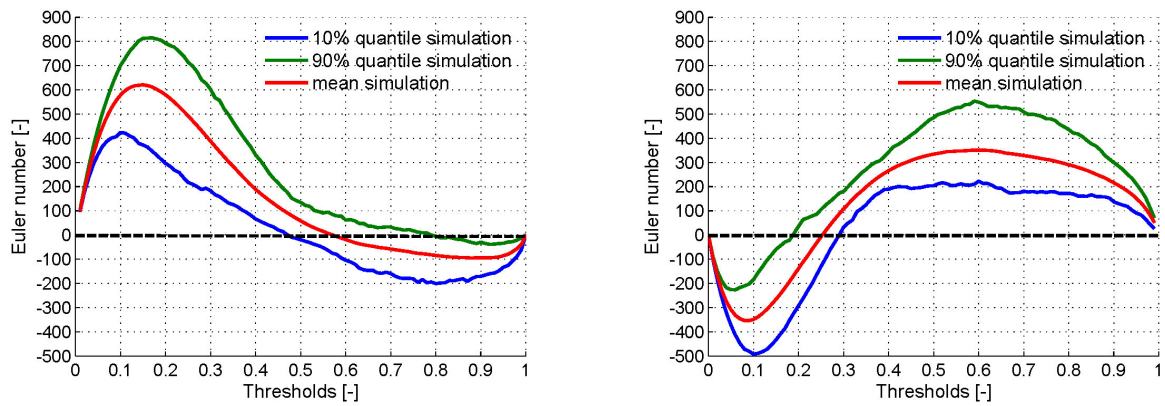
In Chapter 3, the concept of using Euler characteristics to quantify the topological property of porous media has already been reviewed. In the following, the measure of Euler's number as a function of threshold value is studied for the 100 unconditional realizations from the three copula models. Figure 7.23 shows the 80% statistical boundaries and the mean



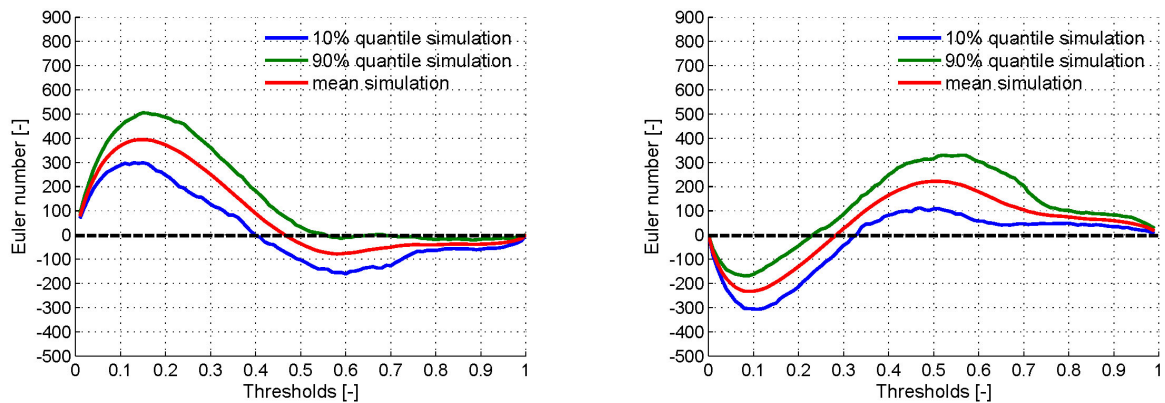
behaviour of Euler's characteristic curves. The left column corresponds to the calculation starting from the smallest threshold value, while the right one corresponds to the calculation starting from the biggest threshold value.



a). from Gaussian copula



b). from v-transformed normal copula



c). from maximum normal copula

Figure 7.23: 80% statistical boundaries and the mean behaviour of Euler's characteristic curves calculated from 100 realizations of the three copula models.

The results shown in Figure 7.23 indicate that when starting from the small threshold end, the realizations of the maximum normal copula start to be connected first. While starting from the big threshold end, the Gaussian realizations start to be connected first. In addition, the two charts for the Gaussian realizations are just reflections of each other, indicating the same connectivity properties of the low and high quantiles. To the contrast, the left and right charts for the realizations of the other two models are not symmetrical at all. The charts for the v-copula realizations show that the bigger values have thicker bands (the range of the thresholds where the curve goes below zero) which are connected than the small values, while the small values have more redundant connections (the maximal magnitude of the curve below zero). This is also true for the maximum normal copula realizations. But the curves from the maximum normal copula realizations are much flatter (smaller maximal magnitude of the curve) than those from the v-copula realizations, indicating a smoother behaviour of the spatial structures. Among all the three models, the v-copula realizations yield the widest statistical boundaries which may also results in bigger uncertainty in the flow and transport estimations.

In a word, the study of the Euler characteristics demonstrate the clear differences in the topological/connectivity properties of different model realizations. According to the conclusion drawn from the research conducted by *Hans-Jörg Vogel* [87] that the topological characteristic of the porous media has a significant influence on the flow and transport behaviour, the realizations generated from different copula models should also lead to different flow and transport results.

## 7.5 Summary of Chapter 7

In this chapter, simulation approaches for generating realizations of multivariate non-Gaussian dependence are developed for both unconditional and conditional cases. New spatial analysis tools, i.e., empirical copulas, measure of asymmetry and Euler's characteristics have been applied to investigate the spatial structure of hydraulic conductivity fields. Geostatistical analysis of the in-situ measured hydraulic conductivities of the Las Cruces Trench Site show that the spatial field exhibit anisotropic, non-Gaussian dependence. Three theoretical copula models: Gaussian copula, v-transformed normal copula and maximum normal copula, are parameterized based on the dataset using the parameterization method described in Chapter 5. One hundred unconditional realizations are drawn from each model for statistical hypothesis test to check the goodness-of-fit of the model, results of which indicate that the Gaussian copula is most likely to be rejected, while the maximum normal copula is the most suitable model as the prior random function of the dataset among the three selected models. The study of the Euler characteristic of the realizations from the three copula models shows the evidence that the spatial dependence described by copulas has a significant influence on the connectivity property of the resultant realizations and thereby on the flow and transport behaviour.

## 8 Observation Network Design

Statistical methods can be applied to evaluate compliance with regulatory standards by means of providing a reasonable degree of assurance that certain criteria is met. In this context, using statistics one can estimate the impact of additional measurements on error reduction before the measurement are taken in order to decide where to collect additional measurements so that the objectives of *monitoring* are met in the most cost-effective way. *Kitanidis* [1997] [54]

In this chapter, a method for environmental network design using the framework of spatial modeling with copulas is introduced. This methodology has been published in *Proceedings of VIII International Geostatistics Congress* [12]. The context of this chapter is similar to the published paper. Some improvement and modifications have been done.

### 8.1 The Problem of Network Design

A monitoring network is usually required to measure specific environmental variables and to reduce the uncertainty about the variables of interest. Different applications usually require different designs. For example to measure contaminant concentrations, it is important to take into account the location of point sources or hot spots contributing to these concentrations. In the epidemiological applications, specialists are mostly concerned with areas of higher disease incidences and will try to separate the monitoring of these locations from locations with low risk.

Different design strategies are available in the literature, as probability based methods including random sampling or stratified sampling. These designs can be very inefficient since sample sites might be adjacent to each other, creating information redundancy. Model based approaches are usually known as regression based approaches (*Federov and Müller* [1988] [31], *Federov and Müller* [1989] [32]) and random field model approaches.

For the regression based approaches, an underlying regression model of the form  $y(x) = \eta(x, \beta) + \varepsilon(x)$  where  $y$  represents the variable of interest,  $x$  represent the selected measurement points,  $\eta(x, \beta)$  is a known non-linear function which depends on  $x$  and unknown parameters  $\beta$  and  $\varepsilon(x)$  is assumed independent for each  $x$ . The optimality criterion is defined in terms of the efficiency in the estimation of  $\beta$ , and the selection of  $\eta$  can be also difficult specially in moderate size problems.

The random field model approaches are usually linked to Geostatistics which focuses on the prediction of an unobserved quantity of a spatial random field regarded as a realization of

a random function. It entails a good understanding of the spatial dependence. Traditional geostatistical tools such as the variogram and the covariance function are normally used to quantify this spatial dependence. Since in many occasions, a single realization of the spatial fields is available, weak or strong stationarity assumptions have to be made for the analysis. The variogram/covariance is defined as a function of the separation vector among sampling locations. However the spatial dependence might also vary for the different quantiles of the marginal distributions of the observations, and this behavior is not appropriately considered with standard geostatistical tools, in which dependence is solely described by the variogram/covariance function only.

Changes in the spatial dependence for different distribution quantiles might have important consequences for network design, especially for extreme events detection. For example, for a water or air quality parameter, a particular network design might be adequate to monitor mean trends but a denser network might be required to detect noncompliance values which may affect human health. This problem has been discussed by *Chang and Zidek* [2007] [16]. However, the variogram/covariance function is only able to describe the dependence in the averaged sense. This gives the genesis of the proposed approach, in which copulas is used to describe the spatial dependence. Thanks for the properties of copulas, the advantage of this approach is that it does not only consider the spatial configuration of the gauged and ungauged locations, but it also considers the observed values at the gauged locations and their influence on the potential values at unobserved locations. Thus the uncertainty in the estimation at an unobserved location which might differ depending on whether neighboring observations correspond to high quantiles or low quantiles is better quantified.

The first step in this approach is to consider a random field in where variations in time can be neglected and the variable of interest is static in nature. The spatial dependence of the natural variable under interest is modeled using copulas. The uncertainties in the estimation of unsampled locations are described with the help of the conditional copula. Then a utility function which is weighted by the conditional copula of each possible state of nature, is developed for decision optimization in where false positives or negatives are penalized with a certain cost, while right decisions are favored with a certain gain.

In the following sections, after the description of the methodology, a synthetic example is presented to better explain the methodology. Then a real application on water quality parameters in the south-western region of Germany is discussed to show its potentialities.

## 8.2 Description of the proposed methodology

We denote  $Z(\mathbf{s})$  a spatial random field of interest indexed by location  $\mathbf{s}$  in an  $n$  dimensional space and  $\mathbf{s} \in S$ , where  $S$  is the study domain. In a purpose oriented network design, we might be interested in measuring at locations where the value of the variable  $Z(\mathbf{s})$  is above a certain threshold  $\beta$  such that:

$$\theta(\mathbf{s}) = \begin{cases} \theta_0(\mathbf{s}) & Z(\mathbf{s}) < \beta \\ \theta_1(\mathbf{s}) & Z(\mathbf{s}) \geq \beta \end{cases} \quad (8.1)$$

where  $\theta(\mathbf{s})$  is a dichotomic spatial random field representing the state of nature at each location  $\mathbf{s}$ .

From the point of view of decision theory, we propose a very simple approach by using decision theory tools. At any location  $\mathbf{s}$ , we can take the decision  $d_i$  of taking water or not depending on our judge on whether the unknown location gets a value below the threshold  $\beta$  or not. The positive decision (taking water) is denoted as  $d_0$ , and the counterpart is denoted as  $d_1$ . We can define a utility function  $U_s(\theta_i, d_i)$ , depending on the state of nature  $\theta_i(\mathbf{s})$  at each location and the decision to be taken. The term *utility* is a measure of the relative satisfaction from consumption of various goods and services or attainment of goals [44]. The defined utility function for our case is presented in Table 8.2.

$U_s(\theta_i, d_i)$	$\theta_0$	$\theta_1$
$d_0$	$k_{00}$	$k_{01}$
$d_1$	$k_{10}$	$k_{11}$

Table 8.1: Utility matrix

The entries of coefficient  $k_{ij}$  can be defined in relative units. Negative  $k_{ij}$  values represent costs, while positive ones represent gains. Note that the diagonal elements  $k_{00}$  and  $k_{11}$  correspond to gains if a correct decision is taken, while the off diagonal ones correspond to losses when a wrong decision is made. It should be mentioned that the gains and costs can be very different for different cases. For instance, exceeding a threshold of a certain pollutant concentration in groundwater can have a severe adverse effect if used for drinking purpose. On the other hand, to decide not to use the water, even if in reality, the threshold is not exceeded leads, generally, to a relatively small loss. But in the case of gold mining, the reverse holds true, i.e., the loss of deciding not mine but actually there is gold might be higher than deciding to mine but no gold is present.

The expected utility at each location  $\mathbf{s}$  for a certain decision  $d_i$  can then be calculated with the help of the coefficients of utility function as:

$$E(U_{\mathbf{s}}|d_i) = k_{i0}p(\theta(\mathbf{s}) = \theta_0) + k_{i1}p(\theta(\mathbf{s}) = \theta_1) \quad i = 0, 1 \quad (8.2)$$

The decision should be taken based on the estimated probability  $p = p(\theta(\mathbf{s}) = \theta_0(\mathbf{s}))$ . If it exceeds a certain limit  $p_l$  then  $d_0$  is taken, else  $d_1$  should be taken. That means if the probability of pollutant concentration or gold grade below the critical threshold is big enough so that the risk of taking water is sufficiently low or it is not worth mining, the decision  $d_0$  is chosen and vice versa. The aforementioned terms *positive decision* and *negative decision* are relative to the specific design purpose. In the project of water quality control, if  $p = p(Z(\mathbf{s}) < \beta)$  is greater than a certain limit, the positive decision/ taking water will be taken. While in the project of gold mining, in this case the negative decision/ not mining will be made. Since we are mainly dealing with environmental variables, in the following, we will comply with the definition for the water quality control case.

The value of the limit  $p_l$  should be specified according to the principle that the expected utility is maximized. That means, if  $p(\theta(\mathbf{s}) = \theta_0(\mathbf{s})) > p_l$  then the expected utility of the positive decision is greater than that of the negative decision:

$$k_{00}p + k_{01}(1 - p) > k_{10}p + k_{11}(1 - p) \quad (8.3)$$

else if  $p(\theta(\mathbf{s}) = \theta_0(\mathbf{s})) < p_l$  then:

$$k_{00}p + k_{01}(1 - p) < k_{10}p + k_{11}(1 - p) \quad (8.4)$$

For  $p(\theta(\mathbf{s}) = \theta_0(\mathbf{s})) = p_l$ , the utilities of the two opposite decisions should be equivalent:

$$k_{00}p + k_{01}(1 - p) = k_{10}p + k_{11}(1 - p) \quad (8.5)$$

which leads to:

$$p_l = \frac{k_{11} - k_{01}}{k_{00} - k_{01} - k_{10} + k_{11}} \quad (8.6)$$

Note that a continuous utility function depending on the exact value could also be defined. In order to take the appropriate decision, the probabilities  $P(Z(\mathbf{s}) < \beta)$  at an unsampled location have to be estimated from the available observations. This can be done by using appropriate interpolation procedures of the random field  $Z(\mathbf{s})$  at a given set of unsampled locations denoted by  $\mathbf{S}^* = (s_1^*, \dots, s_N^*)$ . For this purpose, the approach described in Chapter 6 is applied. The conditional copula at the unsampled locations  $C_{s_j^*, n}(u_j^* | u_1, \dots, u_n)$  for  $j = 1, \dots, N$ ,  $u_j^* = F_Z \beta$ , conditioned on the  $n$  observations is calculated as the probability  $P(Z(\mathbf{s}) < \beta)$ , i.e.,  $p(\theta(\mathbf{s}) = \theta_0)$  required by Equation (8.2) at the unsampled location.

Suppose now that we have a set of locations  $\mathbf{S}' = (s'_1, \dots, s'_m)$  which is a subset of the unsampled locations  $\mathbf{S}^*$ , as candidates for the new measurement. We will select one location from the set  $\mathbf{S}'$  which, among the others, brings the biggest utility once it is included into the observation network. Since the concept of copulas is adopted in the spatial interpolation process, the decision of adding any new measurement location into the existing network depends not only on the spatial configuration of the observing set and the spatial location of the new candidate but also on the given values of the spatial variable at the gauged locations. In this procedure, the observed values have to be first transformed into the copula space with the appropriate copula function.

Let  $s'_i$  be any of the new location candidates for some  $i = 1, \dots, m$ . We can calculate the conditional copula :

$$C_{s'_i, n} = C(u'_i | U_1 = u_1, \dots, U_n = u_n) \quad (8.7)$$

at the location  $s'_i$  conditioned on the  $n$  existing observations according to Equation (2.11). Note that  $U$  refers to the uniform random variable on  $[0, 1]$  transformed from the original variable under study.  $u'_i$  denotes the value of the uniform variable at the candidate location

$s'_i$  which ranges from 0 to 1. Thus the full conditional distribution at  $s'_i$  can be obtained by calculating the conditional copula for all the possible values of  $U$  at  $s'_i$ .

By adding a candidate to the existing observation set, the interpolated values on the interpolation grid can be re-estimated. The conditional copula at the any interpolation location  $s_j^*$  can be calculated using the old observations and the newly added candidate. For any possible value  $u'_i$  of  $U$  at the candidate location, the conditional copula at any  $s_j^*$  is:

$$C_{s_j^*} = C(u_{s_j^*} | u'_i, u_1, \dots, u_n) \quad (8.8)$$

where  $u_{s_j^*}$  refers to any possible value of the uniform variable  $U$  at  $s_j^*$ . If  $u_{s_j^*}$  is corresponding to the threshold  $\beta$ , i.e.,  $u_{s_j^*} = F_Z(\beta)$  with  $F_Z$  indicating the empirical marginal distribution of the spatial variable  $Z$  under study, then the expected utility at  $s_j^*$  after adding the candidate  $s'_i$  with a certain  $u'_i$  value as a new observation can be computed by:

$$E[U_{s_j^*} | u'_i] = \begin{cases} k_{00}C_{s_j^*} + k_{01}(1 - C_{s_j^*}) & \text{if } C_{s_j^*} > p_l \\ k_{10}C_{s_j^*} + k_{11}(1 - C_{s_j^*}) & \text{if } C_{s_j^*} \leq p_l \end{cases} \quad (8.9)$$

The expected total utility of the whole interpolation grid after adding the candidate  $s'_i$  as a new observation is then the integration of the sum of the utility at every interpolation point described in Equation (8.9), over all the possible values of  $u'_i$  of the uniform variable  $U$  at the candidate location  $s'_i$ . The mathematical expression is described as follows:

$$E_T = \int_0^1 \sum_{j=1}^N E[U_{s_j^*} | u'_i] dC_{s'_i, n} \quad (8.10)$$

Equation (8.9) and (8.8) have to be inserted into Equation (8.10). The candidate  $s'_i$  which belongs to the proposed set  $S'$  and yields the biggest total utility  $E_T$  is to be sought. To summarize this procedure, the following steps are listed for clarification.

1. Transform the existing observations  $z_1, \dots, z_n$  into cumulative probability values  $u_1, \dots, u_n$ .
2. For each candidate location  $s'_i$  in the set  $S'$ , calculate the conditional copulas for all possible values of  $u'_i$  of the uniform variable  $U$  at  $s'_i$  conditioned on the  $n$  existing observations. Practically, the range of  $U$ , i.e.,  $[0,1]$ , is divided into hundreds of equal intervals and the conditional copula is calculated for the mid value of each interval.
3. Calculate the conditional copula at each interpolation point in the set  $S^*$  at the  $u_{s_j^*}$  value corresponding to the threshold  $\beta$ , conditioned on the old observations  $u_1, \dots, u_n$  and the proposed value  $u'_i$  at the candidate location. This calculation has to be done for all the discretized values of  $U$  at the candidate location  $s'_i$ .
4. At each interpolation location  $s_j^*$ , calculate the expected utility using Equation (8.9) and sum up for all the interpolation points. This step has to be repeated for all the possible values of  $u'_i$  at the candidate location.

5. Calculate the integration described in Equation (8.10).
6. Repeat the steps 2 to 5 for the all the proposed candidates and select the one which produces the highest total expected utility  $E_T$ .

### 8.3 One Synthetical Example

In this section, we demonstrate the methodology with a synthetical example where the Gaussian copula and the v-transformed normal copula (see Chapter 4) are used to describe the spatial dependence of the studied domain. The main parameters of the v-transformed normal copula are:  $m = 0.0$ ,  $k = 1.0$  and  $\alpha = 1.0$ . The dependence is postively skewed which means high quantiles are more dependent than the low quantiles (see Section 3.3.2 in Chapter 3). In Figure 8.1, a spatial region where four observations are available as monitoring locations is shown. The numbers next to the four closed circles indicate the observation values as probabilities. The five open circles represent the new candidates (set  $S'$ ). The interpolation points are shown as the grid. A new measuring location will be selected from this set, which achieves the highest total expected utility  $E_T$  for the estimated random field. The value of the expected utility at each location depends on the estimated probability  $p(\theta(\mathbf{s}) = \theta_0)$  and the utility matrix.

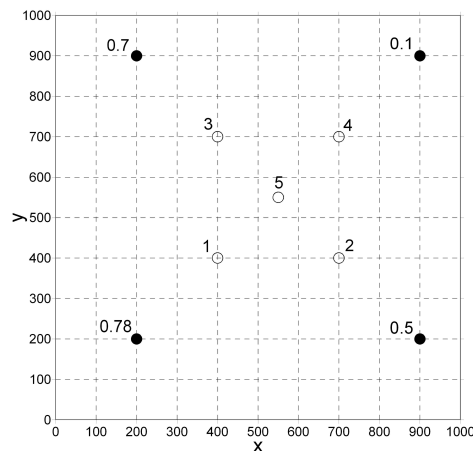


Figure 8.1: Synthetical example for network expansion. Closed symbols are monitoring locations. Open symbols are candidate locations. Grid points are new estimation locations

Steps 2 to 6 as described before were carried out for both the Gaussian copula model and the v-transformed normal copula model. Step 1 is not necessary since the initial data values are given as cumulative probabilities. The proposed entry values for the utility matrix are  $k_{00} = 0.0$ ,  $k_{01} = -2.0$ ,  $k_{10} = -1.0$ ,  $k_{11} = 0.0$ . Negative values are interpreted as costs. In this case, the cost of taking water for drinking purpose but in fact the pollution exceeds the acceptable limit, i.e.,  $\theta = \theta_1$  is twofold the cost of not taking water but the degree of pollution is within the safe range, i.e.,  $\theta = \theta_0$ .



For the specified spatial configuration, if the critical threshold  $\beta$  corresponding to the probability  $p(Z(s) < \beta) = 0.8$  is chosen, the above criteria selects candidate 1 as the best new measuring location for both copula models. Table 8.2 shows the optimization results for both models. The first column shows the ID number of the candidates. The abbreviations *pos. Deci.* and *neg. Deci.* refer to the number of positive decisions and number of negative decisions made. The grid shown in Figure 8.1 consists of 121 nodes, since the number of cutoffs of the uniform variable  $U$  at the candidate location is 100 (see step two), the total number of decisions to be made is 12100. This is, however, not clear. Therefore the maps of the percentage of positive decisions resulted from both models are shown in Figure 8.2. The maps are according to results after adding the optimal candidate to the measurements.

Cand. ID	Gaussian copula			V-transformed normal copula		
	pos. Deci.	neg. Deci.	$E_T$	pos. Deci.	neg. Deci.	$E_T$
1	11895	205	-26.46	11579	521	-40.96
2	11922	175	-27.66	11697	403	-42.20
3	11902	198	-27.13	11617	483	-41.52
4	11939	161	-28.36	11785	315	-42.59
5	11881	219	-27.22	11619	481	-41.80

Table 8.2: Results of optimization of synthetical study field with threshold  $\beta = 0.8$ , coefficients of utility function:  $k_{00} = 0.0, k_{01} = -2.0, k_{10} = -1.0, k_{11} = 0.0$ .

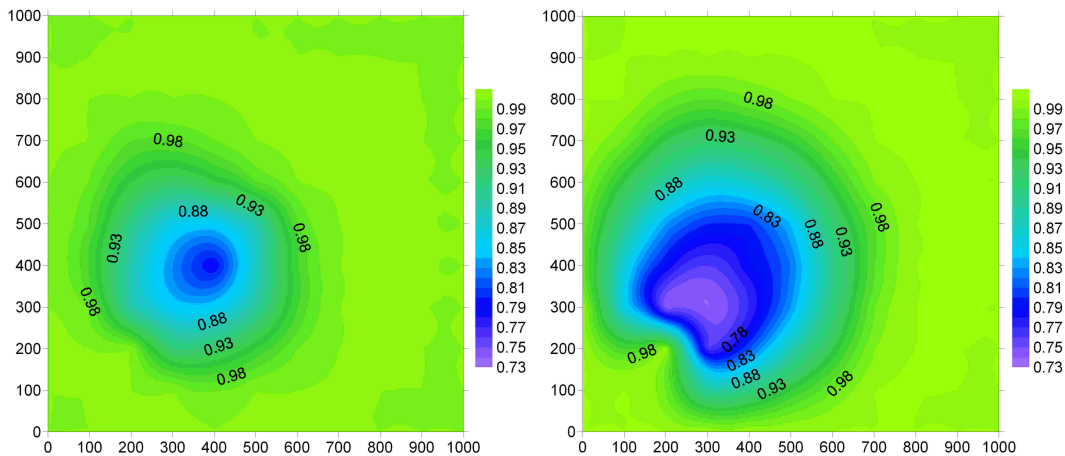


Figure 8.2: Maps of percentage of positive decisions resulted from the Gaussian copula (left) and the v-transformed normal copula (right) after adding the optimal candidate to the measurements ( $p(\theta = \theta_0) = 0.8, k_{00} = 0.0, k_{01} = -2.0, k_{10} = -1.0, k_{11} = 0.0$ ).

This result is expected since this candidate location is nearest to the measurement with the value closest to the threshold. This is the point whose surroundings are most hung in doubt. The maps in Figure 8.2 show that all four observed locations are 100% accepted as drinking water usage. This is straight forward, because all measured values are below the threshold. However, the maps show that the observed point with coordinates (200, 200) can be at the margin of a pollution source. The plume of the possible pollution source is bigger for the

v-transformed normal copula. The maps indicate that although both models result in the same optimal candidate location, the ratios between the positive decisions (using water) and the negative decisions (not using water) are different. The v-transformed normal copula yields a more conservative decision. If the Indicator Kriging is applied in this case, all the observations get the same indicator value since they are all below the threshold, and there is no information on where to measure. Therefore IK will not be a possible alternative.

If the probability of the critical threshold  $\theta = \theta_0$  is reduced to 0.7, point 1 remains as the optimal candidate for both copula models. However, the maps of the percentage of positive decisions are changed dramatically. Figure 8.3 shows the maps. The left two observed points are 100% rejected for drinking water usage. The most dangerous places are also shifted to the surroundings of these two points.

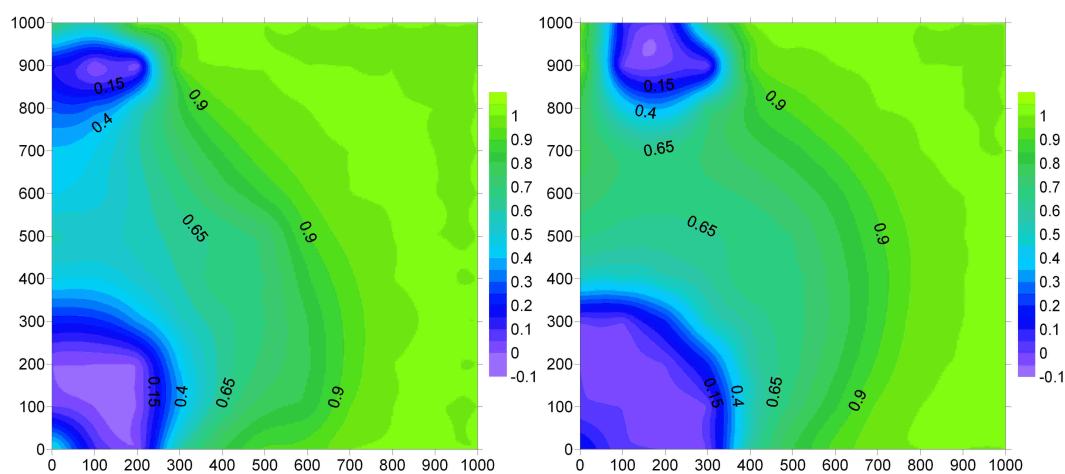


Figure 8.3: Maps of percentage of positive decisions resulted from the Gaussian copula (left) and the v-transformed normal copula (right) after adding the optimal candidate to the measurements ( $p(\theta = \theta_0) = 0.7$ ,  $k_{00} = 0.0$ ,  $k_{01} = -2.0$ ,  $k_{10} = -1.0$ ,  $k_{11} = 0.0$ ).

Further, if the coefficients of the utility function are also changed to  $k_{00} = 0.0$ ,  $k_{01} = -4.0$ ,  $k_{10} = -1.0$ ,  $k_{11} = 0.0$ , which means a higher cost of misuse of the water is imposed, the Gaussian model chooses point 3 as the optimal candidate, while the v-transformed normal copula chooses point 2. Much more conservative results are obtained for both models (see Figure 8.4).

This example demonstrated that the copula based approach is able to give information in the case that indicator approach fails. The final decision is influenced by the chosen dependence model, the critical threshold and the coefficients specified in the utility function.

## 8.4 Application to Water Quality parameters

The proposed methodology is applied to optimize the new observation locations for groundwater chloride concentration in a sub-region of Baden-Württemberg in Germany. One hundred and three points distributed uniformly over the study area are selected as new observation candidates. Five thousand and five hundred unsampled points are interpolated.

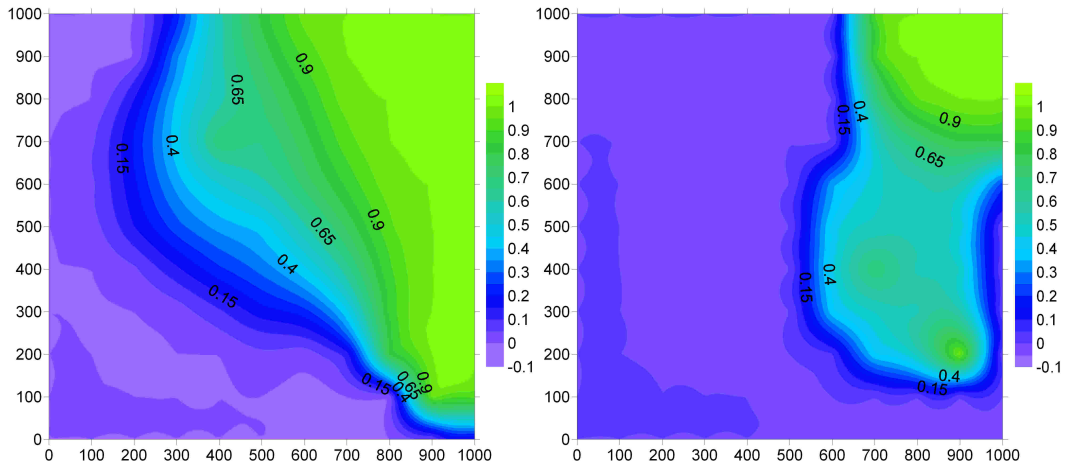


Figure 8.4: Maps of percentage of positive decisions resulted from the Gaussian copula (left) and the  $v$ -transformed normal copula (right) after adding the optimal candidate to the measurements ( $p(\theta = \theta_0) = 0.7$ ,  $k_{00} = 0.0$ ,  $k_{01} = -4.0$ ,  $k_{10} = -1.0$ ,  $k_{11} = 0.0$ ).

Figure 8.5 shows the existing measurements and the interpolation grid as well as the new measurement candidates with corresponding utility values. The upper 20% of the distribution of the existing observations in the whole region of Baden-Württemberg are considered as pollution range. In Figure 8.5, this class is represented by the six-pointed star. The corresponding threshold value is  $47 \text{ mg/l}$ . The entry values for the utility function are selected as  $k_{00} = 0.0$ ,  $k_{01} = -8.0$ ,  $k_{10} = -2.0$ ,  $k_{11} = 0.0$ . In this case the cost of not measuring when there is pollution is much higher than the opposite condition, which means a very strict restriction is imposed here.

The  $v$ -transformed normal copula is used to calculate the conditional distribution which is described in Equation (2.11). The values of the model parameters for the  $v$ -transformed normal copula are taken from the parameterization results presented in Table 6.2.

The results of the total expected utilities calculated with each candidate (see Equations (8.9)) are shown in Figure 8.5. From the figure, we can see that in the south-western part, some locations achieved high utilities. In this region, a contradiction occurs which makes the decision difficult, because the surrounding measurements are somewhere above the threshold and elsewhere far below. On the western boundary of the study area, the measured concentrations are mostly beyond the threshold and the calculated utilities for the nearby candidate points do not reach the highest level. This can be explained by the fact that from the existing observations, it is already clear that the region close to those observations is very dangerous to take water for drinking purpose. It does not make much difference for the decision if an extra observation is added in the vicinity. Moving a bit to the east, there are also some candidate points with high utilities. This is because there are several measurements whose values are very close to the threshold. Candidate locations with low utilities are in the regions either with high measurement densities or with relatively clearer implications of being below the threshold. The utilities pick up to the middle level in the south-eastern part of the domain. This can be interpreted by the fact that the measurement density is very low in this region and its neighborhood, which is not displayed in Figure 8.5.

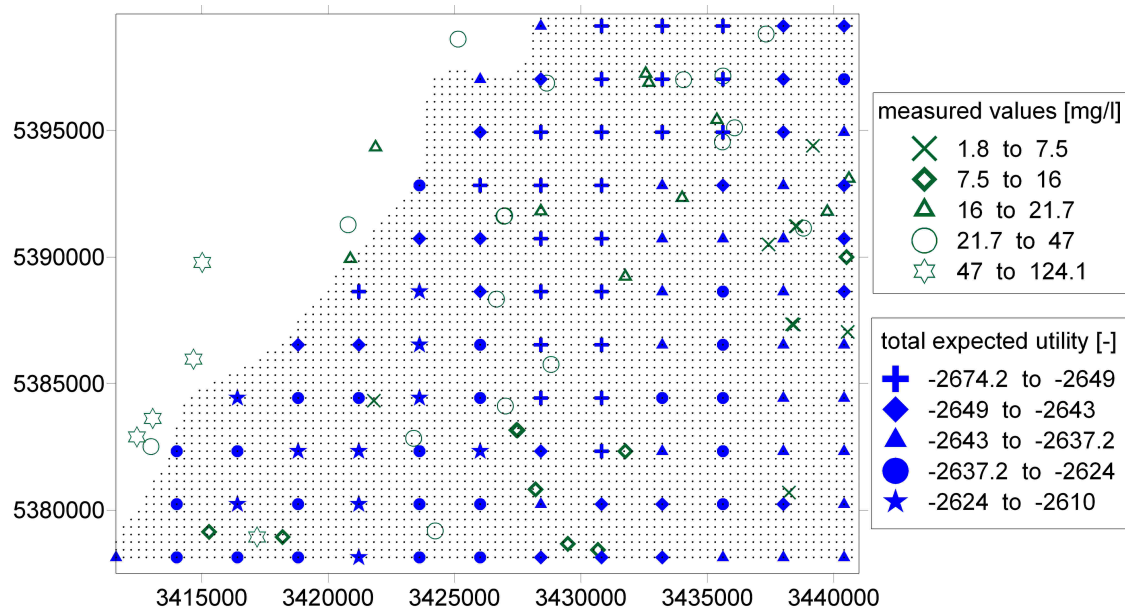


Figure 8.5: Old measurements and new candidate points in the sub-region of Baden-Württemberg in Germany. Different symbols represent different concentration levels for the old measurements and utility levels for the candidates. Small dots represent the interpolation grid.

In summary, the achieved optimal new monitoring locations are in the places where the dilemma whether allow water use or not occurs. However, the traditional Kriging will lead the best points to the places with lowest measurement densities.

## 8.5 Summary of Chapter 8

This chapter has presented a novel approach for monitoring network design based on copulas. The advantages of the proposed approach is that it utilizes the property of copulas to describe the value-dependent spatial dependence to improve the estimation at unsampled location. Since interpolation with copulas yields the full conditional distribution, the uncertainties considered in the decision making process is a function of not only the spatial configuration and measurement density but also the measured values. This is important specially if we are interested in purpose oriented network design (pond) as for example the detection of noncompliance with water quality standards, the detection of higher quantiles in the marginal probability distributions at ungauged locations, the presence or absence of a geophysical variable as gold, hydrocarbons, soil contaminants and so on. Conditional copulas describing the probabilistic structure of the values at the unsampled locations are embedded into the utility function, based on which the optimization scheme is established. The configuration which would maximizes the expected gain defined by the utility function will be chosen to expand the monitoring network. In addition, compared with the indicator approach, the copula based approach is more informative. The synthetical example

and the application to the groundwater quality parameter presented at the end show the potentialities of this method in purpose oriented network design.

# 9 Summary and Outlook

## 9.1 Summary

In this study, the concept of copulas is applied as a new geostatistical tool to describe and model the spatial dependence structure of natural variables. Specific strategies of using copulas to improve the applications of geostatistical approaches in spatial analysis, spatial interpolation, spatial simulation and decision making procedures are developed. Theoretical non-Gaussian copulas are developed which are suitable for spatial modeling. Study of empirical copulas of datasets provide the insight into how the spatial dependence looks like aiming at orienting the selection of a model which well resembles the dependence structure. Once the type of the model is selected, the proposed model inference method which considers multiple point statistics is used to parameterize the selected model. The goodness of fit tests done in the applications in Chapter 6 and 7 show that the parameterized the models are able to reproduce the traditional statistics (variogram/covariance) and the non-Gaussian models produce better crossvalidation results in the interpolation application and reproduce better the asymmetry structure in both the interpolation and simulation applications.

In the spatial analysis and interpolation part, the empirical variogram is replaced by the empirical copulas which take both the data geometry and data values into account in the description of dependence to reveal the change in the dependence structure with the change in the measured values of the parameter under interest as well as with the change in the distance between the measurements. Analysis of the empirical copulas of two dataset: the groundwater quality parameters in Baden-Württemberg, Germany and the hydraulic conductivity measurements in the Las Cruces Trench Site, uncovers the non-Gaussian behaviour of the spatial dependence of these parameters. Scale invariant measures are applied towards the quantification of spatial dependence structure. Advantages of the selected measures compared with the variogram/covariance are demonstrated. Translation invariant spatial copulas that are flexible in describing both Gaussian and non-Gaussian dependence are developed as random functions for interpolation purpose. In the interpolation procedure, estimators of unsampled locations are drawn from the conditional copula at the target locations conditioned on the neighbouring measurements. A method of inference of theoretical copulas from observations combining the maximum likelihood method and multiple-point statistics is proposed. The application of the copula based interpolation approach to the groundwater quality parameters in Baden-Württemberg, Germany give better cross validation results than ordinary and indicator kriging. As the interpolation procedure provides the full conditional distribution, it allows the estimation of confidence intervals. Cross validation results indicate that these intervals are more realistic than those derived from the estimation variance obtained from ordinary and indicator kriging.

On the simulation side, algorithms of generating spatial realizations of various non-Gaussian dependence are developed in both unconditional and conditional cases in order to reproduce heterogeneous spatial structure which governs the subsequent flow and transport simulations performed on these spatial realizations. The unconditional simulation approach is based on non-monotonic transformations of multi-Gaussian processes and provides unlimited possibilities to generate interesting and fantastic spatial patterns. The pitfall of this approach is that it is difficult to predict the right transformation which produces the desired structure. For the conditional simulation case, the principle of sequential simulation is adopted which requires analytical solution to the conditional distribution. The conditional distribution is calculated as the conditional copula. The newly derived multivariate non-Gaussian spatial copulas are implemented as the prior random functions for simulations. The theoretical spatial copula models are validated based on the dataset of hydraulic conductivity measurements of the Las Cruces Trench Site. Statistical tests are conducted to check the goodness-of-fit of the parameterized copula models, results of which indicate that the Gaussian copula is most likely to be rejected, while the maximum normal copula is the most congenial model among the chosen models as a prior random function of this dataset. Significance of spatial dependence on the flow and transport behaviour is investigated indirectly for the realizations of different copulas by studying their Euler characteristics which quantifies the connectivity property of porous media. The results show the evidence that besides the histogram and covariance structure, the multivariate spatial dependence described by copulas has a significant influence on the connectivity property of the resultant realizations and thereby on the flow and transport behaviour.

For the decision making process, statistics is useful in deciding where to collect the additional measurements so that the objectives of monitoring (such as detection of noncompliance with water quality standards) are met in the most cost-effective way [54]. In this framework, spatial modeling with copulas is adopted to facilitate the design of monitoring network for environmental variables. The value-dependent spatial dependence which is important for detection of extreme events is reflected in the estimation uncertainties evaluated from the conditional copula at the candidate locations. The estimated conditional copulas describing the probabilistic structure of the values at the unsampled locations are embedded into the utility function which is to be maximized in order to optimize the expansion of a monitoring network. Compared with the indicator approach, the copula based approach is more informative. The application to expand the observation network of groundwater quality parameters in Baden-Württemberg show the potentialities of this method in purpose oriented network design.

In a nutshell, this study demonstrates that copula provides a more informative and rational way to describe spatial dependence structures of natural variables. Methodologies of using copulas for interpolation, simulation and design purposes are developed and successfully applied to real datasets. Compared with traditional geostatistical tools, the proposed copula based approaches give better validation results in the applications and proved to be more suitable for the spatial analysis and modeling of datasets which exhibit multivariate non-Gaussian behaviour.

## 9.2 Outlook

Although the study has shown the promising aspects of the developed approach as an alternative to traditional geostatistical methods, some suggestions on further improvement can be put forward.

First, as analysis of empirical copulas of several datasets show the variant dependence structure deviates from the amiable Gaussian dependence, theoretical non-Gaussian spatial copulas are developed to better resemble the natural structure. However, if the phenomenon under study is a mixture of different physical processes, the dependence of the mixture cannot be modeled by one continuous random function. For example, the following Figure shows different horizontal planes from the X-ray tomography of the bulk density of a soil column [13].

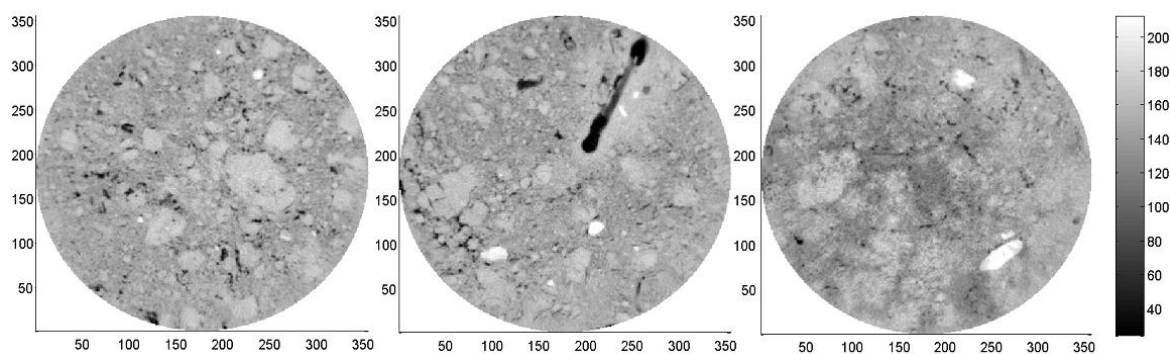


Figure 9.1: *Horizontal planes from the X-ray tomography of the bulk density of a soil column [13].*

Figure 9.1 shows that it is very likely that there are several processes going on in the studied domain, e.g., the matrix as the background image, the big stones unevenly distributed in the field, the variant-sized pores with different connectivity properties scattered in the field. In this case, if the target is to simulate the three dimensional field which mimics the structures of this dataset, any of the existing theoretical copulas, including the maximum normal copula which is the combination of two Gaussian processes, may fail to capture and reproduce the complicated spatial dependence. Existing approaches to handle such problem are suggested as Boolean Algorithms [20] [72] [83], training image approach [36] [84] [43]. However, in the framework of copula based approach, the prior random function is searched. A possible solution to this problem is to parameterize the non-monotonic transformation which allows the creation of various dependence structures.

Figure 9.2 illustrates an example of non-monotonic transformation and the key transformation parameters. As suggested above, the transformation parameters should be parameterized. Hence the parameterization method has to be adapted to realize this target. For the interpolation and conditional simulation of such natural phenomenon, either the analytical solution or a numerical scheme should be derived or applied to solve the conditional distribution function of the complex dependence.



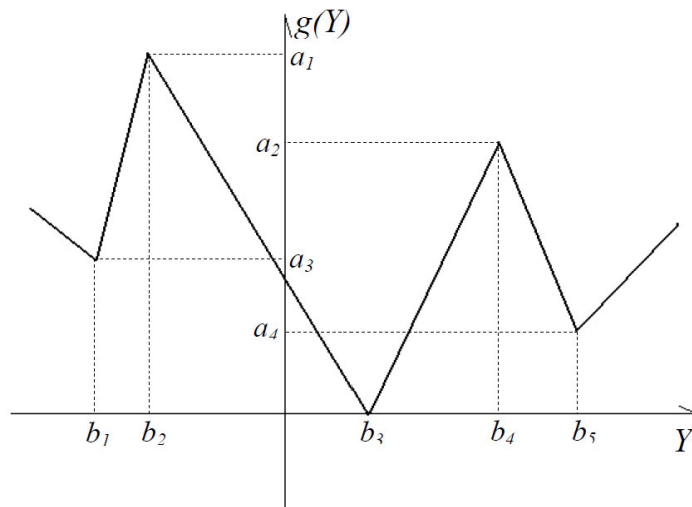


Figure 9.2: An example of more complicated non-monotonic transformation with transformation parameters to be estimated from datasets.

Another aspect which can be studied in the future is already mentioned at the end of Chapter 3, namely, the systematic study of the relationship between the topological characteristics of the porous media and the underlying copulas. This study aims at enabling the selection and parameterization of the copula model to respect the connectivity property of the spatial fields as well as to reproduce it in the interpolation or simulation results.

So far, the copula based geostatistical approach has only been applied to continuous random variables. Further advances can be done to extend the application to categorical variables. In this case, different strategies of spatial analysis and modeling using copulas have to be developed.



## Appendix: v-transformed normal copula

Let  $\mathbf{Y}$  be  $N(\mathbf{0}, \Gamma)$ , an  $n$  dimensional normal random variable with  $\mathbf{0}^T = (0, \dots, 0)$  mean and  $\Gamma$  correlation matrix. All marginals are supposed to have unit variance. Let  $\mathbf{X}$  be defined for each coordinate  $j = 1, \dots, n$  as:

$$X_j = \begin{cases} k(Y_j - m)^\alpha & \text{if } Y_j \geq m \\ m - Y_j & \text{if } Y_j < m \end{cases} \quad (9.1)$$

Where  $k$  is a positive constant,  $m$  and  $\alpha$  are arbitrary real numbers.

The one dimensional marginals of the resultant variable  $\mathbf{X}$  are identical and have the distribution function:

$$\begin{aligned} H_1(x) &= P(X < x) \\ &= P\left(Y < \left(\frac{x}{k}\right)^{\frac{1}{\alpha}} + m\right) + P(Y > x - m) \\ &= P\left(Y < \left(\frac{x}{k}\right)^{\frac{1}{\alpha}} + m\right) - P(Y > x - m) \\ &= \Phi\left(\left(\frac{x}{k}\right)^{\frac{1}{\alpha}} + m\right) - \Phi(-x + m) \end{aligned} \quad (9.2)$$

and density:

$$h_1(x) = \frac{1}{k\alpha} \cdot \left(\frac{x}{k}\right)^{\left(\frac{1}{\alpha}-1\right)} \cdot \phi\left(\left(\frac{x}{k}\right)^{\frac{1}{\alpha}} + m\right) + \phi(-x + m) \quad (9.3)$$

with  $\Phi$  being the distribution function of the standard normal distribution and  $\phi$  the corresponding density function.

The illustration of the derivation of the multivariate v-transformed normal copula starts with an illustration of the bivariate case:

$$\begin{aligned}
H_2(x_1, x_2) &= P(X_1 < x_1, X_2 < x_2) \\
&= P\left(Y_1 < \left(\frac{x_1}{k}\right)^{\frac{1}{\alpha}} + m, Y_2 < \left(\frac{x_2}{k}\right)^{\frac{1}{\alpha}} + m\right) \\
&\quad - P\left(Y_1 < -x_1 + m, Y_2 < \left(\frac{x_2}{k}\right)^{\frac{1}{\alpha}} + m\right) \\
&\quad - P\left(Y_1 < \left(\frac{x_1}{k}\right)^{\frac{1}{\alpha}} + m, Y_2 < -x_2 + m\right) \\
&\quad + P(Y_1 < -x_1 + m, Y_2 < -x_2 + m) \\
&= \Phi_2\left(\left(\frac{x_1}{k}\right)^{\frac{1}{\alpha}} + m, \left(\frac{x_2}{k}\right)^{\frac{1}{\alpha}} + m\right) - \Phi_2\left(\left(\frac{x_1}{k}\right)^{\frac{1}{\alpha}} + m, -x_2 + m\right) \\
&\quad - \Phi_2\left(-x_1 + m, \left(\frac{x_2}{k}\right)^{\frac{1}{\alpha}} + m\right) + \Phi_2(-x_1 + m, -x_2 + m) \tag{9.4}
\end{aligned}$$

The challenge to of extension to higher dimensions is to determine the sign in front of each term and the vector inside  $\Phi$ . As shown in Chapter 4 the multivariate distribution function reads:

$$\begin{aligned}
H_n(x_1, \dots, x_n) &= P(X_1 < x_1, \dots, X_n < x_n) = \\
&= \sum_{i=0}^{2^n-1} (-1)^i \Phi(\zeta_i + \mathbf{m}) \tag{9.5}
\end{aligned}$$

where

$$\zeta_i^T = (b((-1)^{i_1}) x_1^a, \dots, b((-1)^{i_n}) x_n^a) \tag{9.6}$$

$i_j = 0$  or  $1$  and

$$i = \sum_{j=0}^{n-1} i_j 2^j \tag{9.7}$$

and

$$b = \begin{cases} -1 & \text{if } (-1)^{i_j} = -1 \\ \frac{1}{k} & \text{if } (-1)^{i_j} = 1 \end{cases} \tag{9.8}$$

and

$$a = \begin{cases} 1 & \text{if } (-1)^{i_j} = -1 \\ \frac{1}{\alpha} & \text{if } (-1)^{i_j} = 1 \end{cases} \tag{9.9}$$

Equation (9.8) and (9.9) show that the coefficient  $b$  and the exponent  $a$  are determined by  $i_j$ -s which are either 0 or 1. While Equation (9.7) shows that  $i_j$ -s has a relationship with  $i$  which ranges from 0 to  $2^n - 1$  and is used to determinate the signs in front of the individual terms  $\Phi()$  in Equation (9.5). The calculation of the multivariate distribution described in

Equation (9.5) can be considered as the Cartesian product of  $n$  closed intervals, denoted here as  $[c_1, d_1] \times [c_2, d_2] \times \dots \times [c_n, d_n]$ .

Let  $S_1, S_2, \dots, S_n$  be nonempty real subsets, and let  $H$  be an  $n$ -place real function such that  $\text{Dom}H = S_1 \times S_2 \times \dots \times S_n$ . Let  $B = [c, d]$  of an  $n$ -box all of whose vertices are in  $\text{Dom}H$ . Then the  $H$ -volume of  $B$  is given by [66]:

$$V_H(B) = \sum sgn(\mathbf{v})H(\mathbf{v}) \quad (9.10)$$

where the sum is taken over all vertices  $\mathbf{v}$  of  $B$ , and  $sgn(\mathbf{v})$  is given by:

$$sgn(\mathbf{v}) = \begin{cases} 1 & \text{if } v_k = c_k \text{ for an even number of } k\text{-s} \\ -1 & \text{if } v_k = c_k \text{ for an odd number of } k\text{-s} \end{cases} \quad (9.11)$$

For the  $v$  copula,  $c_i = -x_i + m$ , and  $d_i = \left(\frac{x_i}{k}\right)^{\frac{1}{\alpha}} + m$ . The signs in front of  $x_i$  in the vector  $\zeta_i^T$  (see Equation (9.6)) determine whether the component is the lower or upper boundary of the  $n$  intervals, thus also determine the sign ( $sgn(\mathbf{v})$ ) in front of each  $\Phi()$  term in Equation (9.5). In fact, Equation (9.7) is based on the binary decoding of  $i$  in order to find all the combinations of signs in front of  $x_i$ -s in the vector  $\zeta_i^T$ . For example if  $n = 3$ , then  $i$  ranges from 0 to 7. The corresponding binary decoding of  $i$  are shown in the following table:

	$i_j$	sign of $x_1$	sign of $x_2$	sign of $x_3$
$i = 0$	0 0 0	+	+	+
$i = 1$	1 0 0	-	+	+
$i = 2$	0 1 0	+	-	+
$i = 3$	1 1 0	-	-	+
$i = 4$	0 0 1	+	+	-
$i = 5$	1 0 1	-	+	-
$i = 6$	0 1 1	+	-	-
$i = 7$	1 1 1	-	-	-

Table 9.1: Illustration of binary decoding for assigning the signs of the components in the vector  $\zeta_i^T$ .

For the trivariate case, if  $i = 2$ , Table 9.1 shows that  $i_j = 0, 1, 0$  with  $j = 0, 1, 2$ . Applying Equation (9.7), one can get:

$$i = 0 \times 2^0 + 1 \times 2^1 + 0 \times 2^2 = 2$$

which is correct. In this case, the number of the vertices  $v_k$  which equal the lower boundaries  $c_k$  is an odd number, therefore the third  $\Phi()$  in Equation (9.5) gets a negative sign.

For the distribution density, by taking the derivative of Equation (9.4), one gets:

$$\begin{aligned}
h_2(x_1, x_2) = & \frac{1}{k^2 \alpha^2} \left( \frac{x_1 x_2}{k^2} \right)^{\frac{1}{\alpha} - 1} \phi_2 \left( \left( \frac{x_1}{k} \right)^{\frac{1}{\alpha}} + m, \left( \frac{x_2}{k} \right)^{\frac{1}{\alpha}} + m \right) \\
& + \frac{1}{k \alpha} \left( \frac{x_1}{k} \right)^{\frac{1}{\alpha} - 1} \phi_2 \left( \left( \frac{x_1}{k} \right)^{\frac{1}{\alpha}} + m, -x_2 + m \right) \\
& + \frac{1}{k \alpha} \left( \frac{x_2}{k} \right)^{\frac{1}{\alpha} - 1} \phi_2 \left( -x_1 + m, \left( \frac{x_2}{k} \right)^{\frac{1}{\alpha}} + m \right) \\
& + \phi_2(-x_1 + m, -x_2 + m)
\end{aligned} \tag{9.12}$$

Extention to the multivariate case leads to the general formular:

$$\begin{aligned}
h_n(x_1, \dots, x_n) = & \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{\Gamma}^{-1}|^{\frac{1}{2}}} \sum_{i=0}^{2^n - 1} \frac{1}{(\alpha k)^{n - \sum_{j=0}^{n-1} i_j}} \cdot \left( \left( \frac{x}{k} \right)^{\frac{1}{\alpha} - 1} \right)^{n - \sum_{j=0}^{n-1} i_j} \\
& \cdot \exp \left( -\frac{1}{2} (\zeta_i + \mathbf{m})^T \mathbf{\Gamma}^{-1} (\zeta_i + \mathbf{m}) \right)
\end{aligned} \tag{9.13}$$

where  $i_j$  and  $zeta_i$  are also determined by Equation (9.7) and (9.6), respectively.

# Bibliography

- [1] E. Aarts and J. Korst. *Simulated Annealing and Boltzmann Machines*. Wiley, New York, 1989.
- [2] B. Abdous, B. Genest, and B. Rémillard. Dependence properties of meta-elliptical distributions. *Statistical Modeling and Analysis for Complex Data Problems*, 8:1–15, 2005.
- [3] M. Abramovitz and I. Stegun. *Handbook of mathematical functions*. Dover, 1964.
- [4] F. G. Alabert. Stochastic imaging of spatial distributions using hard and soft information: Un-published master's thesis, 1987.
- [5] F. G. Alabert. The practice of fast conditional simulations through the LU decomposition of the covariance matrix. *Mathematical Geology*, 19:369–87, 1987.
- [6] T. Anderson. *An Introduction to multivariate statistical analysis*. Wiley Sons, 1958.
- [7] A. Bárdossy. Introduction to Geostatistics: Lecture Notes, 1997.
- [8] A. Bárdossy. Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research*, 42:doi:10.1029/2005WR004754, 2006.
- [9] A. Bárdossy, J. Giese, J. Grimm-Strele, and K. Barufke. Simik+gis - implementierte interpolation von grundwasserparametern mit hilfe von landnutzungs- und geologiedaten. *Hydrologie und Wasserbewirtschaftung*, 47:13 – 20, 2003.
- [10] A. Bárdossy, U. Haberlandt, and J. Grimm-Strele. Interpolation of groundwater quality parameters using additional information. *geoENV I, Geostatistics for Environmental Applications*, 47:189 – 200, 1997.
- [11] A. Bárdossy and J. Li. Geostatistical Interpolation using Copulas. *Water Resources Research*, 44:doi:10.1029/2007WR006115, 2008.
- [12] A. Bárdossy, J. Li, and L. Guenni. Observation network design for environmental variables using a copula based approach. In *Proceedings of the Eighth International Geostatistics Congress, Julián. M. Ortiz and Xavier Emery, Santiago de Chile, 1-5 December 2008*, pages 207–216.
- [13] A. Bayer, H.-J. Vogel, and K. Roth. Direct measurement of the soil water retention curve using X-ray absorption. *Hydrology Earth System Sciences*, 8:2–7, 2004.

- [14] L. M. Borgman, M. Taheri, and R. Hagan. Three-dimensional frequency domain simulations of geological variables. In *Geostatistics for Natural Resource Characterization*, pages 517–541, 1984.
- [15] Roland N. Bracewell. *The Fourier Transform and its Applications*. McGraw-Hill, 2000.
- [16] H. Chang, A. Qiuyan Fu, ND. Le, and JV. Zidek. Designing environmental monitoring networks to measure extremes. *Environmental and Ecological Statistics*, 14:301–321, 2007.
- [17] U. Cherubini, E. Luciano, and W. Vecchiato. *Copula Methods in Finance*. Willey, 2004.
- [18] J. P. Chilés and P. Delfiner. *Geostatistics: Modeling spatial uncertainty*. Wiley, New York, 1999.
- [19] A. Cirpka. *Stochastic Methods in Subsurface Hydrology, Lecture Notes*, Swiss Federal Institute for Environmental Science and Technology, 2004.
- [20] D. Cox and V. Isham. *Point Process*. Chapman and Hall, New York, 1980.
- [21] M. W. Davis. Production of conditional simulations via the LU triangular decomposition of the covariance matrix. *Mathematical Geology*, 19:91–8, 1987.
- [22] J. P. Delhomme. Spatial variability and uncertainty in groundwater flow parameters: a geostatistical approach. *Water Resources Research*, 15:269–80, 1979.
- [23] C. V. Deutsch. Conditioning reservoir models to well test information. In *Geostatistics Tróia '92, A. Soares, 1993, Kluwer, Dordrecht*, pages 505 – 18, 1993.
- [24] C. V. Deutsch and A. G. Journel. *GSLIB: Geostatistical software library and user's guide, 2nd edition*. Oxford University Press, New York, 1998.
- [25] P. Embrechts, A. J. McNeil, and D. Straumann. *Correlation and Dependency in Risk Management: Properties and Pitfalls*. Cambridge University Press, New York, 2001.
- [26] X. Emery. Contional simulation of nongaussian random functions. *Mathematical Geology*, 34(1):79–100, 2002.
- [27] H.-B. Fang, K.-T. Fang, and S. Kotz. The meta elliptica distribution with given marginals. *Journal of Multivariate Analysis*, 82:1–16, 2002.
- [28] C. Farmer. Numerical rocks. In *Joint IMA/SPE European Conference on the Mathematics of Oil Recovery*. Robinson College, Cambrige University, 1989.
- [29] C. L. Farmer. The generation of stochastic fields of reservoir parameters with specified geostatistical distributions. In *Mathematics in Oil Production*, pages 235 – 252, 1988.
- [30] A. C. Favre, S. E. Adlouni, L. Perreault, N. Thiémonge, and B. Bobée. Multivariate hydrology frequency analysis using copulas. *Water Resources Research*, 40:W01101, 2004.
- [31] V. Federov and W. Müller. Two approaches in optimization of observing networks. *Optimal Design and Analysis of Experiments*, 1:1–2, 1988.



- [32] V. Federov and W. Müller. Comparison of two approaches in the optimal design of an observation network. *Statistics*, 3:339–351, 1989.
- [33] C. Genest and R. MacKay. Archimedean copulas and families of bidimensional laws for which the marginals are given. *The Canadian Journal of Statistics*, 14:145–159, 1986.
- [34] J. J. Gómez-Hernández and X.-H. Wen. To be or not to be multi-gaussian? a reflection on stochastic hydrogeology. *Advances in Water Resources*, 21(1):47 – 61, 1998.
- [35] P. Goovaerts. *Geostatistics for natural resources evaluation*. Oxford University Press, New York, 1997.
- [36] F. Guardiano and R. M. Srivastava. Multivariate geostatistics: Beyond bivariate moments. *Geostatistics-Troia, Vol. 1*, pages 404–412, 1993.
- [37] R. Hilfer. Local-porosity theory for flow in porous media. *Physical Review B*, 45:7115–7121, 1992.
- [38] R. Hilfer. Transport and relaxation phenomena in porous media. *Advances in Chemical Physics*, 92:299–424, 1996.
- [39] R. Hilfer, T. Rage, and B. Virgin. Local percolation probabilities for a natural sandstone. *Physica A*, 241:105–110, 1997.
- [40] R. G. Hills, D. B. Hudson, and P. J. Wierenga. Spatial variability at the las cruces trench site. In M. Th. van Genuchten and F. J. Leij, editors, *Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils*, Proceedings of the International Workshop on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils, pages 529 – 539. Riverside, California, October 11-13 1989.
- [41] R. G. Hills, P. J. Wierenga, D. B. Hudson, and M. R. Kirkland. The second las cruces trench experiment: Experimental results and three-dimensional flow predictions. *Water Resources Research*, 27:2709 – 2718, 1991.
- [42] M. Hollander and D. A. Wolfe. *Nonparametric Statistical Methods*. Wiley, 1973.
- [43] L. Y. Hu and T. Chugunova. Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review. *Water Resources Research*, 44:doi:10.1029, 2008.
- [44] J. E. Ingersoll. *Theory of Financial Decision Making*. Rowman and Littlefield, 1987.
- [45] A. Journel and T. Zhang. The necessity of a multi-point prior model. *Mathematical Geology*, 48:591 – 610, 2006.
- [46] A. G. Journel. Geostatistics for conditional simulation of orebodies. *Economic Geology*, 69:673–80, 1974.
- [47] A. G. Journel. *Fundamentals of geostatistics in five lessons, Vol. 8: Short course in geology*. American Geophysical Union, Washington, 1989.

- [48] A. G. Journel. Deterministic geostatistics: A new visit. In *Geostatistics Wollongong, Vol. 1: Kluwer Academic, Dordrecht*, pages 174 – 187, 1997.
- [49] A. G. Journel and F. Alabert. Non-gaussian data expansion in the earth sciences. *Terra Nova*, 1:123 – 134, 1989.
- [50] A. G. Journel and F. Alabert. Non-gaussian data expansion in the earth sciences. *TERRA Nova*, 1:123 – 134, 1989.
- [51] A. G. Journel and Clayton V. Deutsch. Entropy and spatial disorder. *Mathematical geology*, 25(3):329 – 355, 1993.
- [52] A. G. Journel and CH. J. Huijbregts. *Mining Geostatistics*. Academic Press, New York, 1978.
- [53] S. L. King. Sequential Gaussian simulation vs. simulated annealing for locating pockets of high-value commercial trees in Pennsylvania. *Annals of Operations Research*, 95:177–203, 2000.
- [54] P. K. Kitanidis. *Introduction to Geostatistics: Applications in Hydrogeology*. Cambridge University Press, Cambridge, United Kingdom, 1997.
- [55] N. T. Kottegoda and R. Rosso. *Statistics, Probability, and Reliability for Civil and Environmental Engineers*. McGraw-Hill, 1997.
- [56] W. H. Kruskal. Ordinal measures of association. *Journal of the American Statistical Association*, 53(284):814 – 861, 1958.
- [57] C. Lantuéjoul. Iterative algorithms for conditional simulation. *Geostatistics Wollongong'96*, 1997.
- [58] L. S. Lasdon, R. L. Fox, and M. W. Ratner. Nonlinear optimization using the generalized reduced gradient method. *RAIRO*, 3:73 – 104, 1974.
- [59] L. S. Lasdon and A. D. Waren. Generalized reduced gradient software for linearly and nonlinearly constrained problems. In *Design and Implementation of Optimization Software*, pages 335–362, 1978.
- [60] E. L. Lehmann. Some concepts of dependence. *The Annals of Mathematical Statistics*, 37(5):1137 – 1153, 1966.
- [61] E. L. Lehmann. *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco, 1975.
- [62] G. Le Loc'h, H. Beucher, A. Galli, B. Doligez, and Heresim Group. Improvement in the truncated gaussian method: Combining several gaussian functions. In *Topic B: Heterogeneity Description and Assessment of Uncertainty, ECMOR IV, Fourth European Conference on the Mathematics of Oil Recovery, Norway, June 7-10 1994*.

- [63] A. Mantoglou and J. L. Wilson. The turning bands method for simulation of random fields using 1ne generations by a spectral method. *Water Resources Research*, 18:1379–94, 1982.
- [64] G. Matheron. *The theory of Regionalized Variables and Its Applications*. Ecole de Mines, Fontainebleau, France, 1991.
- [65] J. M. Mejia and I. Rodríguez-Iturbe. On the synthesis of random field sampling from the spectrum: An application to the generation of hydraulic spatial processes. *Water Resources Research*, 10:705–11, 1974.
- [66] R. Nelsen. *An introduction to copulas*. Springer Verlag, 1999.
- [67] I. Neuweiler and A. Cirpka. Homogenization of Richards equation in permeability fields with different connectivities. *Water Resources Research*, 41:doi:10.1029/2004WR003329, 2005.
- [68] I. Neuweiler and H-J. Vogel. Upscaling for unsaturated flow for non-Gaussian heterogeneous porous media. *Water Resources Research*, 43:doi:10.1029/2005WR004771, 2007.
- [69] R. Olea. *Geostatistical Glossary and Multilingual Dictionary*. Oxford University Press, New York, 1991.
- [70] A. Poulin, D. Huard, A. C. Favre, and S. Pugin. Importance of tail dependence in bivariate frequency analysis. *Journal of Hydrologic Engineering*, 12:394–403, 2007.
- [71] S. Rachev, C. Menn, and F. Fabozzi. *Fat-Tailed and Skewed Asset Return Distributions: Implications for Risk Management, Portfolio Selection, and Option Pricing*. Wiley, 2005.
- [72] B. Ripley. *Stochastic Simulation*. John Wiley Sons, New York, 1987.
- [73] V. Robins. Computational topology for point data: Betti numbers and  $\alpha$ -shapes. *Lecture Notes in Physics*, 600:261–274, 2002.
- [74] J. L. Rodgers and W. A. Nicewander. Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42:59 – 66, 1988.
- [75] G. H. De Rooij, R. T. A. Kasteel, A. Papritz, and H. Flühler. Joint distribution of the unsaturated soil hydraulic parameters and their effect on other variates. *Vadose Zone Journal*, 3:947 – 955, 2004.
- [76] G. Salvadori, C. D. Michele, N. Kottegoda, and R. Rosso. *Extremes in nature. An approach using copulas*. Springer, 2007.
- [77] A. Samouëlian, H-J. Vogel, and O. Ippisch. Upscaling hydraulic conductivity based on the topology of the sub-scale structure. *Advances in Water Resources*, 30:1179–1189, 2007.
- [78] B. Schweizer and EF Wolff. On nonparametric measures of dependence for random variables. *Annals of Statistics*, 9(4):879 – 885, 1981.

- [79] M. Shinozuka and C. M. Jan. Digital simulation of random processes and its applications. *Journal of Sound and Vibration*, 25(1):111–128, 1972.
- [80] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Stat. Paris*, 8:229–131, 1959.
- [81] M. Srivastava. Iterative methods for spatial simulation: Stanford Center for Reservoir Forecasting, Rep. No. 5. Technical report, 1992.
- [82] M. Srivastava. An overview of stochastic methods for reservoir characterization. In *Stochastic modeling and geostatistics: Principles, methods, and case studies, Vol. 3: Computer Applications in Geology, Tulsa*, pages 3 – 16, 1995.
- [83] D. Stoyan, W. Kendall, and J. Mecke. *Stochastic Geometry and its Applications*. John Wiley Sons, New York, 1987.
- [84] S. Strebelle. Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, 34(1):1–21, 2002.
- [85] A. H. Studenmunc. *Using Econometrics: A Practical Approach*. Addison-Wesley, Boston MA, USA, 2005.
- [86] H-J. Vogel. Morphological determination of pore connectivity as a function of pore size using serial sections. *European Journal of Soil Science*, 48:365–377, 1997.
- [87] H-J. Vogel. Topological characterization of porous media. In *Morphology and condensed matter - physics and geometry of spatially complex systems. Lecture notes in physics. Lecture notes in physics*, pages 75–92, 2002.
- [88] M. Ward and C. Folland. Prediction of seasonal rainfall in the nordeste of brazil using eigenvectors of sea-surface temperature. *International Journal Climatology*, 11:711 – 743, 1991.
- [89] R. Webster and M. A. Oliver. *Geostatistics for Environmental Scientists, 2nd Edition*. Wiley, United Kingdom, 2007.
- [90] H. White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817 – 838, 1980.
- [91] J. Widjajakusuma, C. Manwart, B. Biswal, and R. Hilfer. Exact and approximate calculations for the conductivity of sandstones. *Physica A*, 270:325–331, 1999.
- [92] P. J. Wierenga, D. Bachelet, J. R. Bilskie, H. Elabd, D. B. Hudson, M. Nash, I. Porro, W. R. Strong, A. Toorman, and J. Vinson. Validation of stochastic flow and transport models for unsaturated soils. Research report 88-ss-03, Dept. of Agronomy and Horticulture, New Mexico State Univ., Las Cruces, NM, 1988.
- [93] P. J. Wierenga, R. G. Hills, and D. B. Hudson. The second las cruces trench site: Characterization, experimental results, and one-dimensional flow predictions. *Water Resources Research*, 27:2695 – 2705, 1991.

- [94] H. Xiao. A description of the behaviour of indicator variograms for a bivariate normal distribution: master's thesis, 1985.
- [95] B. Zinn and C. F. Harvey. When good statistical models of aquifer heterogeneity go bad: A comparison of flow, dispersion, and mass transfer in connected and multivariate gaussian hydraulic conductivity fields. *Water Resources Research*, 39(3):doi:10.1029/2001WR001146, 2003.

# Curriculum Vitae

## Personal Information

Name: Jing Li  
Date of Birth: 15.01.1978  
Place of Birth: Shanghai, China  
Contact Information: yatou1615@hotmail.com

## Educational Background

01/2007 - 01/2009 : Member of the International Research Training Group NUPUS  
(<http://www.nupus.uni-stuttgart.de/>)

01/2006 - 01/2009 : Doctoral study  
Dept. of Hydrology and Geohydrology, Institute of Hydraulic Engineering  
University of Stuttgart, Germany  
Member of the International Doctoral Program ENWAT  
(<http://www.phd.envwater.uni-stuttgart.de/>)

09/2003 - 11/2005: Master of Science  
International Program of Water Resources Engineering and Management  
University of Stuttgart, Germany

09/1996 - 07/2000: Bachelor of Science  
Dept. of Civil and Environmental Engineering  
Tongji University, Shanghai, China

## Working Experience

01/2006 - 01/2009 : Teach and Research Assistant  
Dept. of Hydrology and Geohydrology, Institute of Hydraulic Engineering  
University of Stuttgart, Germany

03/2004 - 11/2005: Research Assistant  
Institute of Physics and Meteorology  
University of Hohenheim, Stuttgart, Germany

01/2002 - 08/2003: Assistant Engineer  
SINGAPORE SALCON, Environmental Engineering Company,  
Shanghai Branch  
Shanghai, China

08/2000 - 12/2001: Customer Representative  
Shanghai Pudong Development Bank  
Shanghai China

### **Publications**

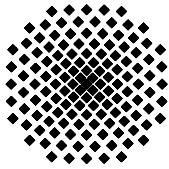
- Bárdossy, A., J. Li (2008), Geostatistical Interpolation using Copulas, *Water Resources Research*, 44, W07412, doi:10.1029/2007WR006115
- Bárdossy, A., J. Li, de Guenni, L. B. (2008), Observation Network Design for Environmental Variables using a Copula based Approach, In J. M. Ortiz editor, *Proceedings of VIII International Geostatistics Congress*, Vol. 1, P. 207-216, December 2008, Santiago de Chile
- Haslauer, C., J. Li, Bardossy, A. (2008), Application of Copulas in Geostatistics, In Atkinson, P. M. and Lloyd, C. D. editors, *Proceedings of 7th International Conference on Geostatistics for Environmental Applications*, September 2008, Southampton, UK

### **Awards**

2008 : CHUNHUI Chinese Government Award for Outstanding PhD Students Abroad 2008

2004: Excellent Master Student Scholarship offered by ZV LW (Zweckverband Landeswasserversorgung)

1998 and 1999: Excellent Undergraduate Student Prize offered by Tongji University



## Institut für Wasserbau Universität Stuttgart

Pfaffenwaldring 61  
70569 Stuttgart (Vaihingen)  
Telefon (0711) 685 - 64717/64749/64752/64679  
Telefax (0711) 685 - 67020 o. 64746 o. 64681  
E-Mail: [iws@iws.uni-stuttgart.de](mailto: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. Silke Wieprecht

### Vorstand (Stand 01.04.2009)

Prof. Dr. rer. nat. Dr.-Ing. A. Bárdossy  
Prof. Dr.-Ing. R. Helmig  
Prof. Dr.-Ing. S. Wieprecht  
Jürgen Braun, PhD  
Dr.-Ing. H. Class  
Dr.-Ing. S. Hartmann  
Dr.-Ing. H.-P. Koschitzky  
PD Dr.-Ing. W. Marx  
Dr. rer. nat. J. Seidel

### 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

Leiter: Prof. Dr.-Ing. Silke Wieprecht  
Stellv.: PD Dr.-Ing. Walter Marx, AOR

### Versuchsanstalt für Wasserbau

Leiter: Dr.-Ing. Sven Hartmann, AOR

### Lehrstuhl für Hydromechanik und Hydrosystemmodellierung

Leiter: Prof. Dr.-Ing. Rainer Helmig  
Stellv.: Dr.-Ing. Holger Class, AOR

### Lehrstuhl für Hydrologie und Geohydrologie

Leiter: Prof. Dr. rer. nat. Dr.-Ing. András Bárdossy  
Stellv.: Dr. rer. nat. Jochen Seidel

### VEGAS, Versuchseinrichtung zur Grundwasser- und Altlastensanierung

Leitung: Jürgen Braun, PhD  
Dr.-Ing. Hans-Peter Koschitzky, AD

## 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 Flächen-tragwerke unter besonderer Berücksichtigung ebener Stauwände, mit Hilfe von Randwert- 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; Gesteungskosten 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 Binnenschifffahrt 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 Binnenschifffahrt 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 Gál, 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; Gál, 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 Binnenschifffahrt durch Schwimmerhebwerke*, 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; Gál, 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 rotations-symmetrisch 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 seegangbelasteter 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 horizontalebene 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 Schußbrinnen 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 Klufdaten zur realitätsnahen Generierung von Klufnetzen 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 Karst-aquiferen 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, ISBN 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 instationärer 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
- 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 Porous 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 Flusstauhaltungen*, 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: Upscaling, 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 Sub-surface - 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 Yesheatesfa: *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, 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 Hochwasserrückhaltebecken-systemen*, 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; Stuhmann, 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 eingeeerdeter 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 3-933761-57-3
- 155 Bielinski, Andreas: *Numerical Simulation of CO<sub>2</sub> 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; Stuhmann, 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; Stuhmann, 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<sub>2</sub> 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<sub>2</sub> 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

Die Mitteilungshefte ab der Nr. 134 (Jg. 2005) stehen als pdf-Datei über die Homepage des Instituts: [www.iws.uni-stuttgart.de](http://www.iws.uni-stuttgart.de) zur Verfügung.