Correlation Dimension analysis of complex hydrological systems: what information can the method provide?

Dissertation

zur Erlangung des Doktorgrades

des Fachbereichs Geowissenschaften

der Freien Universität Berlin

vorgelegt von

Miaomiao Ma

Berlin 2013

1. Gutachter: **PD. Dr. Christoph Merz**

Freie Universität Berlin

Fachbereich Geowissenschaften

Institut für Geologische Wissenschaften and

Leibniz-Zentrum für Agrarlandschaftsforschung (ZALF)

Institut für Landschaftswasserhaushalt

2. Gutachter: **Prof. Dr. Michael Schneider**

Freie Universität Berlin

Fachbereich Geowissenschaften

Institut für Geologische Wissenschaften

3. Gutachter: **Prof. Dr. Gunnar Lischeid**

Universität Potsdam

Institut für Erd- und Umweltwissenschaften and

Leibniz-Zentrum für Agrarlandschaftsforschung (ZALF)

Institut für Landschaftswasserhaushalt

Tag der Disputation: 27.06.2013

Preface

The present thesis was conducted as my Ph.D. research project "Equilibrating data and model complexity" in the framework of a cooperation between Freie Universität Berlin (FUB) and the China Scholarship Council (CSC). The work outlined in this thesis was carried out over a period of three years (September 2010 to August 2013). Financial support was provided by the CSC.

The main chapters of this thesis have been submitted for publications[1], to which I contributed as the first author. I have been in charge of the development and conduction of the research work described. I have been responsible for the scientific content and for the preparation of manuscripts, including literature review, data interpretation, preparation of figures and tables, and manuscript writing. Co-authors have played an advisory and supervisory role.

The thesis contains the contributions by Björn Thomas for collecting and extracting the meteorological and catchment properties data of 63 catchments in the Federal State of Brandenburg, Germany.

---

[1] Ma M, Lischeid G, Merz C, Thomas B., *under review*. "Correlation dimension analysis of observed discharge time series in small catchments: what information the method provide?", *Journal of Hydrology*.

Ma M, Lischeid G, Merz C., *under review*. "Using the Correlation Dimension analysis to evaluate model performance", *Environmental modeling & software*.

Acknowledgements

This thesis summarized my whole Ph.D. research work under the supervision of PD. Dr. Christoph Merz and Prof. Dr. Gunnar Lischeid. With their support, I finally conquered all the difficulties and achieved this research topic. Thanks for their guidance, plenty scientific discussions and many inspiring ideas. Here, I also want to thank Prof. Dr. Asaf Pekdeger, Prof. Dr. Michael Schneider and Dr. Andreas Winkler for giving me the chance to work on a very interesting and challenging project.

I am very grateful to all colleagues at the Institute of Landscape Hydrology and of the Hydrogeology workgroup of the Free University Berlin for their support. In particular, I want to thanks my lovely colleagues: Tobias Hohenbrink, Björn Thomas, Marcus Fahle, Steven Böttcher, Philipp Rauneker, Christian Lehr and Mohamed Omari. Thanks for sharing the academic knowledge and life experiences, for their assistance in resolving the various technical issues and for discussing my papers. Special thanks give to Christian Lehr for translating the summary of this thesis into German language and Prof. Dr. Weimin Bao for his discussions and suggestions of the Vertical-mixed Runoff model.

In addition, I acknowledge the State Office of Environment, Health and Consumer Protection of the Federal State of Brandenburg, Institute of Landscape Hydrology in Leibniz Centre for Agricultural landscape Research (ZALF) and HOHAI University of China for providing the data. Thanks China Scholarship Council for providing the funding.

Declaration of originality

I hereby certify, as the author of this thesis and as one of the main authors of the publications arising, that I was the person responsible for organization, implementation, analysis, evaluation, and manuscript preparation. I declare that this thesis and the work reported herein is the best of my original knowledge, except as acknowledged in the text, and that the work was not submitted previously to any other institution.

---------- ----------

Miaomiao Ma

Summary

A variety of different methods have been suggested to classify catchment runoff or groundwater dynamics, to relate these to catchment or aquifer properties and thus to utilize inherent information of data. To that end, the correlation dimension method, a powerful nonlinear time series analysis method based on the chaos theory, has been suggested to assess the intrinsic dimensionality of time series according to Takens (1981). It can provide an assessment of the minimum number of processes that is required to map the observed dynamics. In the first study, the correlation dimension method was applied to the observed hydrographs of 35 catchments in the Federal State of Brandenburg, Germany. The intrinsic dimensionality of these catchments ranged from 2.2 to 5.8. It was uncorrelated with the results of standard time series analysis methods, such as autocorrelation, the slope of the power spectrum and the Hurst coefficient, revealing that the correlation dimension method captured information independent from these measures. The correlation dimension values did not exhibit any clear spatial patterns, but showed significant correlations with the spatial heterogeneity within the catchments. In addition, the correlation dimension method was applied to groundwater head and lake level data in the biosphere reserve Schorfheide-Chorin region. The intrinsic dimensionality of groundwater level ranged from 0.9 to 5, while lake level exhibited small variations, around 1.57 to 2. The correlation dimension values of groundwater level exhibited no correlation with the screening depth of groundwater wells, but displayed spatial patterns due to the different aquifer conditions (confined or unconfined). It seems that high correlation dimension values indicate partly confined conditions.

Most of the available hydrological models are highly over-parameterized concerning available data and encounter the equifinality problem: different model parameterizations and even different models yield the same best results, which has severe consequences with respect to model uncertainty. However, if these models are used for process identification or as a basis for the modeling of reactive solute transport, they exhibit substantial variety. The same problem exists for model applications to different boundary conditions. Thus, model validation by comparing

measured with simulated time series only is not sufficient. In the second study, we proposed a different approach based on the correlation dimension method. Simulated hydrographs from three hydrological models with increasing complexities were investigated using the correlation dimension method and the relationship between correlation dimension values and Nash-Sutcliffe efficiency values was explored. The correlation dimension method imposes additional constraints to the models and is more powerful to reduce the equifinality problem compared with the traditional Nash-Sutcliffe efficiency criteria. Therefore, the combination of the Nash-Sutcliffe efficiency criterion and the correlation dimension method detects the intrinsic property underlying the system dynamics, but also improves the prediction accuracy, serving as a promising approach for model performance evaluation. In addition, the correlation dimension analyses of model rainfall, evapotranspiration and discharge time series suggested that the hydrological models likely acted as intrinsic dimensionality reducing filters for the high-dimensional model inputs to outputs. The model reduced more intrinsic dimensionalities of simulations, if the higher model complexity was.

Keywords: correlation dimension, dominant process, equifinality, hydrological model, intrinsic dimensionality, groundwater dynamics

Zusammenfassung

Für eine Klassifizierung hydrologischer Einzugsgebiete hinsichtlich ihres Abflussverhaltens, wurden in der Vergangenheit verschiedene Auswertemethoden entwickelt. Gemeinsames Ziel war es, die den Datensätzen inhärente Informationen soweit nutzbar zu machen, dass eine Korrelation der Abflussdynamik mit den Einzugsgebiets- bzw. Aquifereigenschaften möglich ist. Zu diesem Instrumentarium gehört auch die Korrelations-Dimensions-Methode, eine auf der Chaos-Theorie basierende Methode der nicht linearen Zeitreihenanalyse. Sie wurde von Takens (1981) zur besseren Beschreibung hochdynamischer physikalier Systeme entwickelt. Die Methode ermöglicht eine Abschätzung der Mindestanzahl an Prozessen, die zur Darstellung der beobachteten System-Dynamik notwendig sind, der so genannten intrinsischen Dimensionalität. In der vorliegen Arbeit wurde in einem ersten Schritt die Korrelations-Dimension auf gemessene Abflussganglinien von 35 Einzugsgebieten des Bundeslandes Brandenburg (Deutschland) angewendet. Die ermittelten Werte für die intrinsische Dimensionalität dieser Einzugsgebiete lagen zwischen 2.2 und 5.8. Diese Ergebnisse korrelierten nicht mit den Werten der überlicherweise eingesetzten Standardmethoden der Zeitreihenanalyse wie Autokorrelation, der Steigung des Powerspektrums oder dem Hurst-Koeffizienten. Die Ergebnisse zeigen somit, dass mit der Korrelations-Dimensions-Methode zusätzliche bzw. von anderen Auswertemethoden unabhängige Informationen erfasst werden können. Die Werte der Korrelations-Dimension korrelieren signifikant mit der räumlichen Heterogenität innerhalb der Einzugsgebiete und verweisen somit auf strukturelle Einheiten mit unterschiedlicher hydrologischer Komplexität. In einem weiteren Schritt wurde die Korrelations-Dimensions-Methode auf Zeitreihen von Grund- und Seewasserständen im Biosphärenreservat Schorfheide-Chorin angewendet. Hier ergeben sich für die Zeitreihen der Grundwasserstände Korrelations-Dimensions-Werte von 0.9 bis 5. Für die Dynamik der Seewasserstände wurden Werte zwischen 1.57 und 2 ermittelt. Die Korrelations-Dimensionen der Grundwasserstände weisen keinen Zusammenhang mit der Filtertiefe der Grundwassermessstellen auf, zeigen aber räumliche Muster, die mit dem hydraulischem Zustand des Aquifers (gespannt

oder ungespannt) korrelieren. Es konnte gezeigt werden, dass hohe Korrelations-Dimensions-Werte auf gespannte Verhältnisse hindeuten.

Die intrinsische Dimensionalität ist nicht nur für eine weitergehende Interpretation hydrologischer Zeitreihen von großer Bedeutung. Die meisten der verfügbaren hydrologischen Modelle sind hinsichtlich der vorhandenen Daten stark überparametrisiert und unterliegen dem Problem der Äquifinalität, d.h. verschiedene Parametrisierungen von Modellen, sogar verschiedene Modelle liefern qualitativ die gleiche Ergebnisse. Werden die Modelle allerdings zur prozessbasierten Modellierung z.B für den reaktiven Stofftransport oder für szenariobasierte Berechnungen genutzt, zeigen sich grundlegende Abweichungen. Dasselbe Problem besteht bei der Anwendung von Modellen mit unterschiedlichen Randbedingungen. Vor diesem Hintergrund ist die Validierung von Modellen durch den bloßen Vergleich von gemessenen und simulierten Zeitreihen nicht ausreichend. Im Rahmen dieser Arbeit wurde daher ein auf der Korrelations-Dimensions-Methode basierender Ansatz zur Lösung dieses Problems entwickelt und getestet. Simulierte Abflussganglinien von drei hydrologischen Modellen zunehmender Komplexität wurden mit dem Korrelations-Dimensions-Ansatz ausgewertet und die Ergebnisse mit dem etablierten Nash-Sutcliffe-Effizienz-Kriterium verglichen. Es konnte gezeigt werden, dass die Korrelations-Dimension-Methode für die Modelle zusätzliche Beschränkungen der Freiheitsgrade setzt und hinsichtlich der Reduzierung des Problems der Äquifinalität effektiv einzusetzen ist. Die Kombination aus Nash-Sutcliffe-Effizienz-Kriterium und Korrelations-Dimension reduziert den zeitlichen Aufwand der Analyse und verbessert die Vorhersagegenauigkeit. Das vorgestellte Verfahren scheint für die Beurteilung der Leistungsfähigkeit von Modellen ein vielversprechender Ansatz zu sein. Darüber hinaus deuten die Korrelations-Dimensions-Analysen simulierter Niederschlags-, Evapotranspirations- und Abfluss-Zeitreihen darauf hin, dass hydrologische Modelle für hoch-dimensionale Modelleingaben als Filter für die intrinsische Dimensionalität wirken. Es zeigt sich, dass die intrinsische Dimensionalität der Simulationsergebnisse in dem selben Maße abnimmt, wie die Komplexität des Modells zunimmt.

Table of contents

List of figures

List of tables

1 Introduction

1.1 Background and Objectives

The complex nature of geophysical phenomena has been realized for centuries. They exhibit significant variability on temporal and spatial scales and are governed by inherently nonlinear and inter-dependent mechanisms. In order to understand and describe complex geophysical phenomena, the "system" concept was introduced. A widely accepted classification is that of deterministic systems and stochastic systems. The deterministic supporters argue that geophysical phenomena are deterministic in nature. On the other hand, the stochastic supporters argue that geophysical phenomena are obviously random in nature due to the highly variable complex nature of geophysical phenomena and our limited ability to understand such variations. Indeed, most geophysical phenomena are neither steady nor periodic. Some are not as irregular and as complex as the others. Their behavior can be classified between regular and stochastic behavior, called "chaotic" motion. Their describing theory is known as "chaos theory".

The chaotic system was first found by Lorenz (1963) when he studied a simplified system of convection rolls in the atmosphere to investigate the predictability of the weather. He found that this system was neither stochastic nor periodic and never settled down to equilibrium or to a periodic state. However, this system is sensitively dependent on the initial conditions. For instance, small errors in initial conditions (such as those due to rounding errors in numerical computation) yield widely diverging outcomes, rendering long-term prediction impossible in general. This is in contrast to the usual understanding of the physical laws, where small perturbations cannot produce arbitrarily large consequences (Rodriguez-Iturbe et al., 1989). Thus, chaotic motion should be viewed as a separate motion regarding the deterministic and stochastic motions. Sivakumar (2004a) suggest that the chaos theory might connect deterministic and stochastic systems theory, taking into account the advantages of the two systems as well as their limitations.

The fundamental idea of the chaos theory is that many phenomena appear to be complex but may be prone to very few degrees of freedom. Since the development of the chaos theory, it has been applied to a variety of geophysical time series, including oxygen

isotope records, tree-ring thicknesses, surface pressure, sunshine duration, wave amplitude, geo-potential values, temperature, rainfall and wind velocity (e.g. Sivakumar, 2004a; Sivakumar, 2009). Among them, hydrological phenomena (e.g. rainfall, runoff, groundwater flow) are often regarded as paramount examples for highly non-linear and apparent complex systems. However, they can be regarded as the outcome of simple systems with only a few nonlinear interdependent variables with sensitive dependence on initial conditions (Sivakumar, 2004a). Therefore, using the chaos theory to explore the useful information of hydrological systems will forward our understanding of the systems.

Hydrologic phenomena arise as a result of interactions between climate input and landscape characteristics that occur over a wide range of space and time scales. Due to tremendous heterogeneities in climate inputs and landscape properties, such phenomena may be complex at all scales. Consequently, the “complexity" becomes an appropriate indicator to represent the hydrological systems’ characteristics. Most of the available hydrological models are highly over-parameterized concerning available data. On the other hand, a minimum of model complexity is required to account for the expected changes of, e.g., land use, water management strategies or climates. This is a serious obstacle to using models for water resources planning and management. Thus, there is urgent need to equilibrate data and model complexity.

In spite of our technological advances, there are still many constraints in regards to data measurements, some of which are associated with unknown physical mechanisms knowledge. Even if we have all the right technologies for data measurements, we still may not be able to actually obtain the required data for reasons associated with political and societal pressures (Sivakumar, 2008). These difficulties in obtaining more accurate data force us to shift our thinking to extract more information from the available data. In fact, one of the purposes of measuring data is to learn about such mechanisms from the data themselves in an “inverse” manner (Sivakumar, 2008). Various methods have been developed to abstract different kinds of information in the data besides time series plots, such as autocorrelation analysis, power spectrum analysis, Hurst analysis and principal component analysis (PCA) etc. The autocorrelation analysis (e.g. Porporato and Ridolfi, 1997) provides a mathematical tool for displaying the temporal dependency and finding

repeating patterns in the time series, while the power spectrum analysis (e.g. Akselrod et al., 1981; Tsonis et al., 1994; Porporato and Ridolfi, 1997; Kirchner et al., 2000; Sivakumar, 2001) is often used to quantify the degree to what a catchment buffers the high-frequency part of the input signal. The Hurst coefficient (e.g. Lange, 1999) quantifies the phenomenon of persistence and reveals the long-term memory of the data. PCA (e.g. Jolliffe, 2002) can be thought of revealing the internal structure of the data in a way which best explains the variance of the data. However, it is a linear dimensionality reduction method that cannot adequately represent nonlinear relationships. Therefore, diverse nonlinear time series analysis methods based on the chaos and dynamic theory were developed, such as the correlation dimension (CD) method (Grassberger and Procaccia, 1983a), Kolmogorov entropy (Grassberger and Procaccia, 1983b), recurrence quantification analysis (Marwan et al. 2007), and the Lyapunov exponent (Wolf et al., 1985). These methods motivate scientists to explore the hidden information in the data from the special view of the "complexity" underlying the system dynamics.

General speaking, the complexity of a system with deterministic dynamics depends on the number of degrees of freedom and on how many of them are associated with sensitive dependence on initial conditions (Koutsoyiannis, 2006). The former is assessed by the dimensional analysis, while the latter is quantified by positive values of the so called Lyapunov exponents. The dimensional analysis has great generality and mathematical simplicity. It offers an approach for reducing complex physical problems to the simplest form and obtaining a quantitative answer (Sonin, 2001). At the heart of dimensional analysis is the concept of similarity, which refers to some equivalence between two things or phenomena that are actually different. A certain type of dimensional analysis method based on the temporal similarity of different periods in a data set is used to assess the "intrinsic dimensionality" or the "degrees of freedom" of data, which is defined as the minimum number of free variables required to generate the data without any significant information loss. In this catalog, the CD method (Grassberger and Proccacia, 1983a) has attracted considerable interest of scientists in atmospheric and hydrologic sciences (e.g. Sivakumar, 2004a). The obtained CD value is extremely sensitive to slight changes in the complexity of the underlying deterministic structure (DeCoster and Mitchell, 1991). The higher the CD values, the more complex the underlying system dynamics appear to be.

This information is helpful for classifying catchments regarding the observed runoff behavior and for distinguishing different dynamics in groundwater aquifer systems.

On the other hand, different rainfall runoff models or even the same model parameterized in different ways often give indistinguishable results with respect to the simulated hydrograph (Beven and Freer, 2001). However, if these models are used for process identification or as a basis for the modelling of reactive solute transport, they exhibit substantial variety, regardless of superior performance on a calibration or validation data set. The same problem exists with respect to model applications to different boundary conditions. This phenomenon is called "equifinality" and it has severe consequences with respect to model uncertainty. Thus, model validation by comparing measured with simulated time series only is not sufficient and additional constraints are needed.

Since it is impossible to model everything due to the interactions with climate and other systems in nature, therefore, the "dominant processes" (corresponding to the minimum number of independent processes) model framework becomes more practical. In that case, the number of dominant processes might be a reliable additional constraint for models. The CD method can assess the dominant processes that are required to reproduce the observed dynamics. This makes it as an ideal candidate to investigate model structure uncertainties and evaluate model performance.

Based on the advantage of the CD method to explore the underlying the system dynamics with no pre-defined assumptions, this thesis aims at applying the CD method to hydrological phenomena and develop new promising approaches for classifying catchment behavior, evaluating model performance and assessing groundwater dynamics through estimating the complexity or the number of dominant processes of observed and simulated data. The thesis addresses the following topics:

1) Which different information does the CD method provide compared with autocorrelation analysis, power spectrum analysis, Hurst analysis and principal component analysis (PCA)?

2) What are the reasons for the different dimensionalities of observed hydrographs? Are they related to anthropogenic pressure, catchment properties or meteorological data?
3) How does the CD method tackle the equifinality problem and set up a blueprint for model evaluation?
4) How does the intrinsic dimensionality change from the input to the output of conceptual rainfall runoff models?
5) What are the possible strengths and weaknesses of the CD method with respect to groundwater head and lake level time series?

1.2 Outline of the thesis

This PHD thesis is composed of eight chapters. Chapters 4 and 5 have been written as manuscripts for publication in international peer-reviewed scientific journals (Appendix I).

Chapter 1 is the introduction of this thesis. It describes the background, problem definition and the objectives of the thesis.

Chapter 2 describes the principle, calculation steps and some methodological problems of the CD method and then provides an overview on previous studies of the CD analysis.

Chapter 3 introduces the principle of three rainfall runoff models (i.e. abc model, VM model and HBV model) and the standard time series analysis methods (i.e. variance, autocorrelation analysis, power spectrum analysis, Hurst analysis and PCA). These three hydrological models with different complexities were employed as the basis of model performance evaluation in Chapter 5. While the standard time series analysis methods were used to check the redundancies between these measures and CD results in Chapter 4 and 5.

Chapter 4 describes the first CD method application, which aims at classifying the catchments regarding the runoff behavior and exploring the reasons for different dimensionalities of observed hydrographs.

Chapter 5 describes the second CD method application, which uses the CD method to reduce the equifinality problem and consequently proposes a promising evaluating approach for model performance.

Chapter 6 describes the third CD method application, which applies the CD method to assess the complexity of groundwater head and lake level and explore the independent information of groundwater dynamics.

Chapter 7 discusses the possible strength and weakness of the CD method on assessing the number of dominant processes underlying the system dynamics.

Chapter 8 gives a conclusion of this thesis. It highlights the major outcomes of this work and provides recommendations for future research.

2 Theory and applications of the Correlation Dimension method

2.1 Theory of the Correlation Dimension method

The correlation dimension (CD) is a measure of the number of "active nodes" in the system or of the effective number of degrees of freedom (Theiler, 1986). In physical terms, the CD gives a lower bound on the effective number of degrees of freedom activated in a physical process (Ding et al. 1993). It is a special case of the Renyi dimension D_q (Renyi, 1970; Wang and Gan, 1998), defined as

$$D_q = \lim_{\varepsilon \to 0} \frac{\log(\sum_{i=1}^{M(\varepsilon)} p_i^q)}{(q-1)\log \varepsilon} \tag{2-1}$$

Where $M(\varepsilon)$ is the number of boxes with a discrete probability distribution $P_i > 0$; ε is the side length of the box; q is any real number, but usually an integer number; D_0 is the box-counting dimension; D_1 is the information dimension; D_2 is the correlation dimension (CD) and $D_0 \geq D_1 \geq D_2$ (Ding et al., 1993; Wang and Gan, 1998).

The box-counting dimension (i.e. the capacity dimension or the Hausdorff dimension) was first proposed by Mandelbrot (1977) to define the dimension of a fractal set. The simple concept of this dimension is counting the number of boxes that contain one or more data points. This dimension is commonly based on narrow and intuitive definitions (Wan and Gan, 1998). The information dimension makes a step forward and takes amount of information contained in each box into consideration. This is achieved by measuring the irregularity or entropy based on the Shannon formula for a discrete probability (Wan and Gan, 1998). These two dimensions described are impractical to implement on high dimensional data, because the number of boxes increases exponentially with the dimensionality of the data, and consequently requires heavy computational load (Buchula et al., 2005). In contrast, the CD can better deal with high dimensional data compared with box-counting and information dimensions. Therefore,

the CD attracts more interests of scientists. The CD method introduced by Grassberger and Proccacia (1983a) can assess the "intrinsic dimensionality" or "degree of freedom" of time series according to Taken (1981), that is, the minimum number of dominant processes required to map the observed dynamics.

The CD method can be used as a nonlinear time series analysis method based on the theory of dynamic systems. The dynamical system can be defined as a deterministic mathematical prescription for evolving the system state forward in time (Ott, 1993). It is characterized by: (1) a phase space in which the motion of the system takes place; (2) a time evolution law that describes what future states follows from the current state; (3) a time set that can describes the moments at which movements from one position to another take place (Koutsoyiannis, 2006). The phase space is a coordinate system with the coordinates representing the dominant variables to completely describe the state of the system at any moment (Sivakumar, 2007). The elements or points of the phase space represent possible states of the dynamic system. The path in phase space followed by the system as it evolves with time is so called "orbit" or "trajectory". In dissipative dynamic systems, the trajectory of the system, after some transient time, is attracted to some subsets of the phase space. This set itself is invariant under the dynamical evolution and is called the attractor of the system (Kantz and Schreiber, 2004). One goal of estimating the intrinsic dimensionality is to extract the properties of underlying attractors based on observed or simulated time series and a reconstruction of the phase space.

2.1.1 Phase space reconstruction using the time delay embedding method

The idea behind such a phase space reconstruction is that a nonlinear system is characterized by self-interaction, and that time series of a single variable can carry the information of the entire multi-variable system dynamics. Takens' theorem (1981) guarantees that original and reconstructed attractors can be considered to represent the same dynamical system in different coordinate systems.

The time delay embedding method is a commonly used phase space reconstruction method. It takes a scalar time series $x(t_i)$ and its successive time delays to embed in the m dimensional phase space defined by

$$Y(t)=\{x(t_i),x(t_i+\tau),x(t_i+2\tau),x(t_i+3\tau),\cdots,x(t_i+(m-1)\tau)\} \quad (2\text{-}2)$$

where τ is delay time or time lag; and m is the embedding dimension. For practical purposes, the most important embedding parameter is the product $m\bullet\tau$ rather than any single parameter, because $m\bullet\tau$ is the time span represented by an embedding vector (Kantz and Schreiber, 2004). In this section, the parameters τ and m are described separately.

2.1.1.1 Time lag

The time lag τ is used to "decorrelate" a time series into a number of vectors (Wang and Gan, 1998). Each column in the matrix $Y(t)$ stands for the trajectory for the underlying system in the embedding phase space. If τ is too small, then each column of $Y(t)$ is not independent from others and little new information is included, resulting in an underestimation of the correlation dimension. On the contrary, if τ is too large, all relevant information contained by the original system and by the phase space reconstruction is lost, resulting in almost uncorrelated vector' components. Thus the data are (seemingly) randomly distributed in the embedding space (Hegger et al. 1999; Sivakumar, 2000). Various methods for selecting an appropriate τ are proposed, such as the autocorrelation function (e.g. Tsonis and Elsner, 1988; Jayawardena and Lai, 1994), the mutual information (e.g. Frazer and Swinney, 1986; Moon et al., 1995; Hegger et al. 1999), the phase portrait (e.g. Kantz and Schreiber, 2004). Among these methods, autocorrelation function is the most commonly used due to its simple concept and computational ease. It is suggested using a value of τ at which the autocorrelation function first attains a certain value, such as 0 (e.g. Holzfuss and Mayer-Kress, 1986), 0.1 (e.g. Tsonis and Elsner, 1988) or 0.5 (e.g. Schuster, 1988). Although this method displays a regular behavior representing the effect of the seasonal characteristics of the time series (e.g. Porporato and Ridolfi, 1997), these given values are too arbitrary and depend on personal choice. Frazer and Swinney (1986) pointed out that the autocorrelation function method measures the linear dependency among successive points and thus may not be appropriate for nonlinear dynamics. They proposed the mutual information method. There τ should be chosen to coincide with the first minimum of mutual information of the data

set. They also showed that the mutual information method is superior to the autocorrelation function method for choosing τ. But this method requires a large number of data points (unless the dimension is small) and is computationally cumbersome (Sivakumar, 2000). In addition, the mutual information method always involves the parameter "partition" (i.e. the number of bins) and then the estimated time lag τ depends on the parameter values. There are no clear approaches to identify the value of partition. Hence, the estimation of τ using mutual information method is uncertain. The phase portrait method is another way to identify the proper time lag τ, i.e., plotting $x(t)$ versus $x(t+\tau)$, and subsequently increasing τ until the pattern stabilizes. It is a kind of visual inspection and the τ obtained by this method is arbitrary. In summary, none of aforementioned methods can be the undisputed rule for choosing τ from our investigation. Therefore, the heuristic method with experimenting different τ is preferred in this study.

2.1.1.2 Embedding dimension

No optimal method to select the embedding dimension m exists. For a long enough data string with low enough noise, the plateau onset occurs at $m = ceiling(CD)$ (Ding et al. 1993). In principle m should be equal or greater than CD, but a too large m will add redundancy and lack of statistics, and hence m should be chosen carefully. One method to obtain the optimum embedding dimension m is the false nearest neighbors approach (e.g. Kennel et al., 1992; Rhodes and Morari, 1997; Hegger, 1999). Its principle is eliminating false neighbors by checking the neighborhood of points embedded in projected manifolds of increasing dimension (Das et al., 2002). However, the results highly rely on the involved parameters which are hard to identify. The same situation is faced by the correlation integral method, which is another approach to obtain the optimum embedding dimension m. For example, one of the common parameters for the false nearest neighbors and the correlation integral methods is the Theiler window, which is applied to exclude temporal correlated points from the pair counting. This value is usually identified by the space time separation plot (Provenzale et al., 1992). The parameter "time lag τ" is employed in this plot, which is difficult to be estimated as described in the previous section. Thus, the estimation of the Theiler window based on τ

is incredible. There are no definitive clear-cut guidelines for choosing the embedding dimension m and hence a practical approach to experiment with different τ and m is a better choice to get a reliable CD value (e.g. Tsonis et al., 1993, Sivakumar, 2000).

2.1.2 Correlation Dimension estimation

Some systems may be constrained in the case of two variables that are highly correlated where all data points plot closely to a common regression line. That means the data points do not fill the whole 2D phase space more or less homogeneously. The CD method in fact checks how homogeneous the phase space is filled by the observations, accounting for highly non-linear relationships as well. According to Taken (1981), this method can be extended from multi-variate high-dimensional data sets to uni-variate time series. That is, the time series can be represented by a set of the same time series, but with different time lags. Then, again, this new reconstructed vector usually does not fill the available phase space homogeneously. That means, the next data point cannot be any value given a series of observed subsequent data points (e.g., discharge data). Instead, the more constrained, the lower degrees of freedom are.

The CD method assumes that the data are spatially correlated in the respective phase space. The more correlations are in a data set, the smaller is the effective dimensionality of the data compared to the specified number of variables, so the smaller the inter-points distances tend to be (Lischeid and Bittersohl, 2008). The correlation integral $C(r)$ is employed to measure the spatial correlation of a data set in the phase space. It is defined as the number of pairs whose distance is smaller than the threshold r. The equation is:

$$C(r) = \lim_{n\to\infty} \frac{1}{N^2 - N} \sum_{a \neq b} H(r - \|Y(t_a) - Y(t_b)\|) \qquad (2\text{-}3)$$

where $H(u)$ is the Heaviside function with the function $H(u) = \begin{cases} 0 & u < 0 \\ 1 & u \geq 0 \end{cases}$ and $u = r - \|Y(t_a) - Y(t_b)\|$. $\|Y(t_a) - Y(t_b)\|$ denotes the Euclidian distance between two geometrical points in the reconstructed phase space, r stands for the radius of a sphere

centered on $Y(t_a)$ or $Y(t_b)$ and N is the number of data points in the spatial series (Grassberger and Proccacia, 1983a).

For a large data set, there is a relationship between the correlation integral $C(r)$ and the radius r as:

$$C(r) \approx \alpha r^{d} \tag{2-4}$$

where α is a constant and d is the correlation exponent. The value of d is determined as the mean slope of scaling range in $\log C(r)$ versus $\log r$ plot.

$$d = \lim_{r \to 0, N \to \infty} \frac{\log C(r)}{\log r} \tag{2-5}$$

Plotting $\log C(r)$ against $\log r$ yields a curved line that can usually be subdivided into three parts: (i) the depopulation range (an irregular pattern) for small values of $\log r$, (ii) the scaling range (a linear part) for intermediate values of $\log r$, (iii) the saturation range (slope approaches zero) for large values of $\log r$. The correlation exponent value is estimated from the slope of the scaling range. It must be noted that the exact delineation of the scaling region can be difficult and often requires visual inspection. Moreover, the scaling region becomes smaller and smaller with increasing of m, and eventually vanishes for large m (e.g. Ding et al., 1993; Hossain and Sivakumar, 2006). Hence the estimation of the correlation exponent partly is an empirical exercise. The correlation exponent is identified from the scaling range of $\log C(r)$ against $\log r$ plot for different embedding dimensions. Then the values of the embedding dimension m are plotted versus the correlation exponent $d(m)$. The estimated CD value typically increases with m and reaches a plateau on which the dimension estimate is relatively constant for a range of large enough m. This saturation value is the estimated CD of the analyzed signal, while the embedding dimension corresponding to the plateau onset is sufficient to estimate the dimension of the attractor. That is to say, the nearest integer above the CD provides the minimum dimension of the phase space essential to embed the attractor, while the value of the embedding dimension at which the saturation of the correlation exponent occurs

provides an upper bound on the dimension of the phase space sufficient to describe the motion of the attractor (Fraedrich, 1986). If there is no plateau in the $d(m)$ curve, it indicates that the data could be stochastic in nature or severely affected by noise. In that case the CD value cannot be estimated. Therefore, the CD method is able to distinguish chaotic motion from a simple system and stochastic motion (Theiler, 1986).

In this study, we tested different methods to identify the proper time lag and embedding dimension, such as the autocorrelation function method, the mutual information method, the phase portrait method, the false nearest neighbors and the correlation integral method. However, none of these methods was effective. Thus different τ and m were tested for each hydrograph. In this procedure, two vital plots need to be considered carefully in the application of CD method. One is the $\log C(r)$ against $\log r$ plot, whose scaling range slope gives the correlation exponent. The other is the $d(m)$ plot, whose saturation range can finally achieve the reliable CD values. These two plots directly determine the reliability of the final CD results and hence have to be evaluated thoroughly.

The scaling range is defined as the linear part of $\log C(r)$ against $\log r$ plot. Noise in the data will render a portion of the plot for low $\log r$ unusable. The finiteness of the data set will further curtail the extent of the nearly linear range (DeCoster and Mitchell, 1991). In addition, the slope is often not absolutely constant within the scaling range for finite real-world data sets. Since no existing technique is able to identify the scaling range precisely, we suggest using the distinct windows technique (Figure 2.1) to determine the slope of the scaling range. In our case, the slope has been determined for 50 subsequent windows separately (Figure 2.1). This approach makes the scaling range more visible and subsequently helps to improve the visual inspection.

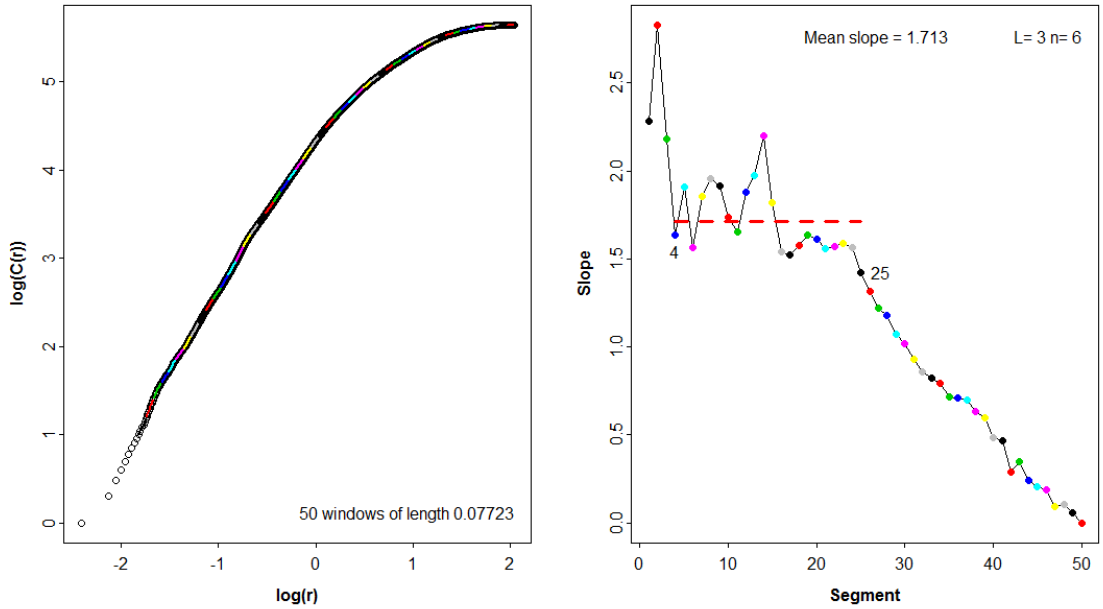

Figure 2.1 Distinct windows technique of $\log C(r)$ against $\log r$ plot. 50 windows marked by different colors have been used to determine the slope of the scaling range (left figure). The slopes of these 50 segments are given in the right figure. Order of the segments is the same in both graphs, facilitating identification of the scaling range and subsequent determination of the corresponding mean slope for these segments.

The saturation range of the $d(m)$ plot determines the estimated CD value of the analyzed signal. This saturation range should be visible for at least three different time lags (Rodriguez-Iturbe et al., 1989; Sivakumar, 2000). The reliable CD value could be estimated by the mean value of the saturation range of $d(m)$ plot. The analysis was performed using the R package (R Development Core Team 2006). No automatic algorithm was found that could accomplish this task. Therefore, the CD algorithm remained a semi-automatic program requiring some expertise. In summary, the main steps of the CD method were:

1) Reconstructing the phase space using the time delay embedding method to represent the underlying system dynamics.
2) Calculating the Euclidian distances among the points in the reconstructed phase space.

3) Plotting the $\log C(r)$ versus $\log r$ and determining the scaling range slope in this plot using the distinct windows technique.
4) Plotting correlation exponent d versus embedding dimension m for different values of time lags, and then estimating the CD value as the mean value of the saturation range in $d(m)$ plot.

2.1.3 The minimum data size

The CD method is assuming that the time series is infinite. Lacking of sufficient data will not only delay the plateau onset, but also make the deviation from the plateau behavior occur at smaller values of m, thus shortening the plateau length from both sides (Ding et al., 1993). A small data size may result in a significant underestimation of the CD value. Sivakumar (2000) investigated the reliability of the CD estimation in short hydrological time series. The results reveal that the accuracy of the dimension estimation depends primarily on whether the length of the time series is sufficient to represent the changes that the system undergoes over a period of time, rather than the data size in terms of the number of values in the series. The minimum data size was first tackled by Smith (1988), who concluded that this number is a function of the embedding dimension and equal to 42^m, where m is the smallest integer above the dimension of the attractor. Nerenberg and Essex (1990) demonstrated that Smith's procedure to obtain the 42^m was flawed and that the data requirements might not be so extreme. They suggested that the minimum number of points required for the dimension estimation is $N_{\min} \approx 10^{2+0.4m}$. In fact, these requirements would impose serious restrictions to any application of hydrological data. However, the mentioned research is not absolutely correct according to Lorenz (1991). He pointed out that those different variables could yield different estimations of CD values and selected suitable variables could sometimes yield a fairly good estimate even if the number of the points were not large. A similar conclusion was drawn by Sivakumar (2000; 2005b). He suggested that the minimum data size required for the CD estimation may largely depend on the type and dimension of the attractor. Therefore, one reasonable way to determine the minimum sufficient data size is to compute the CD for different sample sizes until no significant changes can be observed.

2.1.4 How to handle noisy data?

The CD method is designed under the assumption that the time series is noise free. Inaccurate estimations may be due to the presence of noise (e.g. Schertzer et al., 2002; Kantz and Schreiber, 2004). However, all hydrological or real measurements are to some extent contaminated by noise. Thus, the severity of the influence of noise on data analysis techniques depends largely on the level and the type of noise. There are two types of noise, measurement noise and dynamical noise. Measurement noise refers to the corruption of observations by errors, which are independent of the dynamics and can be reduced by diverse techniques. For example, noise in discharge data might result from the impact of wind, wetting, gage exposure, instrumentation, round-off errors and human errors in manually reading discharge meters (Sivakumar, 2000). Dynamical noise, in contrast to measurement noise, is a feedback process where the system is perturbed by a small random amount at each time step. This type of noise arises from the propagation of minor random fluctuations in the settings of the main system parameters causing random-like fluctuations that are not specific to the system (Sivakumar et al., 1999). It might also be caused by the influence of intrinsic system events taking place at random (Schouten et al., 1994). Dynamical noise directly influences the evolution of the system in time. Currently, there are no techniques available to remove dynamical noise and it is debatable whether dynamical noise should be removed at all.

The presence of noise influences the estimation of the CD primarily from the identification of the scaling region. Noise may corrupt the scaling behavior at all length scales, but its effects are significant especially at smaller length scales. It has been found that even small level of noise could conceal possible scaling behavior and hence significantly complicate estimates of CD (Schreiber and Kantz, 1996). It is necessary to remove the noise using noise reduction techniques without altering the delicate and fundamental nonlinear interactions in the signal (Porporato and Ridolfi, 1997). Diverse noise reduction methods are developed, such as the simple noise reduction method (e.g. Schreiber and Grassberger, 1991), the local projective noise reduction method (e.g. Schreiber, 1993) and the systematic noise reduction method (e.g. Sivakummar et al., 1999). Elshorbagy et al. (2002a) reported that the commonly used algorithms for noise

reduction in hydrological data might remove a significant part of the original signal and introduce an artificial chaotic behavior to the data. They recommended that the current noise reduction techniques should be applied only for better estimation of chaotic invariants and that the raw data should always be the basis. Previous studies (e.g. Sivakumar, 2000, Schouten et al., 1994) concluded that noise reduction methods might not be effective when the noise level is high. From our investigations of the above methods, none of these approaches is effective to reduce complex noise hidden in the data, if the noise level is high. As these methods always imply sophisticated procedures including the selection of new parameter values and the noise level identification, we did not follow that path. Instead, given that the tolerant noise level of the CD method is small (Kantz and Schreiber, 2004), we preferred less noisy data for the CD analysis.

2.2 Applications of the Correlation Dimension method in the literature

The inherent nonlinear nature of hydrologic system and the associated processes make these systems complex. Thus complexity is a central and fundamental characteristic of the hydrological system. To assess the complexity of hydrological system dynamics, the CD method was suggested. Multiple CD method applications from previous studies are summarized as follows.

2.2.1 Chaos investigation

The CD method has been used to detect the low-dimensional chaotic process of time series (e.g. Rodriguez-Iturbe et al., 1989; Jayawardena and Lai, 1994; Porporato and Ridolfi, 1997; Kim et al., 2001; Sivakumar, 2002; Hossain and Sivakumar, 2006; Gaume et al., 2006), since the CD values are extremely sensitive to slight changes in the complexity of the underlying deterministic system dynamics. The results indicate that:

1) The phase-space reconstruction with time delay embedding approach (Section 2.1.1) has the ability to represent the dynamics of geophysical phenomena and the CD is an useful indicator to reflect the properties (e.g. variability, irregularity and complexity) and uniqueness of a particular geophysical time series (Sivakumar, 2004a; Sivakumar and Singh, 2012);

2) The CD values of river flow are much easier to be obtained than rainfall time series. A possible implication may be that the mechanisms occurring in the aquifers and catchments are more deterministic in nature and more predictable than that occurring in the atmosphere (Sivakumar, 2004a).
3) The CD analysis has encountered some methodological problems, such as insufficient data size, the identification of proper time lag and embedding dimension, noise level determination and reduction, intermittency, the presence of zero rainfalls and high autocorrelation (e.g. Tsonis et al., 1993, 1994; Sivakumar, 2000; Sivakumar, 2005a; Koutsoyiannis, 2006).

It should be noted that the CD method poses inherent limitations when it is used for chaos identification. For example, observation of the finite CD value can only be taken as a preliminary indicator of the presence of chaos instead of as strong evidence, because finite CD value could be observed even for linear stochastic systems as well (e.g. Osborne and Provenzale, 1989).

Since the beginning of the 1990s, a period of reflection has ensured, corresponding with a critical analysis of the results obtained, and many phenomena originally imputed to be chaotic have actually been shown not to be so after more accurate investigations (e.g. Provenzale et al., 1994). One reason is the lack of investigative methods which provide sufficient conditions to identify the existence of a chaotic dynamics. Emblematic of this is the convergence of the correlation integral, usually understood as the principle if not unique sign of deterministic chaos. It has shown how some phenomena, though not being chaotic, present a convergent correlation integral (Porporato and Ridolfi, 1997). The other reasons are the limitation of the available sample, the insufficient sampling frequency, and the presence of noise etc.

With further developments in nonlinear dynamics in the 1990s, such as the surrogate data method, the false nearest neighbor algorithm, noise reduction and missing data estimation, attempts have been made to improve the identification of chaos in geophysical phenomena. The inverse approach to identify the chaos is proposed (e.g. Porporato and Ridolfi, 1997; Sivakumar, 2007). It uses the local approximation method (Farmer and Sidorovich, 1987; Casdagli, 1989) to identify the transformed function from starting point

to future behavior and then performs nearest neighbor searching procedure to predict the time series. Finally it checks the following criteria (i.e. prediction accuracy, prediction accuracy against the embedding dimension, against the lead time and against the number of neighbors) to assessing the general performance of nonlinear deterministic local approximation method. This inverse approach is generally much more reliable than the CD method for the chaos identification, because it is essentially based on prediction accuracy (Sivakumar, 2007).

2.2.2 Disaggregating the rainfall and stream flow series to different temporal scales and reconstructing the missing data

When chaos is found in hydrological time series, the chaotic dynamic approaches based on the CD method are suggested to estimate the missing data and to disaggregate the rainfall and stream flow time series from one temporal scale to another (e.g. Sivakumar et al., 2001; Sivakumar et al., 2004; Elshorbagy et al., 2002b). For these mentioned studies, the CD method is applied to investigate chaos, which occurred in the data transformed weights distribution series. They are extensions of chaos investigation using the CD method.

2.2.3 Assessing the number of dominant processes

The CD method has been used to assess the number of dominant processes (e.g., Tongal et al., 2012). The dominant processes concept was proposed (Grayson and Blöschl, 2000) to capture the essential features of hydrological systems. It assumes that simple models with only a few dominant parameters could capture the essential features of a given catchments response to precipitation events. This assumption is consistent with the fundamental idea of chaos theory (i.e. geophysical phenomena may be governed by a very few degrees of freedom). A logical way to identify the dominant processes governing a system is to evaluate the sensitivity of the system to each individual process through a multi-variable sensitivity analysis, and then select those have a ‘noticeably significant’ influence on the system sensitivity analysis (Sivakumar, 2004b). In that case, a promising procedure is proposed (Sivakumar, 2004b):

1) Determine the number of dominant processes using the CD method.
2) Identify the dominant processes of the system via expert knowledge.
3) Conduct sensitivity analysis to arrange the dominant processes in the order of their extent of dominance on the system.

2.2.4 Classification of catchments

Based on the number of dominant processes, Sivakumar et al. (2007) used the "region of attraction of trajectories" to classify the system as potentially low-, medium-, or high-dimensional. Considering the complexity and nonlinearity in hydrologic systems, the system complexity is an appropriate basis for a classification framework and nonlinear dynamic concepts constitute a suitable methodology for assessing system complexity. A catchment classification framework with an aim to subdivide catchments into different groups and sub-groups on the basis of their salient characteristics (e.g. data and process complexity) is invoked to provide directions for model developers on the level of model complexity (Sivakumar and Singh, 2012). This research is still in a state of infancy, but it is a new way to classify catchments. The results indicate that the regionalization approach, as arguably one of the most important aspects of extrapolation/interpolation of hydrologic data and for predictions in ungauged basins, may not always be the right way to classification.

In summary, most previous studies of the CD method we discussed above were employed on uni-variate time series. The CD value obtained stands for the number of dominant processes underlying the system dynamics. In the following chapters, we still apply the CD method to the uni-variate time series for classifying and assessing the complexity of observed and simulated data in hydrological systems.

3 Rainfall-runoff models and standard time series analysis

3.1 Rainfall-runoff models

For this study, three rainfall-runoff models of different complexity (Table 3.1) were applied to the same data sets.

Table 3.1 Comparison of abc, VM and HBV model structures

Classification	abc model	VM model	HBV model
Input data	Precipitation	Precipitation	Precipitation & Temperature
		Pan evaporation	Potential evapotranspiration
Modules	Water balance	Three layers evapotranspiration	Snow accumulation and melt
		Vertical-mixed runoff production	Soil moisture accounting
		Runoff separation with free water storage reservoir model	Response function
		Rive routing and response	Rive routing
Sensitive parameters	3	7	8
Complexity	Simple	Medium complex	Higher complex

3.1.1 abc Model

The abc hydrological model (Fiering, 1967) is a simple linear water balance calculation model relating precipitation to evapotranspiration, groundwater storage, groundwater outflow and stream flow with only the precipitation as input. It assumes that the watershed behaves like a linear reservoir. The abc model (Figure 3.1) calculates the runoff as:

$$Q_t = R_{quick} + R_{base} = (1-a-b) \bullet P_t + c \bullet S_{t-1} \quad (3\text{-}1)$$

$$S_t = a \bullet P_t + (1-c) \bullet S_{t-1} \quad (3\text{-}2)$$

Where Q_t, P_t, and S_t are the overall runoff, precipitation and groundwater storage at time t respectively [mm/day]; R_{quick} is the quick flow runoff [mm/day]; R_{base} is the base flow runoff [mm/day]; a is the proportion of precipitation entering the

groundwater storage; b is the proportion of precipitation lost as evapotranspiration; c is the proportion of the groundwater storage discharging into the runoff.

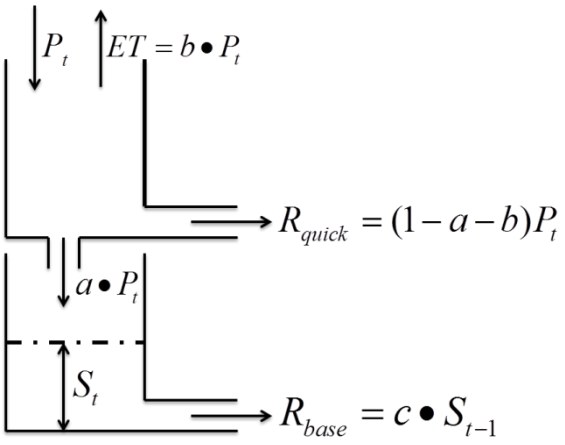

Figure 3.1 The schematic diagram of abc model referred to Fiering (1967). Where ET is the actual evapotranspiration [mm/day].

There are three parameters (i.e. a, b and c) in this model. The constraints of parameters are:

$$\begin{aligned} &0 \le a,b,c \le 1 \\ &0 \le a+b \le 1 \\ &S_t \ge 0 \end{aligned} \qquad (3\text{-}3)$$

3.1.2 Vertical-Mixed runoff (VM) model

Xinanjiang model (Zhao, 1992) has been successfully and widely applied in China since its development. This model is only suitable to humid and semi-humid regions due to its runoff formation on saturation of storage (i.e. Dunne runoff). However in practice, runoff formation in excess of infiltration (i.e. Horton runoff) always exists concurrently with the Dunne runoff in the rainfall-runoff transfer processes. Therefore, both formations should be considered and incorporated into the Xinanjiang model. The ways to distinguish the portion of these two runoff mechanisms in the mixed runoff have a significant influence on the quality of model applications. In that case, Bao and Wang (1997) proposed the vertical-mixed method to solve the problem of co-existence of two runoff mechanisms

and then constructed the VM model. This model can be used to simulate runoff in different climate zones and has achieved successful applications in many Chinese catchments. Compared with the original Xinanjiang model, the VM model improves the runoff production module. The other three modules (i.e. evapotranspiration, runoff separation and runoff concentration) keep the same principles with the original Xinanjiang model.

When the rainfall drops on the ground, the runoff is generated and divided into overland flow (i.e. surface water) and infiltration water by the improved Green-Ampt infiltration capacity distribution curve. Afterwards, the infiltration water is in preference to compensate the soil water content until it reaches the field capacity. Then the left water percolates down into the groundwater.

The calculation of the VM model (Figure 3.2) is divided into surface runoff and underground runoff. The surface runoff is calculated as the Horton runoff formation and determined by rainfall intensity and antecedent soil water capacity, whereas the underground runoff, including the interflow and groundwater, is calculated as the Dunne runoff formation and determined by antecedent soil water capacity and actual infiltration volume (Qu et al., 2007). The tension water capacity curve is introduced to reflect non-uniform distribution of tension water capacity in the basin. It is combined with the improved Green-Ampt infiltration curve to calculate the surface runoff. The formulas can be expressed as:

$$RS = P - E - FA \tag{3-4}$$

$$\begin{cases} FA = FM - FM[1 - \dfrac{P - E}{FM(1 + BF)}]^{1+BF} & P - E < FM(BF + 1) \\ FA = FM & P - E \geq FM(BF + 1) \end{cases} \quad \text{for all} \tag{3-5}$$

$$FM = FC(1 + KF\frac{WM - W}{WM}) \tag{3-6}$$

Where RS is the surface runoff [mm/day]; P is precipitation [mm/day]; E is the actual evapotranspiration [mm/day]; FA is the actual infiltration volume [mm/min]; FM is the areal mean infiltration capacity [mm/min]; BF is the exponent of the free water capacity curve, reflecting the spatial distributional characteristic of infiltration capacity; FC is the constant infiltration rate [mm/min], KF is the sensitive coefficient of influence of soil moisture deficit to infiltration rate; WM is the areal mean tension water capacity [mm]; W is the areal actual soil water capacity [mm].

The formulas to calculate the underground runoff are of the form:

$$RR = \begin{cases} FA + W - WM + WM(1 - \dfrac{FA + aa}{WM(1+B)})^{1+B} & FA + aa < WM(1+B) \\ FA + W - WM & FA + aa \geq WM(1+B) \end{cases} \quad \text{for all (3-7)}$$

$$aa = WM(1+B)[1 - (1 - \frac{W}{WM})^{\frac{1}{1+B}}] \qquad (3\text{-}8)$$

Where RR is the underground runoff [mm/day]; B is the exponent of the tension water capacity distribution curve; aa is the value of Y-coordinate corresponding to the initial mean tension water capacity W .

The total runoff is the sum of surface runoff RS and underground runoff RR. The inputs of VM model are precipitation and pan or potential evaporation. The output hydrograph is first simulated from each sub-basin and then routed down the channels to the main basin outlet. There are 16 parameters in total and 10 sensitive parameters (Table 3.2) in the VM model. Among these parameters, flow routing parameters (storage coefficient with units of time KE and flow weighting factor XE) are not sensitive for the daily runoff simulation results, since the basin was not divided into sub-basins in this study. The outflow coefficient of the free water storage to interflow KI can be calculated by the bidirectional structural constraint, which is an empirical equation based on hundreds of model applications in China, defined as:

$$KI + KG = 0.7(\frac{1001}{1+A})^{0.052} \quad (3\text{-}9)$$

Where A is the catchment area [km^2], KG denotes the outflow coefficient of the free water storage to groundwater.

Consequently 7 sensitive parameters were left in this study.

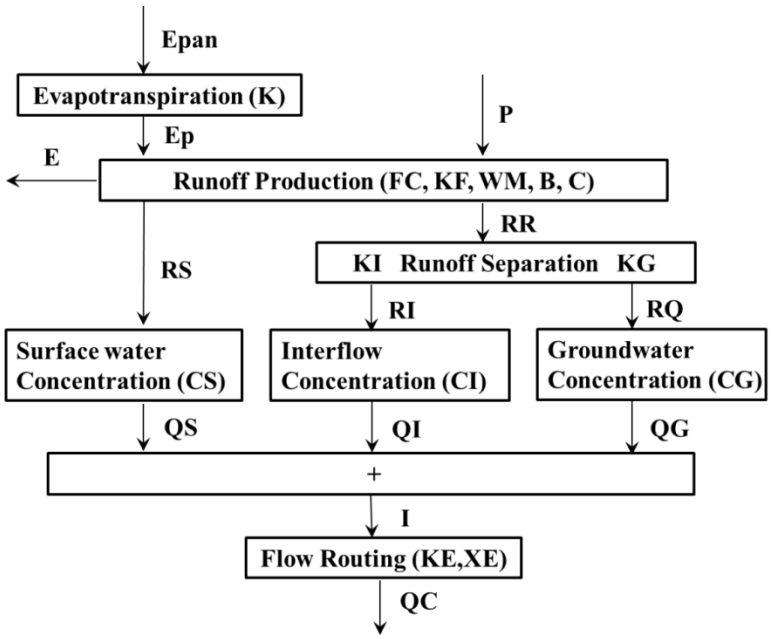

Figure 3.2 Computation flow chart of the VM model referred to Bao and Wang (1997). Note that all symbols inside the blocks are module names and parameters (Table 3.2), while those outside the blocks are inputs, outputs and internal variables. Where *Epan* is pan evaporation [mm/day]; *Ep* is potential evapotranspiration [mm/day]; *E* is the actual evapotranspiration [mm/day]; *P* is precipitation [mm/day]; *RR* is infiltration water volume [mm/day]; *RS* , *RI* and *RG* represent three runoff separation components, with surface water [mm/day], interflow [mm/day] and groundwater [mm/day] respectively; *QS* , *QI* and *QG* are simulated surface runoff [m^3/s], interflow [m^3/s] and groundwater discharge [m^3/s] from sloping surfaces respectively; *I* is the total runoff discharge from catchment sloping surfaces [m^3/s]; *QC* is the calculated or simulated runoff discharge [m^3/s].

Table 3.2 Parameters of the VM model. The bold parameters are sensitive parameters.

Component	Method	Parameter	Physical meaning	Empirical range
Evapotranspiration	Three layers evapotranspiration pattern	**KC**	Ratio of potential evapotranspiration to pan evaporation	0.1-1.1
		WUM	Average soil moisture storage capacity of the upper layer	5mm for deforested area and 20mm for forested area
		WLM	Average soil moisture storage capacity of the lower layer	60mm for deforested area and 90mm for forested area
		C	Coefficient of the deep evapotranspiration	0.09-0.2. It depends on the proportion of the catchment area covered by vegetation with deep roots
Runoff production	Vertical-mixed runoff formation	**WM**	Areal mean tension water capacity	120-250 mm
		KF	Coefficient of influence of soil moisture deficit to infiltration rate	
		B	Exponent of the tension water capacity distribution curve	0.1-0.4, Varying from 0.1 for catchment area less than 10 km^2 to 0.4 for catchment area larger than 1000 km^2
		FC	Constant infiltration rate	
Runoff separation	Free water storage reservoir model	BF	Exponent of the free water capacity distribution curve	1-1.5
		KI	Outflow coefficient of the free water storage to interflow	Calculated by bidirectional structural constraint
		KG	Outflow coefficient of the free water storage to groundwater	The sum of KI and KG is between 0.7-0.8
Runoff concentration	Linear reservoir; Muskingum successive routing method	**CS**	Recession coefficient of surface water	0-1 and $CS < CI < CG$
		CI	Recession coefficient of interflow	0-1 and $CS < CI < CG$
		CG	Recession coefficient of groundwater	0.95-0.998
		KE	Storage coefficient with units of time	Determined by hydraulic formulas
		XE	Flow weighting factor	0.0-0.5, determined by hydraulic formulas

3.1.3 HBV model

The HBV (Hydrologiska Byråns Vattenbalansavdelning) model, a conceptual rainfall-runoff model, has been successfully applied to many catchments in Sweden and abroad (e.g. Zhang and Lindström, 1996). It has different versions and here we used the HBV-light version and the standard model structure described by the Lindström et al. (1997). This model uses the daily rainfall, temperature and potential evaporation as inputs to simulate the daily discharge. It contains snow accumulation and melts, soil moisture accounting, response function and river routing modules (Figure 3.3).

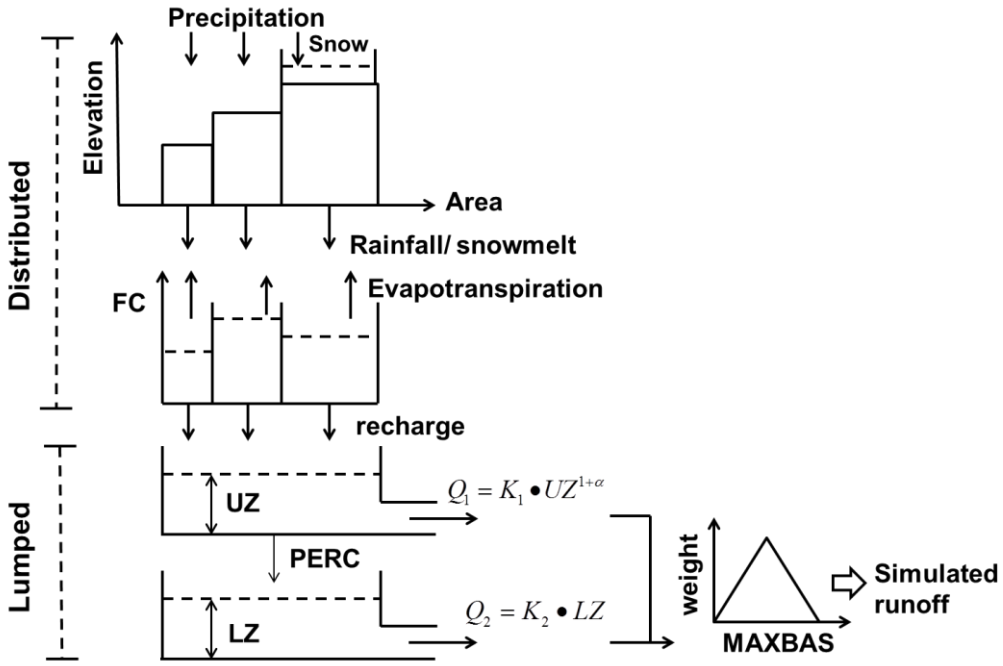

Figure 3.3 The HBV model structure referred to Lindström et al. (1997). Where Q_1 and Q_2 are outflow from upper and lower groundwater response boxes respectively [m^3/s]; UZ and LZ stand for upper and lower ground water storage respectively [mm]; K_1, K_2, FC, $PERC$, $MAXBAS$ and α are the model parameters (Table 3.3).

Precipitation is simulated to be either snow or rain depending on the threshold temperature TT. All precipitation simulated to be snow is multiplied by a snowfall correction factor SFCF. Snowmelt is calculated by a degree-day method (Equation 3-10).

Melt water and rainfall is retained within the snowpack until it exceeds a certain fraction CWH of the water equivalent of snow. Liquid water within the snowpack refreezes (Equation 3-11). Rainfall and snowmelt (P) are divided into water filling in the soil box and groundwater recharge depending on the relation between soil moisture content SM and its maximum value FC (Equation 3-12). Actual evaporation from the soil box equals the potential evaporation if SM/FC is above LP, while a linear reduction is used when SM/FC is below LP (Equation 3-13). Groundwater recharge is added to the upper groundwater box and the water percolates from upper to the lower groundwater box. Runoff from the groundwater boxes is computed as the sum of two linear outflows by linear reservoir function (Equation 3-14). The recession components threshold of upper groundwater box is defined by a nonlinear drainage equation. The runoff is finally transformed by a triangular weighting function to give the simulated runoff (Equation 3-15) (Seibert, 2005).

$$melt = CFMAX(T(t) - TT) \tag{3-10}$$

$$refreezing = CFR \bullet CFMAX(TT - T(t)) \tag{3-11}$$

$$\frac{rech\arg e}{P(t)} = (\frac{SM(t)}{FC})^{\beta} \tag{3-12}$$

$$E_{actual} = E_{potential} \bullet \min(\frac{SM(t)}{FC \bullet LP}, 1) \tag{3-13}$$

$$Q_{GW(t)} = Q_1 + Q_2 = K_1 \bullet UZ^{1+\alpha} + K_2 \bullet LZ \tag{3-14}$$

$$Q_{sim(t)} = \sum_{i=1}^{MAXBAS} (\int_{i-1}^{i} \frac{2}{MAXBAS} - \left| u - \frac{MAXBAS}{2} \right| \frac{4}{MAXBAS^2} du) \bullet Q_{GW(t-i+1)} \tag{3-15}$$

Where $P(t)$, $T(t)$, $SM(t)$, $Q_{GW(t)}$ and $Q_{sim(t)}$ are precipitation, temperature, soil moisture, groundwater discharge and simulated discharge at time t. $CFMAX$, CFR, FC, LP, K_1, K_2, α and $MAXBAS$ are model parameters.

For both the snow and soil routine, calculations are performed for each different elevation zone, but the response routine is a lumped representation of the catchment. In this study, we ignored the different vegetation zones and elevation zones due to lack of corresponding data. Noted the previous research on the HBV model parameters sensitivity analysis (e.g., Harlin and Kung, 1992; Abebe et al., 2010), 8 sensitive parameters are employed in this model (Table 3.3).

Table 3.3 Parameters of HBV model. The bold parameters are sensitive parameters.

Module	Parameter	Definition	Empirical range
Snow accumulation and melt	TT (℃)	Threshold temperature	[-2, 0]
	SFCF	Snowfall correction factor	[0.2, 1]
	CFMAX (mm/ ℃/d)	Degree-day factor	[1, 4]
	CFR	Refreezing coefficient	0.05
	CWH	Water holding capacity	0.1
Soil moisture accounting	**FC** (mm)	Maximum soil moisture content	[200, 850]
	LP (mm)	Limit for potential evapotranspiration	[0.2, 1]
	BETA	Shape empirical coefficient	[1, 4]
Response function	**Alpha**	Response box parameter	[0, 0.5]
	K1(1/d)	Recession coefficient from upper storage	[0.07, 0.2]
	K2 (1/d)	Recession coefficient from lower storage	[0.005, 0.07]
	PERC (mm/d)	Percolation from upper to lower response box	[1, 2.5]
River routing	**MAXBAS** (d)	Transformation function parameter	[2, 5]

3.1.4 The Nash-Sutcliffe efficiency criterion (NSc)

The NSc efficiency criterion is used widely in hydrology. It measures the fraction of the variance of observed flows explained by the model in terms of the relative magnitude of the residual variance (‘noise’) to the variance of the flows (‘information’) (Nash and Sutcliffe, 1970; Yapo et al, 1996). The optimal value is 1.0 and the reasonable NSc value should be larger than 0.0. Its function is:

$$NSc = 1 - \frac{\sum_{i=1}^{n}(Q_i - q_i)^2}{\sum_{i=1}^{n}(Q_i - \overline{Q})^2} \quad \overline{Q} = \frac{1}{n}\sum_{i=1}^{n} Q_i \tag{3-16}$$

Where Q_i is the observed discharge at the time i; q_i is the simulated discharge at the time i; $\overline{Q}$ is the observed average discharge; i is the index of time steps; n is the total number of time steps.

3.2 Standard time series analysis methods

Various standard time series methods are developed to abstract different kinds of information in the data besides time plots, such as variance, autocorrelation analysis, power spectrum analysis, Hurst analysis and PCA etc.

3.2.1 Variance

In probability theory and statistics, variance is one of several descriptors of a probability distribution, describing how far the numbers lie from the mean (expected value). The variance of a random variable X is its second central moment of the expected value of the squared deviation from the mean $\mu = E[X]$. The corresponding equation is:

$$Var(X) = E[(X - \mu)^2] \quad \mu = E[X] \tag{3-17}$$

Where E is the expected value operator; μ is the mean value of time series X.

3.2.2 Autocorrelation analysis

The autocorrelation analysis investigates the temporal correlation of a series with itself (Equation 3-18). It can be used as a measure for the “smoothness” of a time series.

$$R(\tau) = \frac{E[(X_t - \mu)(X_{t+\tau} - \mu)]}{\sigma^2} \quad \mu = E[X] \tag{3-18}$$

Where τ is the time lag; E is the expected value operator; μ is the mean value of time series X ; t is the time scale; σ^2 is the variance of time series and calculated from equation 3-17.

In general, the autocorrelation function of a random process fluctuates about zero, indicating that the process at any instance of time has no memory of the past at all, while the autocorrelation function of a periodic process is also periodic, indicating the strong relation between values that repeat over and over again (Sivakumar, 2001). For a chaotic process, the autocorrelation function decays exponentially with increasing lag, because the points are not independent of each other and self-similarity is present (Sivakumar, 2001). In this study, autocorrelation values were determined for time lags equal to 1 day, 5 days, 10 days and 30 days. The R language function "acf" was used for the autocorrelation analysis.

3.2.3 Power spectrum analysis

Power spectrum is the Fourier transform of the autocorrelation function and can be seen as a conversion from time domain to frequency domain. It is often used to quantify the degree to what a catchment buffers the high-frequency part of the input signal (e.g. Akselrod et al., 1981; Tsonis et al., 1994; Porporato and Ridolfi, 1997; Kirchner et al., 2000; Sivakumar, 2001) (Figure 3.4). In this study, the R language function "spectrum" was used for the power spectrum analysis and its function is:

$$E(f) \propto f^{-\beta} \tag{3-19}$$

Where f is the frequency; β is an exponent; E is the expected value operator.

The linear slope of regression line fitted to the spectrum in the logarithmic plot was compared to the CD values, where $1/300$ day was taken as the minimum frequency. The

slope can be used as a measure to what degree the catchment behaves as a low-pass filter of the precipitation input signal. In general, for a random process, the power spectrum oscillates randomly about a constant value, indicating that no frequency explains any more variance of the sequence than any other frequency, while for a periodic or quasi-periodic sequence, only peaks at certain frequencies exist (Sivakumar, 2001). It should be noted that the identification of scaling regime and the estimation of spectral exponents depend on individual judgment. Therefore, discrepancies and uncertainties are unavoidable and the caution is needed when uses the power spectrum analysis.

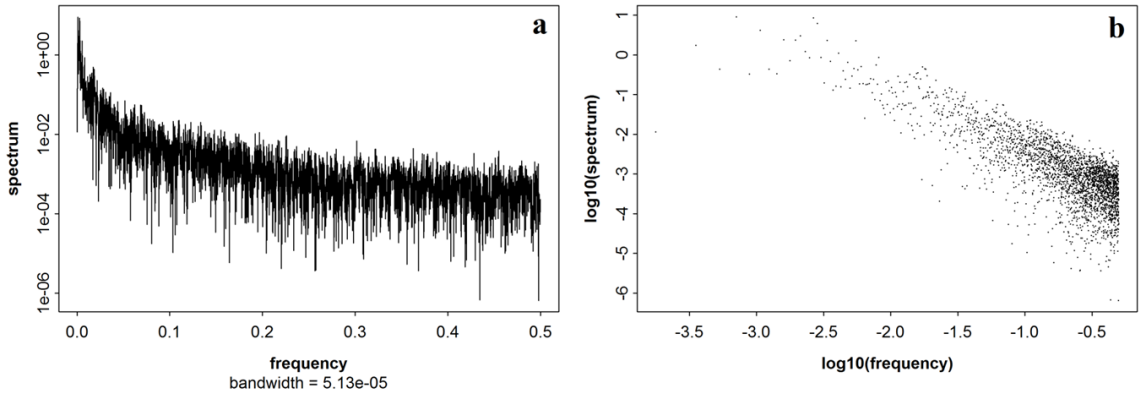

Figure 3.4 (a) the power spectrum density versus frequency plot; and (b) its corresponding logarithm power spectrum plot

3.2.4 Hurst analysis

The phenomenon of persistence (long term memory within a time series) is quantified by calculating the Hurst exponent H. A common method is the rescaled range or R/S statistics (e.g. Mandelbrot and Wallis, 1969; Lange, 1999). Here, R denotes the total range of deviation from a linear increase and S denotes the standard deviation of the series at a certain time scale.

Given a time series X of length n and a time scale k, the quantity q versus k for various values of n is plotted and it is defined as:

$$q(n,k)=\frac{R(n,k)}{S(n,k)}=\frac{\max\limits_{0\le i\le k}D(n,i,k)-\min\limits_{0\le i\le k}D(n,i,k)}{\sqrt{\frac{1}{k}\sum\limits_{i=n+1}^{n+k}(X(t_i)-\overline{X}(n,k))^2}} \tag{3-20}$$

$$D(n,i,k)=Y_{n+i}-Y_n-\frac{i}{k}(Y_{n+k}-Y_n)\quad Y_n=\sum_{i=1}^{n}X(t_i) \tag{3-21}$$

Where Y is the partial sums from the original series X; D is the deviation from a linear increase; and i is the index of time scale.

The expected persistence behaviour is

$$q\propto k^H \tag{3-22}$$

Where H is Hurst exponent. If Hurst scaling is found, then the expected H values are $0\le H\le 1$. A value in the range $0<H<0.5$ indicates anti-persistent behavior (i.e. a time series with long-term switching between high and low values in adjacent pairs). This is hardly observed in hydrological data. $H=0.5$ is ordinary Brownian motion, revealing that the differences between subsequent data points are uncorrelated. A value in the range $0.5<H<1$ is fractal Brownian motion and exhibits persistent behavior. It indicates a time series with long-term positive autocorrelation, meaning that a high value in the series will probably be followed by another high value and the values will also tend to be high for a long time.

3.2.5 Principal component analysis (PCA)

PCA is a distance preservation method that uses the simple criteria (i.e. maximizing the variance preservation or minimizing the reconstruction error) combined with a basic linear model. It can be thought of revealing the internal structure of data in a way which best explains the variance of data. It performs an orthogonal transformation to convert a set of observations of correlated variables into few components which are independent from each other. This transformation is defined in such a way that the first principal component explains most variance in the data and each succeeding component explains

most of the remaining variance under the constraint that it is uncorrelated with the preceding components.

Mathematically PCA is based on the single value decomposition of the correlated matrix of z-normalize data (i.e. normalized to zero mean and unit variance). It extracts axes (eigenvectors) within a multidimensional data set which are along the main directions of the data (i.e. explain most of the variance of the data). The square eigenvalues gives the proportion variance in the high dimensional data explained by the single principal component (Thomas et al., 2012). The principal components are sorted according to their total explained variance. The projections of the data on the new axes are called scores, while correlations of the scores with the measurements yield the loadings of the components, which describe the positive or negative importance of principal components to explain the input data (Thomas et al., 2012). Thus, the loadings can be used to analyze the spatial pattern of single component. The squared loadings give the amount of explained variance at a gauge with respect to the principal components (Thomas et al., 2012).

Thomas et al. (2012) applied the PCA to the 15 years daily discharge data of 37 catchments in the Federal State of Brandenburg, Germany. However, PCA assumes that observed variables are linear combinations of the latent ones and yields a linear projection of the observed variables. Thus, it cannot adequately represent nonlinear relationships and is often used as a benchmark of nonlinear time series analysis methods (e.g., Lischeid and Bittersohl, 2008; Lischeid, 2009). In this study, we employed the CD method, one of nonlinear time series analysis method, on the similar discharge data in order to find out the relations between the CD and PCA results.

4 The Correlation Dimension analysis of observed discharge time series in small catchments

Hydrological models are often developed for specific situations and thus their extensions and generations to other situations are difficult. Therefore, it is necessary to classify the catchments regarding their available meteorological data and catchment properties. This chapter describes the first study of the CD method applications. In this study, the CD method is applied to observed discharge time series in order to classify the catchments regarding the runoff behavior, check the reasons for different dimensionalities of observed hydrographs and explore the independent information underlying the system dynamics. We assumed that the degrees of anthropogenic pressure (e.g. land use, changing groundwater level and urbanisation) might have an impact on the dimensionality of observed hydrographs. The main steps of this study were that:

1) Observed discharge data (63 small catchments (<500 km^2) in the Federal State of Brandenburg, Northeast Germany) were checked quickly using empirical cumulative distributed function to exclude the high data with insufficient resolution of the sensor. Catchment properties data were extracted from digital maps.
2) The intrinsic dimensionality of the observed hydrographs was assessed by the CD method.
3) The relationships between the CD values, catchment properties and meteorological data were investigated and then linear regression analysis and cross validation were employed to check whether their relationship was significant.
4) The relationships between the CD values, the variance of time series, autocorrelations for different time lags, the slope of power spectrum, the Hurst coefficient, the variance of PCA scores and the PCA loadings were investigated in order to check for redundancy of the CD approach with other standard time series analysis approaches.

4.1 Study area and data processing

Time series of observed discharge from 63 small catchments (<500 km^2) located in the Federal State of Brandenburg, Northeast Germany, were chosen. The Federal State of Brandenburg covers an area of 29500 km^2 and has a population of 2.5 million. It is characterised by a glacially and post-glacially formed landscape which is dominated by Pleistocene sandy and loamy sediments, while Holocene organic sediments are limited to riversides. In this region, forest area contributes to 35% of the total area where Scots Pine represents the dominated tree species. Agricultural land is another main land use type with 34% cropland and 9% pasture. The region receives an amount of precipitation with 610 mm per year on average and annual mean temperature from 1951 to 2000 was between 7.8° and 9.5℃. The landscape is characterized by a large number of lakes, kettle holes and wetlands. Evapotranspiration from these surface water bodies is highly over-proportional. Due to high climatic water demand, the evapotranspiration here is approximately 510 mm per year, only leaving 100 mm per year as runoff (Lischeid and Natkhin, 2011). The runoff exhibits substantial spatial variability, depending on local meteorological conditions. About 80 out of 100 mm runoff per year occurs as groundwater flow, whereas surface runoff plays only a minor role, accounting for less than 5% of total runoff (Merz and Pekdeger, 2011).

The region exhibits a wide array of anthropogenic impacts on the freshwater systems. These include weirs, dams and locks, flood protection by levees, navigation, loading by waste water and industrial pollutants (Nützmann et al., 2011). These effects result in extensive use and alteration of regional freshwater quantity and quality. Drainage and sub-irrigation systems are common in peat-lands. The groundwater level is decreasing due to the combined effects of land cover change (e.g. increasing area and aging of pine forests), drainage measures for agricultural purposes (e.g. melioration of the nearby Nuthe-Nieplitz river valley) and climate change (e.g. warming) (Germer et al., 2011).

Observed daily discharge data covering a 15 year period (from 1991 to 2006) of 63 small catchments were used to for the CD analysis. These data were provided by the State

Office of Environment, Health and Consumer Protection of the Federal State of Brandenburg, Germany. The data had been quality checked by the water resources authority. However, visual inspection revealed a cascading structure in some data sets that were due to limited resolution of the measurements. These artifacts are very likely to have an impact on the CD assessment. Thus data were checked by studying the empirical cumulative distribution functions (ECDF) of the data sets. The ECDF describes the probability that a real-valued random variable X with a given probability distribution will be found at a value less than or equal to X. A stepwise increase of the ECDF gives evidence for insufficient resolution of the data and changing rating curve from water level to the discharge (Figure 4.1). After the quick check of data using ECDF figures, 14 catchments with distinct steps in the ECDF were discarded, leaving 49 catchments for the CD analysis. However, only 35 catchments got reliable CD values. Their corresponding locations are shown in Figure 4.2 and Table 4.2.

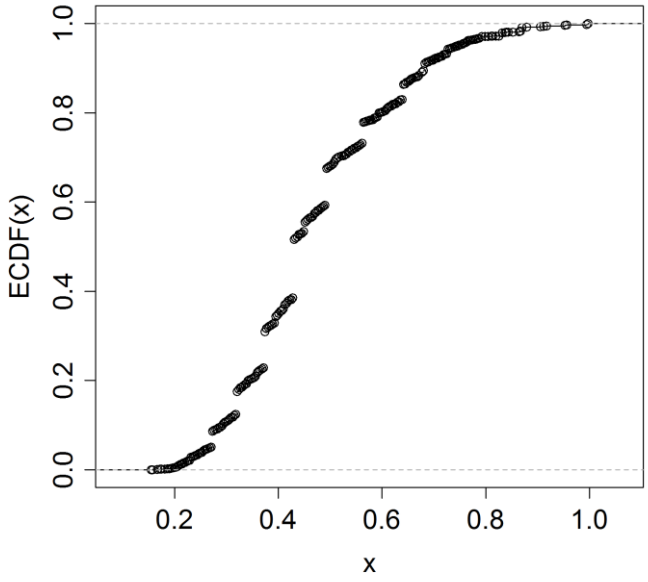

Figure 4.1 Empirical cumulative distribution function (ECDF) of catchment with strong stepwise.

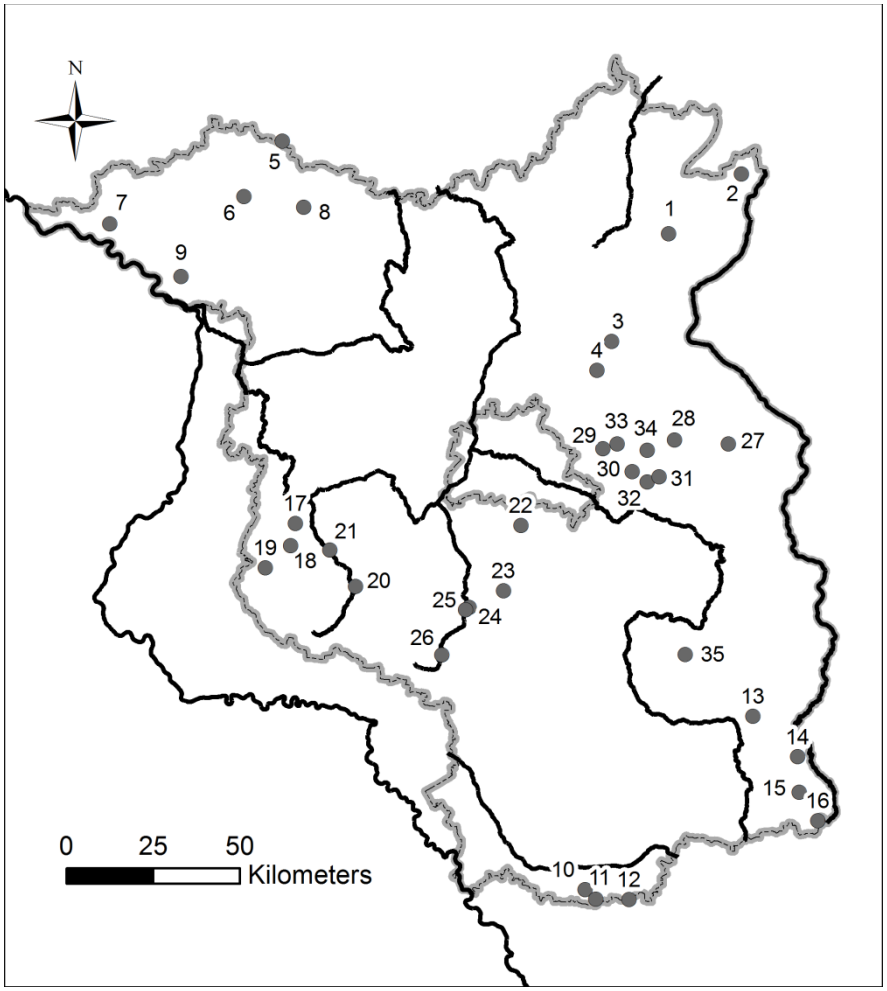

Figure 4.2 Locations of 35 catchments in Brandenburg, Germany, with federal state borders (grey bold dotted lines), rivers (black solid lines), gauges (points), and gauge codes (numbers).

Differences between observed discharge time series are likely linked to climate forcing (through precipitations, radiation, etc.) as well as to rainfall-runoff transformation mechanisms which are specific for each catchment (Porporato and Ridolfi, 1997), such as catchment geometrics, hydraulics, hydrological characteristics of watershed, topology, river network, the properties of aquifers, storage capacities including lakes and anthropogenic impacts. The investigated indicators are listed in Table 4.1.

Table 4.1 Investigated indicators of catchment properties.

Indicators	
Catchment longitude and latitude	
Catchment area	
MMP= max discharge over median discharge	
Catchment altitude	min, max, mean, range
Groundwater level	min, max, mean, range
Groundwater depth(differences between groundwater level and altitude)	min, max, mean, range
Annual average precipitation	min, max, mean, range, summer range, winter range
Annual average potential evapotranspiration	min, max, mean, range, summer range
Land use types	forest area (or percentage) agricultural land area (or percentage) Settlement area (or percentage) water surface area (or percentage) wetland area (or percentage) peat land area (or percentage)
Population and population density	min, max, mean, range
Waste water discharge	
River length and mean stream network Density including artificial river network	

Catchment properties were derived from topographical, hydrogeological, hydrometeorological and land use data sources. Mean catchment values were calculated after intersection of these data with surface catchment boundaries (Thomas et al., 2012). We used (i) the hydrogeological map of Brandenburg Hyka50, 1:50,000 (State Office for Mining, 2012a) to set up a digital altitude model of the groundwater table and digital elevation model of Germany to calculate mean depth to groundwater, (ii) the Hydrological Atlas of Germany 1961-1991 (Federal Ministry for the Environment, 2003) to calculate mean potential evapotranspiration, (iii) the CORINE land cover data (Bossard et al., 2000) for calculating percentage coverage of forest and agricultural area, (iv) the surface catchment area (State Office for Mining, 2012b) to calculate mean surface catchment area, (v) river length and density was taken from the GwNet 25 BB database and waste water discharge data was from the municipal wastewater treatment database

(State Office for Mining, 2012a, 2012b), (vi) Population density was calculated on the basis of the VG250 database (Central Basic Geodata Service for Germany, 2012).

4.2 Results

4.2.1 Correlation Dimension results

Out of 63 hydrographs provided for this study, 14 catchments were discarded during screening procedure using the ECDF figure due to insufficient measurements, and another 14 catchments were deleted due to the absence of a saturation range in the $d(m)$ plot. Thus, CD values were calculated for 35 catchments (Table 4.2). The CD values ranged from 2.2 to 5.8. The frequency distribution of CD values was not normal distributed (Figure 4.3). The majority of CD values ranged from 3 to 6 and only one catchment showed a smaller CD value of 2.2. The higher CD values, the larger dimensions of the space required to completely unfold the attractor, and the more complex the underlying dynamical system appears to be. The CD values did not show a pronounced spatial pattern (Figure 4.4). They tended to be slightly higher in the western part, ranging from 4.95 to 5.8.

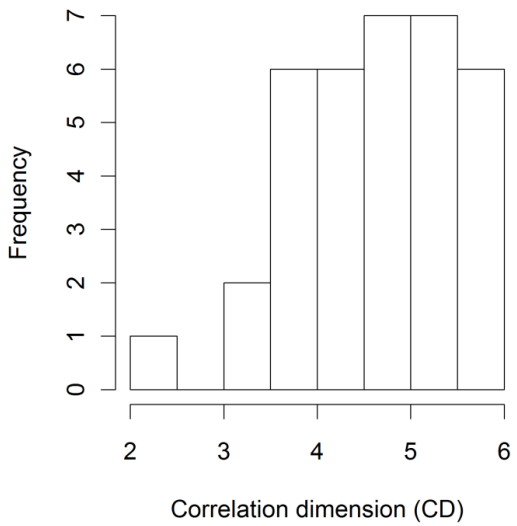

Figure 4.3 The histogram of the CD values frequency distribution.

Table 4.2 The CD values of 35 catchments in Brandenburg

Code	Location	Name	Area (km^2)	CD
1	Northeast	Schönermark	333.5	5.65
2		Tantow	147.0	3.95
3		Eberswalde	125.0	5.4
4		**Grüntal**	**40.2**	**2.2**
5	Northwest	Freyenstein	39.4	4.2
6		Pritzwalk	75.9	3.75
7		Gadow	466.8	4.5
8		Wittstock	73.8	5.3
9		Bad Wilsnack	289.9	3.1
10	Mid-south	Lindenau	258.0	4.6
11		Ortrand	245.0	4.33
12		Lipsa	152.7	5.6
13		Peitz	197.1	5.35
14	Southeast	Mulknitz	120.9	3.8
15		Jocksdorf	28.4	3.9
16		Zschorno	76.3	4.65
17	West	Görisgraben	344.3	4.95
18		Wenzlow	93.7	5.25
19		Birkenreismühle	95.9	5.64
20		Trebitz	226.7	5.78
21		**Golzow**	**416.3**	**5.8**
22	Middle	Blankenfelde	23.3	4.77
23		Mellensee	64.4	4.45
24		Woltersdorf II	209.9	5.13
25		Woltersdorf I	362.9	5.35
26		Jüterbog-Bürgermühle	141.4	4.5
27	East	Gusow	174.7	4
28		Buckow	122.3	4.25
29		Fredersdorf	116.6	5.2
30		Eggersdorf 2	80.7	4.75
31		Kienbaum Straßenbrücke	140.4	4.65
32		Lichtenow	83.1	4.64
33		Garzau	30.9	3.25
34		Kienbaum Neue Mühle UP	16.1	5.58
35		Ressen Stau UP	72.4	3.65

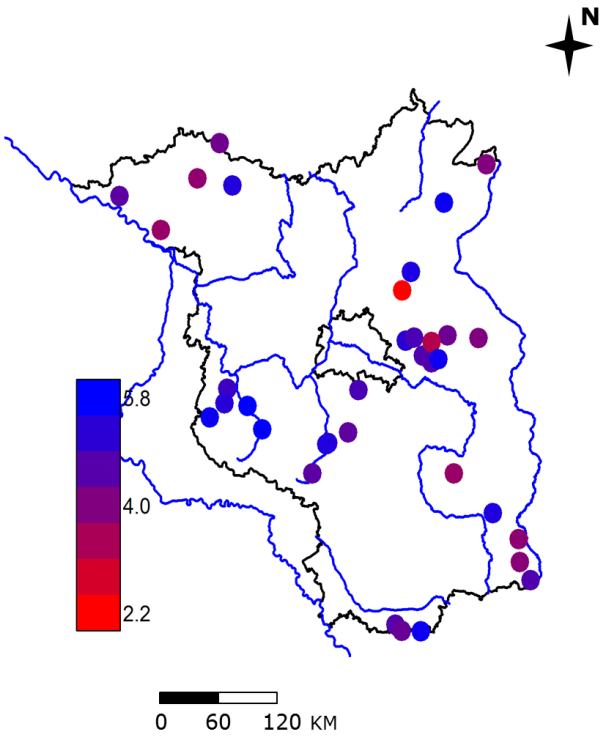

Figure 4.4 Distribution map of CD values (color bar) corresponding to 35 catchments, with federal state borders (dashed lines), rivers (solid lines) and gauges (points).

4.2.2 Relationship between Correlation Dimension values and catchment properties

This section aims at relating the CD values of 35 catchments in Brandenburg State both to catchment properties and to meteorological data. To test for significant correlations, the t-test was used and a cut-off 0.05 was selected as the significance level. The significant correlations between the indicators and the CD values are summarized in Table 4.3. A ranked visualization of all the investigated indicators correlations with CD values is given in Figure 4.5. The obtained results reveal:

1) Only five indicators, including forest area, summer average potential evapotranspiration range, groundwater depth range, annual average potential evapotranspiration range and maximum groundwater depth, show significant correlations

with CD values (i.e. $P < 0.01$). These five indicators show significant colinearity (i.e. the Kendall correlation coefficient between "max groundwater depth" and "groundwater depth range" is 0.89 and the coefficient between "summer average potential evapotranspiration range" and "annual average potential evapotranspiration range" is 0.81).

2) Forest area exhibited the strongest correlation (i.e. 0.4). In contrast, mean, minimum and maximum values of any other indicators do not relate to CD values.

3) Forest area percentage is positively correlated with CD values, whereas agricultural land area percentage is negatively correlated.

4) The CD values exhibit no correlation with catchment size.

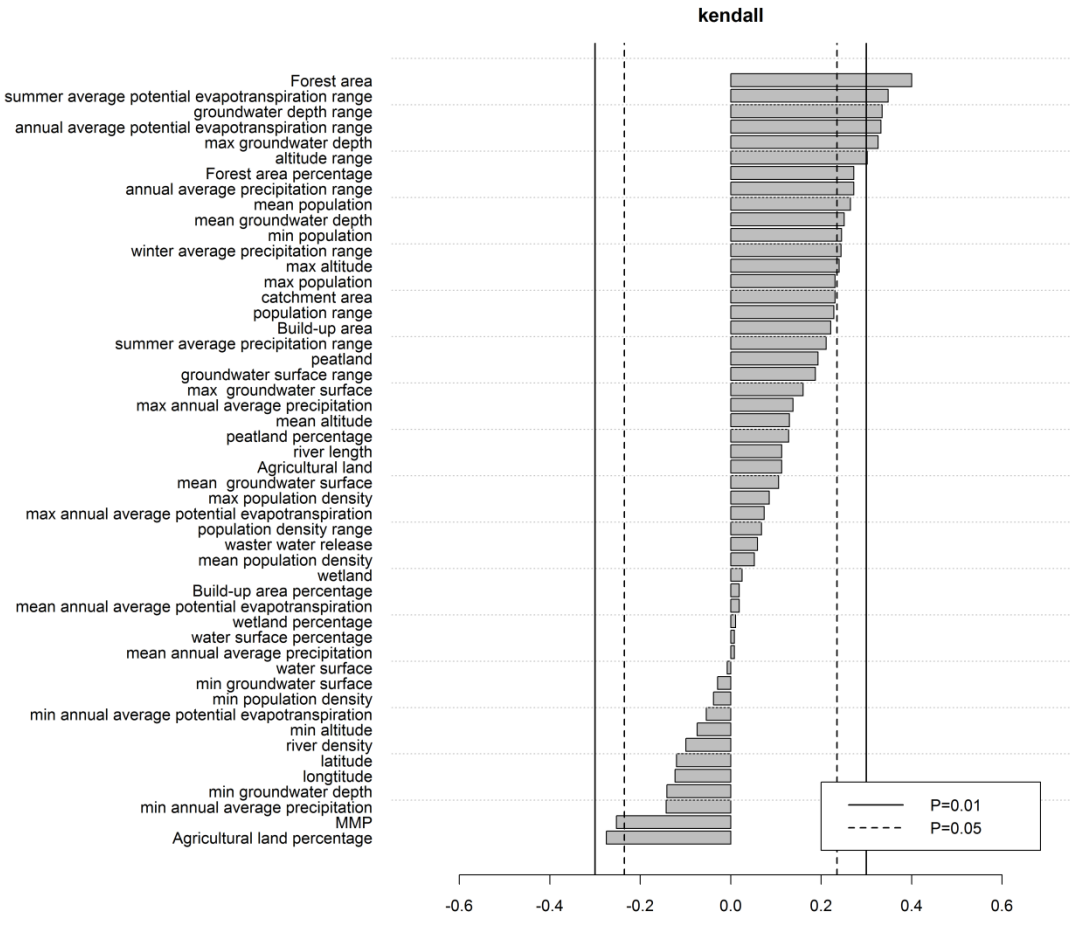

Figure 4.5 Ranked visualization of Kendall correlations between CD values and investigated indicators.

Table 4.3 Kendall correlations between CD values and indicators of catchment properties and meteorological data, whose P values are less than 0.05.

Code	Indicators	Kendall correlation coefficient	P value
1	Forest area	0.4	0.001
2	summer average Ep range	0.348	0.007
3	groundwater depth range	0.335	0.005
4	annual average Ep range	0.332	0.008
5	max groundwater depth	0.326	0.007
6	altitude range	0.302	0.011
7	annual average precipitation range	0.272	0.024
8	Forest area percentage	0.272	0.023
9	mean population	0.265	0.027
10	mean groundwater depth	0.251	0.035
11	min population	0.245	0.041
12	winter average precipitation range	0.244	0.042
13	max altitude	0.24	0.045
14	MMP(max discharge/ median discharge)	-0.253	0.034
15	Agricultural land area percentage	-0.275	0.021

4.2.3 Linear regression analysis and cross validation

The simple linear regression analysis was introduced to model the relationship between CD values and investigated five indicators whose P values are less than 0.01 in order to quantify the strength of these relationships. Then the cross validation, which uses the "training set" to set up the model and the "test set" to assess the model performance, was employed to find out whether the significant relationship was caused by single extremes or not. The results indicate that all the five linear regression equations displayed significant relationships and no relationship was depended on one value. Therefore, these five linear regression models were confirmed to be significant. However, they cannot be used for prediction due to the high residual errors.

In addition, since the CD values are affected by several factors' interactions, multiple linear regression analysis was made to assess which combination of factors can explore more information about coordinate Y than a single factor. In our case, the indicators whose P value during correlation analysis was less than 0.05 (Table 4.3) were

investigated. The multiple regression analysis regards the influence of every factor to be included in a model, but also considers the relationship between factors and eliminates those factors which contribute no new information to the model (Draper and Smith, 1981; Gomez-Plaza et al., 2001). Significance test was provided by F-test and t-test, with a significant level of 0.001 and 0.05 respectively. None of the tested regression models exhibited P values less than 0.05. Therefore, no combination of factors outperformed any univariate regression model.

4.2.4 Relating Correlation Dimension values to standard time series analysis methods

In this section, the CD results of observed discharge were compared to other time series analysis approaches in order to check to what degree the CD method yields additional information. Given that the CD frequency distribution does not appear to be normal distributed (Figure 4.3), the Kendall correlation method was used. Correlations were tested for significance using the t-test. The results are listed at Table 4.4.

Table 4.4 Kendall correlations and corresponding P values between CD values and standard time series analysis.

Method	Indicator	Kendall correlation coefficient	P value
Variance		-0.086	0.477
Autocorrelation	Time lag =1	-0.12	0.306
	Time lag =5	-0.12	0.306
	Time lag =10	-0.13	0.268
	Time lag =30	-0.13	0.293
Power Spectrum	Frequency > (1/300)	-0.04	0.755
Hurst Exponent		-0.24	0.047
PCA scores	Variance of scores	-0.03	0.798
PCA loadings	Component 1	0.147	0.222
	Component 2	0.126	0.293
	Component 3	0.13	0.28
	Component 4	0.295	0.013
	Component 5	-0.052	0.67

For autocorrelation analysis, Kendall correlation coefficients were around -0.12 and P values were around 0.3 (Table 4.4). Thus the relationships between CD values and autocorrelation for different time lags were not significant. The CD method provided different information compared to autocorrelation analysis. Besides, there was no significant correlation between CD values and power spectrum analysis considering the small negative Kendall correlation coefficient (-0.04) and high P value (0.755).

All Hurst exponents of the 35 catchments were within the range $(0.5,1]$ and thus showed persistence in the observed discharge series. The Kendall correlation coefficient between CD values and Hurst exponents was -0.24 with a P value of 0.047. That is, the Hurst exponent had a weak but significant relationship with CD values.

The relationships between PCA loadings or scores and CD values were investigated (Table 4.4). Loadings of component 4 had a significant but still weak correlation with CD values (i.e. correlation coefficient of -0.295 and a P value of 0.013). The variance of the original observed discharge series and PCA scores (transformed component discharge series) exhibited no correlations with CD values.

4.3 Discussion

This study showed that none of the indicators of anthropogenic impacts exhibited significant correlation with CD values. It can only be speculated whether the effects of anthropogenic impact were too small compared to the effects of other catchment properties, or the selected indicators were not suitable. Therefore, it is difficult to judge whether strong anthropogenic pressure on hydrological processes could increase the intrinsic dimensionality of observed hydrographs due to the weak correlation coefficients.

However, the CD values of observed discharge time series displayed significant correlations with five indicators: forest area, summer average potential evapotranspiration range, groundwater depth range, annual average potential evapotranspiration range and maximum groundwater depth. Three out of these five indicators relate to the variation of catchment properties and reflect the heterogeneity of the catchment. Thus, the

heterogeneity of the catchment seems to be the main factor explaining high CD values of the observed hydrographs. In addition, forest area percentage of land use is positively correlated with CD values, while agricultural land area percentage is negatively correlated. That means, more forest area induces higher CD values and increases the complexity of system dynamics, whereas the agricultural land leads to the lower CD value and reduces the complexity. It is assumed that the heterogeneity of the catchment hydrological processes maybe increases with increasing percentage of forest area due to larger species diversity in the forest, resulting in higher CD values. On the contrary, agricultural lands mainly cultivate a few species fit to plants seasons and certain regions, leading to the homogeneity of the catchment and lower CD values. Thus, the reason for different dimensionalities of observed hydrographs relates to the heterogeneity of the catchment. The underlying system dynamics of forest area seems to be more complex than agricultural land with respect to the observed runoff behavior.

5 Using the Correlation Dimension analysis to evaluate model performance

The evaluation of hydrological model performance is commonly assessed by comparison of simulated and observed discharge at the catchment outlet (Krause et al., 2005). Various efficiency criteria, such as the Nash-Sutcliffe (NSc) efficiency, coefficient of determination, the absolute and squared errors, have been used to assess the model's ability to reproduce historic catchment behavior (e.g. Yapo et al, 1996, 1998; Madsen 2000). Different efficiency criteria place emphasis on different systematic and dynamic errors between simulations and observations. For example, the Nash-Sutcliffe (NSc) efficiency criterion measures the fraction of the variance of observed flows explained by the model in terms of the relative magnitude of the residual variance ('noise') to the variance of the observed data ('information') (Nash and Sutcliffe, 1970; Yapo et al, 1996). It employs the function of square errors, which lays stress on larger errors while small errors tend to be neglected. Consequently it leads to an overestimation of the model performance during peak flows and an underestimation during the low flows (Krause et al., 2005). It is often argued that a single evaluating criterion would not be adequate, therefore, multiple efficiency measures were proposed and studied (e.g., Madsen, 2000; Yu and Yang, 2000; Cheng et al., 2002, 2006). In some of these studies, the observed discharge time series were partitioned into peak flow, medium flow and low flow and have been assessed with different criteria. Finally all the efficiency criteria were aggregated into one function. In these studies, the multiple efficiency criteria showed advantages compared to a single criterion.

However, no matter using single or multiple efficiency criteria, many different model structures and many different parameters sets within a chosen model structure often produce a range of equally acceptable solutions. Obviously a global optimum does not exist. These amounts of acceptable "equally good" solutions are called 'non-dominated solutions' (Yapo, et al., 1998; Madsen, 2000) or 'equifinality' (Beven, 1993; Beven and Freer, 2001). Equifinality is a serious obstacle for modelers. The common traditional efficiency criteria are blind to detect the intrinsic properties of the underlying system,

failing to solve the equifinality problem. There is an urgent need to develop an efficiency criterion that accounts for the intrinsic properties of the respective dynamic system through assessing the "complexity" of observed and simulated data independent from any predefined assumptions.

This chapter describes the second study of the CD method application. In this study, the CD method was used to reduce the equifinality problem and consequently improve the hydrological prediction accuracy. To the authors' knowledge, this study is the first one to propose an efficiency criterion based on chaos theory that explores the intrinsic system's properties. This new criterion is compared with the commonly used traditional criterion (NSc) for evaluating model performance. The main steps were that:

1) Three hydrological models of different complexity were fitted to observed data from two catchments (the Karthane catchment in Germany and the Shaowu catchment in China) using a Monte Carlo approach, and performance of the optimal simulations was evaluated based on the NSc value.
2) The intrinsic dimensionality of observed and simulated discharge time series was determined using the CD method, and the relationship between NSc and CD values was investigated, achieving the objective 3 (Page 5).
3) The intrinsic dimensionality of precipitation and evapotranspiration time series, which characterize the input into the system, was assessed and compared to that of the runoff data as the system's output, achieving the objective 4 (Page 5).
4) The CD values were compared to the results of autocorrelation and power spectrum analysis to check for redundancies between these measures.

5.1 Study area and data

Data sets from two catchments (one in Germany and one in China) were used in this study. Their corresponding characteristics are described as follows.

5.1.1 The Karthane catchment, Germany

The Karthane catchment is a small catchment ($289.9\,km^2$) situated in the glacially formed

northeast German lowland, with its gauge in Bad Wilsnack (Table 4.2). It is located in the Northwest of the Federal State of Brandenburg (Figure 5.1). The landscape is dominated by Pleistocene sandy and loamy sediments. Agricultural land and forest area are the major land use types, with the percentage of 78.7% and 19.1% respectively. Annual mean precipitation is around 690 mm per year, while evapotranspiration amounts to 550 mm per year, leaving 140 mm as runoff. The mean temperature is approximately 9℃. Ten years daily monitoring data including precipitation, discharge and temperature data are used in this study (Figure 5.2). Potential evapotranspiration is assessed by the mean values of the respective region in the digital hydrological atlas of Germany from 1961 to 1991 (Federal Ministry for the Environment, 2003).

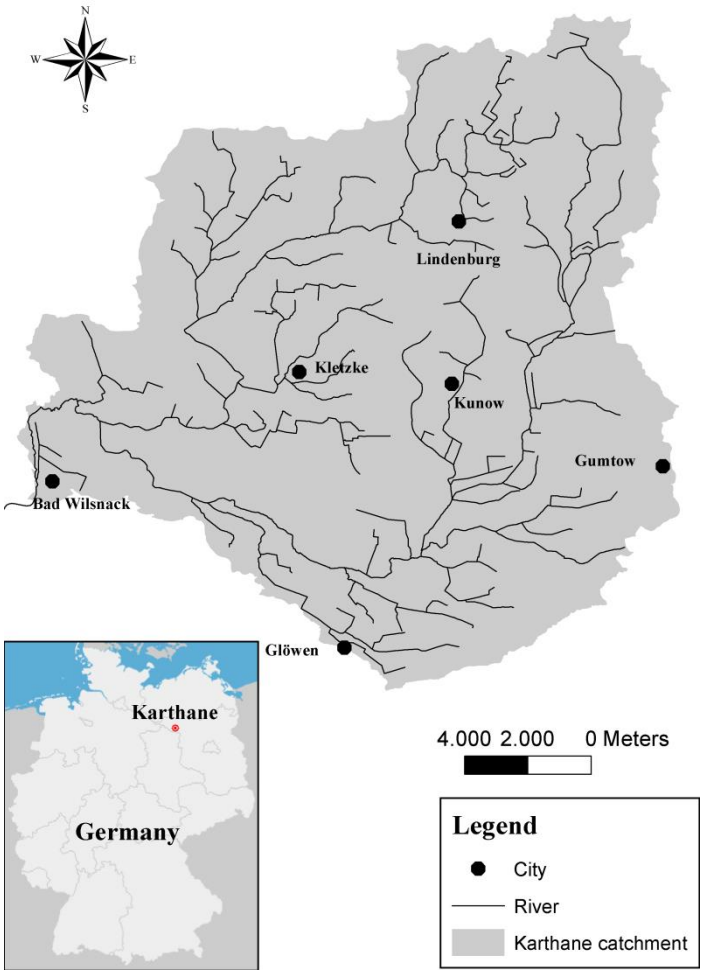

Figure 5.1 Map of the Karthane catchment, Germany.

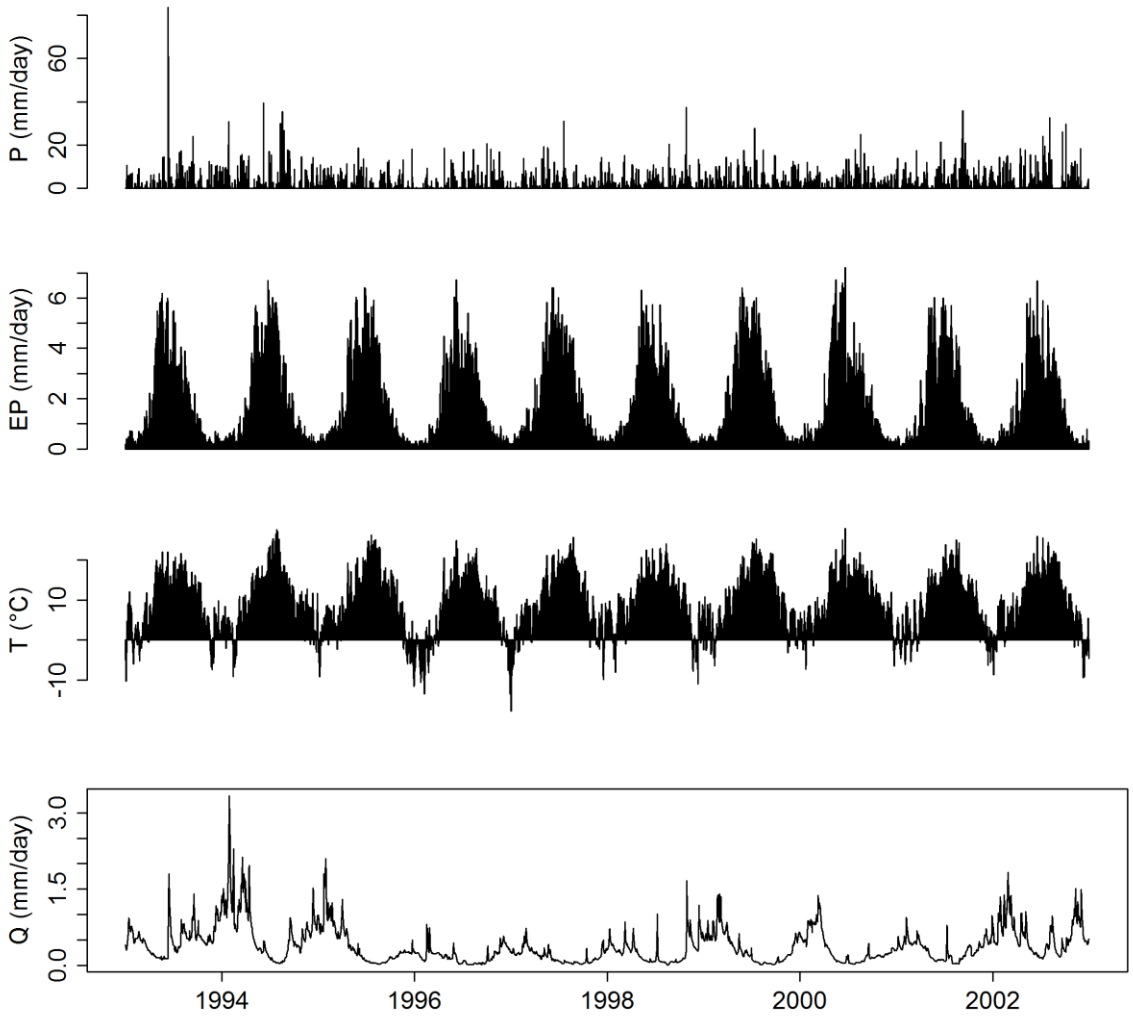

Figure 5.2 Precipitation (P), potential evapotranspiration (Ep), temperature (T) and observed discharge (Q) time series of the Karthane catchment from the year of 1993 to 2002.

5.1.2 The Shaowu catchment, China

The Shaowu catchment with a drainage area of 2745 km^2 locates at the upstream of the Minjiang River in China. It is a sub-basin in the Northwest of the Minjiang basin. The landscape here is mainly medium or low mountains with granite generated from the late Jurassic to early Cretaceous period. The groundwater is mainly bedrock fissure water and has good circulation and recharge condition. Annual mean temperature is 18℃. Annual mean rainfall is around 1940 mm per year, while the mean pan evaporation is

approximately 750 mm per year. There are 9 rainfall gauges (at Shaowu, Gaoyang, Jinkeng, Guangze, Gaojia, Chafu, Zhima, Qiaowan and Siqian), 1 evaporation monitoring station (at Guangze) and 1 discharge measuring station (at Shaowu) in this catchment (Figure 5.3). Ten years daily observed data of rainfall, pan evaporation, discharge and temperature data were used in this study (Figure 5.4). For rainfall, the arithmetic mean values of 9 rainfall gauges were used. Pan evaporation was measured with an E-601 evaporator. About 68.9% of the annual rainfall occurs between March and June, whereas the maximum evaporation is between May and September.

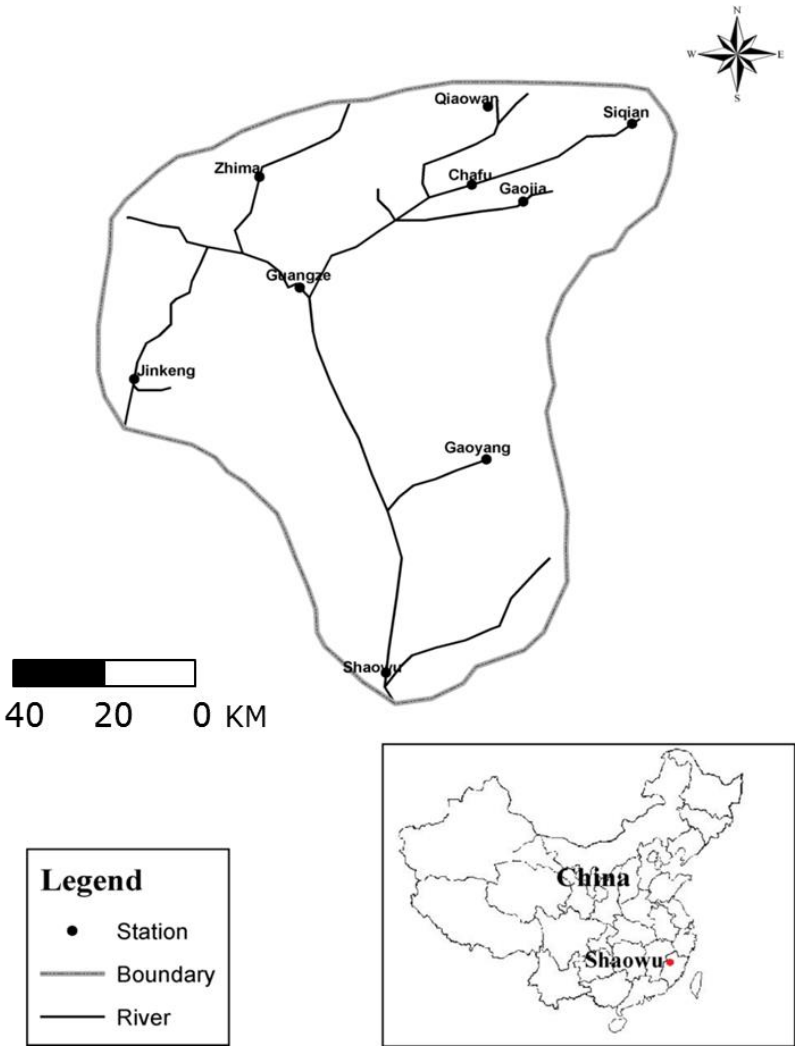

Figure 5.3 Map of the Shaowu catchment, China.

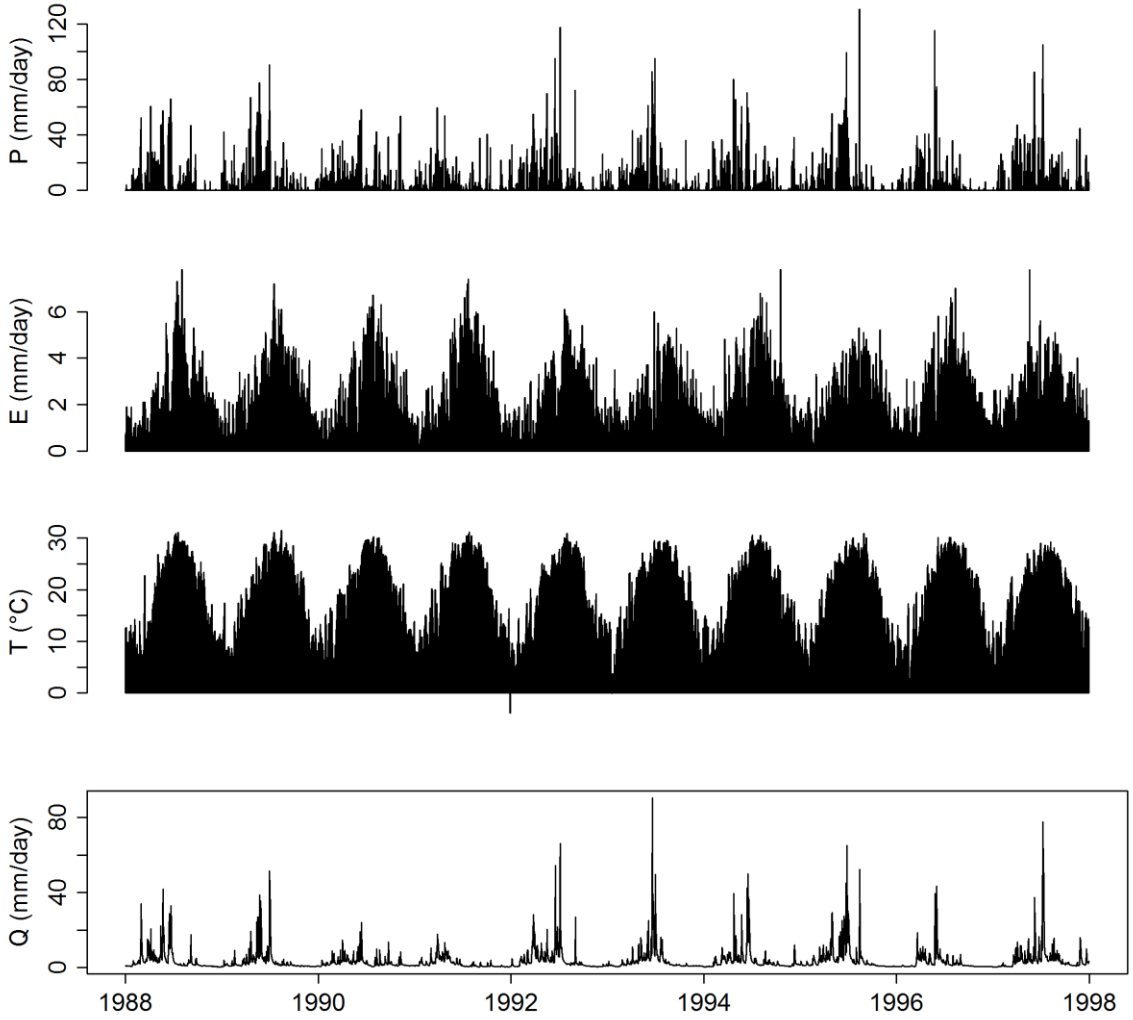

Figure 5.4 Precipitation (P), pan evapotranspiration (Epan), temperature (T) and observed discharge (Q) time series of the Shaowu catchment from the year of 1988 to 1997.

5.2 Results

5.2.1 Model results

For each of the three models, 10000 parameter combinations were generated in a Monte Carlo procedure using random numbers from a uniform distribution within the given ranges of parameter sets (Table 5.1). These ranges were defined by the model constraints and empirical scopes given by the model developers. The model was run with each parameter combination and the NSc value was calculated. Then simulations were ranked

based on the NSc values. As expected the equifinality phenomenon became manifest in the results. In the following we will focus only on the simulations with the best NSc values. For the Karthane catchment only has the potential evapotranspiration data, therefore, the VM model parameter "K" stands for the data conversion coefficient instead of its original meaning and the given range of "K" was selected as 0.1 to 2.

Table 5.1 Parameters range used in the Monte Carlo procedure. The bold parameters are sensitive parameters.

abc model		VM model		HBV model	
Parameters	Range	Parameters	Range	Parameters	Range
a	[0,1]	**KC**	[0.1, 1.1] (Karthane)	TT	[-2, 0]
b	[0,1]		[0.1, 2] (Shaowu)	**SFCF**	[0.2, 1]
c	[0,1]	UM	[5, 20]	**CFMAX**	[1, 4]
		LM	[60, 90]	CFR	0.05
		C	[0.1, 0.2]	CWH	0.1
		WM	[120, 250]	**FC**	[200, 850]
		KF	[0.1, 20]	**LP**	[0.2, 1]
		B	[0.1, 0.4]	**BETA**	[1, 4]
		FC	[1, 80]	**Alpha**	[0, 0.5]
		BF	[1, 1.5]	**K1**	[0.07, 0.2]
		KG	[0.2, 0.4]	K2	[0.005, 0.07]
		CS	[0.1, 0.5]	PERC	[1, 2.5]
		CI	[0.5, 0.9]	**MAXBAS**	[2, 5]
		CG	[0.9, 0.999]		

In this section, we explored whether the equifinality phenomenon relates to the model complexity. The results indicated that all of the studied models exhibited the equifinality phenomenon with different optimal NSc values. For the Karthane catchment, the more complex VM and the HBV model produced acceptable simulations and their optimal NSc values reached up to 0.64, while the simple abc model simulations were not reasonable and the best NSc value was only 0.18. For the Shaowu catchment, the simulations of abc, VM and HBV models obtained very high NSc values, with 0.8, 0.88 and 0.9 respectively.

5.2.2 Correlation Dimension results of observed and simulated discharge series

The CD analyses of observed and simulated discharge time series were made to explore the intrinsic properties of their system dynamics. The CD values of observed discharge time series were 3.1 for the Karthane catchment and 2.55 for the Shaowu catchment (Figure 5.5).

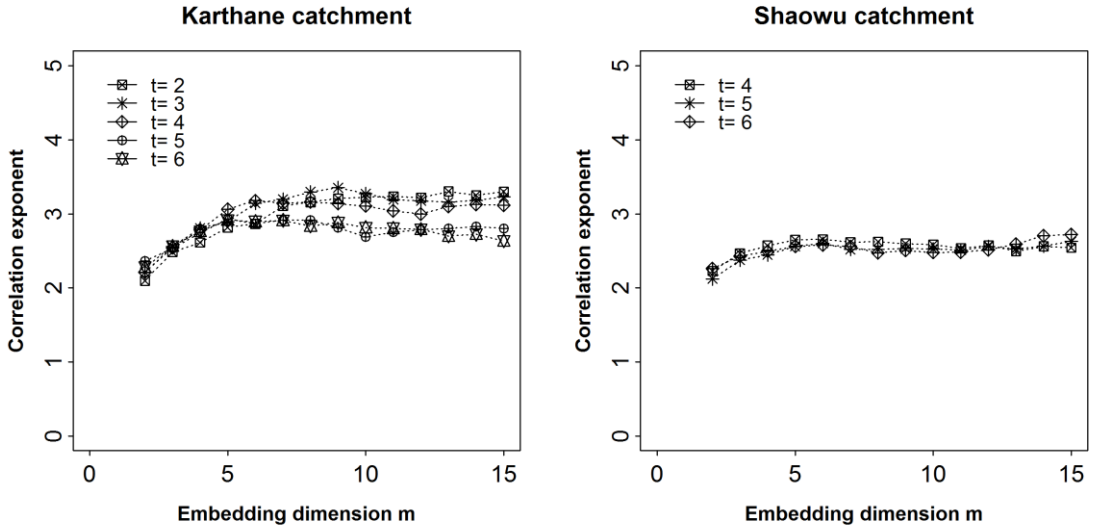

Figure 5.5 The correlation exponent versus embedding dimension of observed discharge time series for the two catchments, where t is the time delay.

For the abc and the VM model, the selected model with the best results were classified into two groups with five simulations each. The first group contains the simulations obtained by very similar parameter combinations, and the second group includes simulations from widely differing parameter values (Table 5.2, 5.3). Thus it can be checked to what extent different parameterizations of the same models fitted to the same hydrograph vary in terms of intrinsic dimensionality. The results revealed that the optimal simulations with very similar parameter values yielded almost similar CD values, whereas those with widely differing parameter values led to totally different CD values. In order to enhance the ability of the CD method to distinguish between the optimal simulations with widely differing parameter values, 20 simulations of the HBV model were investigated via the CD analysis. The CD results covered only a narrow range in the Karthane

catchment, from 2.67 to 2.95 (Table 5.4), and they were very close to the CD value of observed discharge series (Figure 5.6). For the Shaowu catchment, the CD results grouped in two separate ranges (Figure 5.6), that is, between 2 and 2.21 and the other between 3.95 and 4.65 (Table 5.4). Again, different optimal simulations with widely differing parameter values yielded very different CD values. In addition, the correlations between CD values and each parameter in these three models were investigated. The results reveal that there was no correlation between the CD values and model parameters and no special pattern was existed.

Table 5.2 The NSc and the CD values of simulated hydrographs for different parameter values of the abc model.

Study Area	Classification	Num	a	b	c	NSc	CD
Karthane catchment	Group 1	1	0.162	0.833	0.054	0.17	3.5
		2	0.162	0.833	0.059	0.17	3.5
		3	0.162	0.838	0.059	0.16	3.65
		4	0.16	0.833	0.059	0.17	3.5
		5	0.165	0.833	0.054	0.18	3.55
	Group 2	1	0.216	0.774	0.088	0.16	3.75
		2	0.167	0.819	0.093	0.12	3.78
		3	0.172	0.828	0.103	0.12	4
		4	0.152	0.823	0.044	0.11	5.65
		5	0.162	0.819	0.088	0.11	4.18
Shaowu catchment	Group 1	1	0.652	0.284	0.407	0.80	6
		2	0.652	0.289	0.407	0.80	6.03
		3	0.657	0.289	0.407	0.80	6.02
		4	0.657	0.294	0.412	0.80	6.1
		5	0.647	0.28	0.397	0.80	5.85
	Group 2	1	0.652	0.27	0.436	0.80	6.2
		2	0.613	0.309	0.51	0.80	6.58
		3	0.667	0.28	0.451	0.81	6.35
		4	0.676	0.294	0.466	0.80	6.4
		5	0.618	0.333	0.466	0.80	6.13

Table 5.3 The NSc and the CD values of simulated hydrographs for different parameter values of the VM model. The bold columns are sensitive parameters.

Study Area	Classification	Num	**KC**	WUM	WLM	C	**WM**	**KF**	B	**FC**	BF	**KG**	**CS**	**CI**	**CG**	NSc	CD	CD_{Evap}
Karthane catchment	Group 1	1	**1.425**	14.700	89.850	0.159	**215.550**	**13.350**	0.382	**74.400**	1.333	**0.346**	**0.190**	**0.900**	**0.994**	0.64	2.65	
		2	**1.425**	14.750	89.850	0.159	**215.550**	**13.350**	0.384	**74.400**	1.335	**0.346**	**0.192**	**0.900**	**0.994**	0.64	2.60	
		3	**1.420**	14.650	89.700	0.158	**214.900**	**13.255**	0.381	**74.050**	1.330	**0.345**	**0.188**	**0.898**	**0.994**	0.63	2.62	
		4	**1.415**	14.650	89.550	0.158	**214.250**	**13.160**	0.381	**74.050**	1.330	**0.344**	**0.188**	**0.896**	**0.993**	0.62	2.68	
		5	**1.410**	14.600	89.400	0.157	**213.600**	**13.065**	0.379	**73.350**	1.328	**0.343**	**0.186**	**0.894**	**0.993**	0.61	2.69	9.10
	Group 2	1	**1.725**	19.150	60.300	0.131	**242.850**	**16.390**	0.397	**67.750**	1.428	**0.224**	**0.470**	**0.892**	**0.981**	0.62	2.30	
		2	**1.655**	10.750	72.600	0.126	**199.950**	**11.545**	0.213	**52.700**	1.213	**0.398**	**0.370**	**0.854**	**0.999**	0.64	3.35	8.90
		3	**1.55**	10.20	71.70	0.19	**237.00**	**16.96**	0.39	**70.90**	1.17	**0.39**	**0.36**	**0.81**	**0.99**	0.59	2.55	
		4	**1.570**	10.450	72.450	0.189	**240.250**	**17.340**	0.393	**72.650**	1.183	**0.392**	**0.368**	**0.820**	**0.993**	0.60	2.40	
		5	**1.520**	14.850	88.950	0.173	**203.850**	**18.100**	0.343	**45.350**	1.058	**0.343**	**0.314**	**0.822**	**0.999**	0.59	2.10	9.15
Shaowu catchment	Group 1	1	**0.640**	7.025	69.000	0.106	**168.800**	**8.713**	0.238	**26.725**	1.040	**0.220**	**0.448**	**0.830**	**0.970**	0.88	2.63	
		2	**0.635**	6.950	68.850	0.106	**168.400**	**8.614**	0.237	**26.235**	1.038	**0.219**	**0.446**	**0.826**	**0.969**	0.88	2.75	
		3	**0.630**	6.875	68.700	0.105	**168.400**	**8.614**	0.235	**26.235**	1.038	**0.218**	**0.446**	**0.826**	**0.969**	0.88	2.59	10.50
		4	**0.615**	6.650	68.250	0.104	**166.800**	**8.466**	0.231	**25.500**	1.028	**0.215**	**0.438**	**0.820**	**0.967**	0.88	2.77	
		5	**0.660**	7.325	69.600	0.108	**170.400**	**8.862**	0.244	**27.460**	1.050	**0.224**	**0.456**	**0.836**	**0.972**	0.88	2.67	
	Group 2	1	**0.835**	15.800	89.400	0.143	**145.200**	**9.406**	0.399	**20.600**	1.235	**0.317**	**0.468**	**0.760**	**0.972**	0.89	2.69	
		2	**0.900**	8.750	63.150	0.189	**196.400**	**5.347**	0.249	**22.560**	1.225	**0.359**	**0.498**	**0.660**	**0.965**	0.88	2.95	11.10
		3	**0.710**	17.975	85.650	0.130	**135.600**	**8.218**	0.361	**27.705**	1.175	**0.292**	**0.418**	**0.710**	**0.960**	0.88	2.77	10.80
		4	**0.780**	6.050	71.850	0.104	**187.600**	**3.961**	0.357	**25.010**	1.413	**0.392**	**0.498**	**0.748**	**0.973**	0.88	2.90	
		5	**0.900**	18.950	63.900	0.129	**173.200**	**2.476**	0.265	**40.690**	1.398	**0.296**	**0.402**	**0.660**	**0.973**	0.88	2.85	

Table 5.4 The NSc and the CD values of simulated hydrographs for different parameter values of the HBV model. The bold columns are sensitive parameters.

Study Area	Num	TT	**SFCF**	**CFMAX**	**FC**	**LP**	**BETA**	**Alpha**	**K1**	K2	PERC	**MAXBAS**	NSc	CD
Karthane catchment	1	-0.398	**0.9**	**2.499**	**639**	**0.239**	**1.437**	**0.137**	**0.099**	0.07	2.08	**3.036**	0.636	2.82
	2	-1.232	**0.96**	**1.287**	**558**	**0.218**	**1.418**	**0.102**	**0.088**	0.06	2.057	**3.283**	0.636	2.68
	3	-1.648	**0.99**	**1.741**	**574**	**0.265**	**1.556**	**0.166**	**0.057**	0.06	1.23	**3.357**	0.636	2.67
	4	-1.182	**0.97**	**1.496**	**616**	**0.221**	**1.448**	**0.106**	**0.13**	0.08	1.941	**3.382**	0.637	2.76
	5	-1.324	**0.9**	**2.628**	**580**	**0.203**	**1.514**	**0.252**	**0.061**	0.08	1.517	**2.967**	0.636	2.78
	6	-0.65	**0.87**	**1.44**	**589**	**0.216**	**1.378**	**0.202**	**0.079**	0.07	1.561	**3.23**	0.637	2.78
	7	-1.545	**0.98**	**1.125**	**558**	**0.253**	**1.631**	**0.138**	**0.08**	0.07	1.933	**3.424**	0.637	2.69
	8	-0.896	**0.9**	**1.609**	**621**	**0.213**	**1.383**	**0.17**	**0.094**	0.07	1.944	**3.679**	0.638	2.67
	9	-0.803	**0.96**	**2.246**	**656**	**0.221**	**1.335**	**0.223**	**0.053**	0.07	1.367	**2.87**	0.638	2.77
	10	-1.964	**0.97**	**2.88**	**602**	**0.21**	**1.426**	**0.245**	**0.056**	0.08	1.357	**2.86**	0.641	2.79
	11	-0.535	**0.93**	**3.625**	**680**	**0.233**	**1.402**	**0.111**	**0.085**	0.06	1.169	**3.853**	0.635	2.75
	12	-1.181	**0.98**	**3.084**	**578**	**0.223**	**1.512**	**0.251**	**0.089**	0.07	2.35	**3.787**	0.635	2.68
	13	-1.963	**0.88**	**2.551**	**576**	**0.209**	**1.511**	**0.244**	**0.093**	0.07	1.949	**3.635**	0.634	2.73
	14	-1.443	**0.95**	**2.421**	**571**	**0.237**	**1.447**	**0.151**	**0.106**	0.07	2.007	**3.297**	0.634	2.84
	15	-1.339	**0.99**	**3.233**	**640**	**0.305**	**1.698**	**0.232**	**0.06**	0.08	1.714	**3.883**	0.633	2.88
	16	-0.984	**0.98**	**2.234**	**608**	**0.313**	**1.656**	**0.301**	**0.077**	0.06	2.177	**3.196**	0.632	2.8
	17	-1.16	**0.91**	**1.032**	**676**	**0.459**	**2.276**	**0.228**	**0.059**	0.07	1.989	**2.763**	0.631	2.95
	18	-0.801	**0.98**	**3.697**	**605**	**0.306**	**1.585**	**0.146**	**0.099**	0.06	2.317	**4.15**	0.631	2.76
	19	-1.325	**0.92**	**2.099**	**713**	**0.421**	**1.979**	**0.138**	**0.096**	0.07	2.23	**3.11**	0.631	2.95
	20	-0.545	**0.94**	**3.661**	**605**	**0.309**	**1.644**	**0.188**	**0.101**	0.06	1.819	**3.708**	0.631	2.94

Shaowu catchment	1	-1.169	**0.63**	**3.386**	**426**	**0.973**	**2.903**	**0.414**	**0.085**	0.06	3.408	**2.468**	0.915	2.12
	2	-0.019	**0.43**	**1.647**	**264**	**0.93**	**2.976**	**0.425**	**0.102**	0.03	4.025	**2.792**	0.912	2
	3	-0.196	**0.88**	**3.416**	**494**	**0.927**	**2.028**	**0.482**	**0.079**	0.07	4.691	**2.48**	0.917	3.95
	4	-1.259	**0.22**	**2.18**	**345**	**0.915**	**2.224**	**0.481**	**0.095**	0.02	4.727	**3.027**	0.910	2.1
	5	-0.356	**0.28**	**2.302**	**286**	**0.674**	**1.061**	**0.365**	**0.152**	0.08	3.646	**2.997**	0.911	4.5
	6	-0.868	**0.29**	**1.493**	**282**	**0.7**	**1.152**	**0.406**	**0.1**	0.08	4.609	**2.52**	0.920	4.1
	7	-0.765	**0.78**	**2.921**	**494**	**0.883**	**1.63**	**0.294**	**0.151**	0.04	3.508	**2.53**	0.915	2.15
	8	-1.454	**0.68**	**2.154**	**441**	**0.972**	**2.834**	**0.352**	**0.101**	0.05	3.224	**2.439**	0.913	4.3
	9	-0.504	**0.53**	**3.449**	**262**	**0.869**	**2.264**	**0.368**	**0.104**	0.05	3.057	**2.482**	0.911	2.03
	10	-0.275	**0.34**	**2.813**	**499**	**0.977**	**2.784**	**0.361**	**0.12**	0.07	2.761	**2.71**	0.912	4.6
	11	-1.036	**0.42**	**2.892**	**499**	**0.902**	**1.762**	**0.39**	**0.105**	0.07	4.801	**2.777**	0.916	3.95
	12	-0.827	**0.93**	**3.144**	**547**	**0.902**	**1.58**	**0.39**	**0.108**	0.06	4.381	**2.663**	0.919	4.21
	13	-0.981	**0.41**	**1.638**	**554**	**0.934**	**2.246**	**0.367**	**0.121**	0.03	2.752	**2.721**	0.916	2.05
	14	-1.43	**0.51**	**2.066**	**335**	**0.771**	**1.584**	**0.405**	**0.122**	0.03	4.104	**2.854**	0.917	2.12
	15	-0.231	**0.29**	**2.456**	**484**	**0.649**	**1.434**	**0.383**	**0.118**	0.04	3.231	**2.75**	0.915	2.15
	16	-0.957	**0.21**	**3.596**	**276**	**0.716**	**1.577**	**0.397**	**0.118**	0.03	4.602	**2.67**	0.917	2.08
	17	-0.106	**0.65**	**2.805**	**592**	**0.862**	**2.626**	**0.393**	**0.112**	0.06	3.863	**2.757**	0.918	4.18
	18	-0.418	**0.78**	**3.952**	**339**	**0.868**	**1.54**	**0.463**	**0.093**	0.04	4.995	**2.891**	0.918	2.21
	19	-0.106	**0.93**	**1.354**	**519**	**0.737**	**2.466**	**0.478**	**0.087**	0.07	3.708	**2.833**	0.916	4.25
	20	-1.778	**0.36**	**2.705**	**560**	**0.809**	**1.933**	**0.425**	**0.098**	0.06	2.798	**2.475**	0.915	4.65

In this section, the relationship between NSc and CD values was investigated. For the Karthane catchment, the optimal simulations of the abc model could not be improved any more after 10000 iterations of the Monte Carlo procedure, ending up with low NSc values (around 0.18). The CD values of optimal simulation of the abc model ranged from 3.5 to 5.65. They clearly exceeded that of the observed discharge of 3.1. The CD values seemed to be negatively related to the NSc values. However, a corresponding relationship is less clear for the CD results of different models applied in the Shaowu catchment (Figure 5.6). Discharge simulated by the simple abc models exhibited clearly larger CD values compared to those of the VM and HBV model simulations for each catchment. It seemed that there was a convergence of CD values approaching the intrinsic dimensionality of the observed discharge time series with increasing NSc values. To summarize, simulations with similar NSc values differ considerably with respect to CD values, but the CD converges for the best simulation in terms of NSc towards that of the observed discharge.

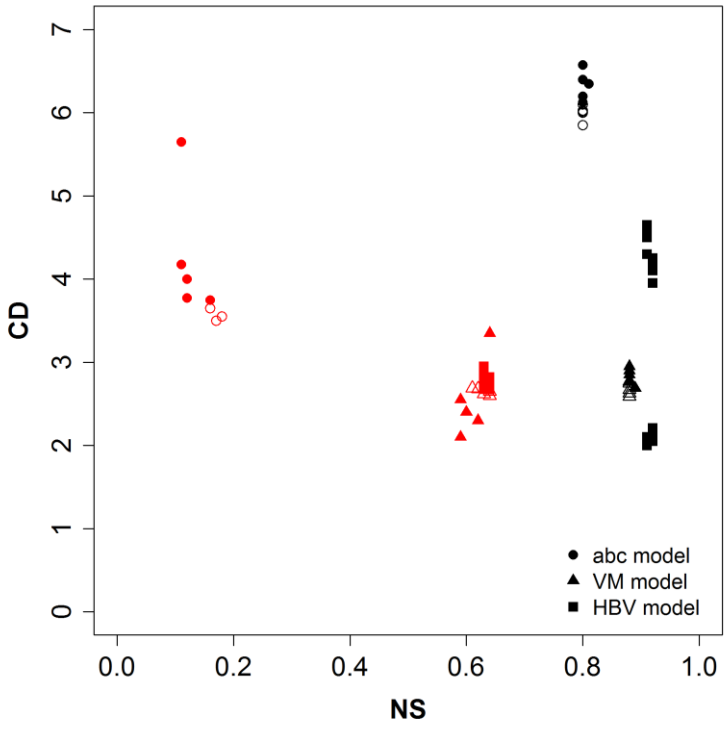

Figure 5.6 The relationship between NSc and CD values of the different models. The red color denotes the Karthane catchment value, while the black color denotes the Shaowu catchment value. The open symbols denote the CD values obtained from similar

parameter values, whereas the solid symbols denote the CD values obtained from widely differing parameter values.

5.2.3 Correlation Dimension results of rainfall and evapotranspiration time series

The correlation dimension values of observed and simulated hydrographs clearly exhibited saturation with increasing embedding dimension. However, no CD value could be obtained from the 10 years daily precipitation data for both catchments. The CD results of observed pan evaporation data in the Shaowu catchment was 10.4, whereas no reliable CD value was obtained from the analysis of potential evapotranspiration data in the Karthane catchment. For the actual evapotranspiration calculated by the VM model, the CD values ranged from 9 to 11, much higher than the values of observed discharge (Table 5.3).

5.2.4 Relating Correlation Dimension values to autocorrelation and power spectrum analysis

The CD method has only recently been applied to characterize hydrological time series, but autocorrelation analysis and power spectrum are much more common. We test to what degree the CD analysis gives additional information that is uncorrelated with those of the common approaches. To that end, Kendall correlations between CD values and autocorrelation were investigated. Correlations were tested for significance using the t-test at a 0.05 significant level. Autocorrelation values were determined for time lags of 1 day, 5 days, 10 days and 30 days (Table 5.5). For the abc model simulations, most Kendall correlation coefficients were negative and ranged from -0.96 to -0.6, excluding the coefficient for the time lags of 30 days in the Karthane catchment (Table 5.5). All of these relationships were statistically significant. For the VM and HBV model, most of the CD values did not exhibit significant relationships with the autocorrelation coefficients (Table 5.5). However, an analysis of a merged data set, that combined the findings for the different models for single catchment or for both catchments, yielded weak but significant negative relationships for nearly all of the considered time lags.

For the power spectrum analysis, the linear slope of the regression line fitted to the spectrum in the double logarithmic plot of spectrum density versus frequency was compared to the CD values, where $1/300$ day was taken as the minimum frequency. The CD values of the abc model simulations indicated significant negative relationships with the slope of the power spectrum for both catchments, but only the Karthane catchment for the HBV model, and none for the VM model. Similar as for the autocorrelation results, merging the results for all the models for single or both catchments resulted in weak but significant relationships (Table 5.5).

Table 5.5 Kendall correlation between CD values, autocorrelations for different time lags and the slope of power spectrum of simulated discharge time series. The bold values stand for the Kendall coefficient whose P values are significant on the 0.05 level.

Study area	Models	Autocorrelation analysis								Power spectrum analysis	
		time lag = 1 day		time lag = 5 day		time lag = 10 day		time lag = 30 day			
		Kendall coefficient	P values	Kendall coefficient	P values	Kendall coefficient	P values	Kendall coefficient	P values	Kendall coefficient	P values
Karthane catchment	abc	**-0.87**	0.001	**-0.83**	0.001	**-0.60**	0.023	-0.41	0.123	**-0.69**	0.008
	VM	0.02	1.000	0.20	0.474	0.20	0.474	0.16	0.592	-0.02	1.000
	HBV	-0.29	0.085	-0.28	0.097	-0.32	0.055	**-0.39**	0.018	**-0.33**	0.047
	Total	-0.19	0.082	**-0.28**	0.011	**-0.35**	0.002	**-0.53**	0.000	**-0.22**	0.046
Shaowu catchment	abc	**-0.73**	0.004	**-0.96**	0.000	**-0.96**	0.000	**-0.91**	0.000	**-0.56**	0.032
	VM	-0.36	0.067	**-0.50**	0.012	**-0.50**	0.012	-0.27	0.181	-0.25	0.215
	HBV	0.07	0.673	0.21	0.205	0.21	0.205	0.29	0.085	0.04	0.820
	Total	-0.18	0.080	**-0.30**	0.004	**-0.38**	0.000	**-0.28**	0.007	**-0.35**	0.001
Both catchments		**-0.22**	0.003	**-0.25**	0.001	**-0.29**	0.000	**-0.35**	0.000	**-0.30**	0.000

5.3 Discussion

The CD values of optimal simulations indicated weak but significant relationships with the results of autocorrelation and power spectrum. This might be partly due to the rather strong relationships found for the abc model but not for the medium complex models. Anyhow, even then only a minor fraction of the variance of the CD values was explained by the more common autocorrelation and power spectrum results, indicating that the CD measure largely gives independent information.

Figure 5.5 shows that the CD value of the Karthane catchment is 3.1, higher than the value of the Shaowu catchment (i.e. 2.55). It means that the dynamic system of the Karthane catchment is more complex than the Shaowu catchment. The higher CD values, the more complex the underlying dynamical system appears to be. The required dominant variables to model the system dynamics are 4 and 3 respectively, and hence the models with less parameter numbers than dominant variables are rejected due to their insufficient model structure.

As conceptual rainfall-runoff models consider mainly the watershed input data and the corresponding output data, the CD analysis of rainfall input and evapotranspiration time series has been performed to check how the intrinsic dimensionality changed in the model from the input to the output. Many previous researches performed a CD analysis of rainfall time series, but they mainly focused on whether the rainfall time series is resulting from stochastic or low-dimensional chaotic processes. The results are controversial. Some research obtained the CD values in the rainfall time series, ranging from 0.5 to 9 (summarized in Table 1 of Sivakumar, 2004a), while the others got no reliable CD values for rainfall data. In this study, no CD value was obtained in the 10 years daily precipitation data for the two catchments. The reasons maybe related to insufficient size (ten years daily data is not long enough) and high noise level of the data sets. These two reasons are the main limitations for applying the CD method. The short data will not only delay the plateau onset, but also make the deviation from the plateau behavior occur at smaller values of the embedding dimension, thus shortening the plateau length from both sides. Moreover, an insufficient data set gives rise to shorter scaling

regions, thus causing the disappearance of scaling regions to occur at smaller values of the embedding dimension (Ding et al., 1993).

From our investigation, the CD values of evapotranspiration time series are much higher than the values of discharge series. That is, the underlying dynamic system of evapotranspiration is more complex than that of discharge. The model seems to be an intrinsic dimensionality reducing filter from input to output. Moreover, the simple abc model cannot reduce the intrinsic dimensionalities of simulations to the necessary dimensionality of observations, always keeping higher CD values (Figure 5.6). The simple model produces more complex discharge time series than the medium complex model (Figure 5.6). Thus instead of raising intrinsic dimensionality, the model reduces the intrinsic dimensionality of simulations when the model complexity increases.

Figure 5.6 shows that the CD values vary between different optimal simulations based on NSc values and make them distinguishable. Thus the CD method can make a better evaluation of the model performance than the other traditional efficiency criteria, like absolute and square errors. However, the CD values also can give similar results if the optimal simulations are obtained from similar parameter combinations. Therefore, it is unnecessary to make the CD analysis on the optimal simulations with similar parameter combinations when evaluating model performance.

It is suggested to combine the CD analysis and NSc efficiency criterion due to the time consuming semi-automatic CD program. Firstly, use NSc as the objective function to find optimal solutions of the model and then apply the CD analysis to distinguish between these simulations. Finally, the comparison of CD values between observed and simulated time series offers the information whether the simulations underestimate or overestimate the intrinsic dimensionalities of observations. If the CD value of a simulation is larger than that of the observation, then this simulation overestimates the intrinsic dimensionalities of observation, and vice versa. If the CD value of a simulation is smaller than that of the observation, then this simulation underestimates the intrinsic dimensionalities of observation. The closer the CD values between the simulation and observation are, the better the simulation is. The proposed evaluating approach employs the traditional NSc efficiency criterion to save time, but also extracts the intrinsic

properties of underlying system dynamics using the CD method. It finally reduces the equifinality problem and might be a step towards improving the model prediction accuracy.

6 Correlation Dimension analysis of groundwater head and lake level

This chapter describes the third CD method application. It is based on the work of Lischeid et al. (2010). The CD method is applied to assess the complexity of groundwater head and lake level and to explore independent information of groundwater dynamics. The main steps were that:

1) The intrinsic dimensionality of groundwater head and lake level was assessed by the CD method and the information provided by the CD method was addressed, achieving the objective 5 (Page 5).
2) The relationship between the CD values and the screening depth of groundwater wells was investigated.

6.1 Study area and data

The study region (Figure 6.1) is located within the biosphere reserve Schorfheide-Chorin, approximately 65 km northeast of Berlin, and 10 km northwest of the town of Angermünde. The region is centered on 53° 03′ 07″ N, 13° 50′ 23″ E. Land use is forest predominantly (52% of the area), although arable land and small settlements cover another 36% of the area. The remaining 12% are covered by lakes and wetlands. Groundwater wells and lakes are located within an area of approximately 40 km^2. The maximum distance between single measurement sites is less than 10 km. In general, groundwater flow direction in the uppermost aquifer is from the West to the East (Lischeid et al., 2010).

Meteorological data were available from Angermünde, about 10 km southeast of the study area. The long-term (1951–2008) mean rainfall at Angermünde amounted to 606 mm per year, and mean air temperature to 8.6 °C (Lischeid et al., 2010). Potential evapotranspiration was about 570 mm per year during that period. In contrast, the year of 2007 was an extraordinary wet year with 747 mm of precipitation, whereas potential evapotranspiration was 547 mm. Consequently, the usual water level depression during

the growing season was virtually absent in some groundwater wells and lakes in 2007 (Lischeid et al., 2010). Pressure transducers were installed at some of sites by the Institute of Landscape Hydrology, ZALF. They yielded hourly water level data. In this study, data of the growing season 2007 (March through August) were used (Figure 6.2). The measurement error usually was less than 1 cm.

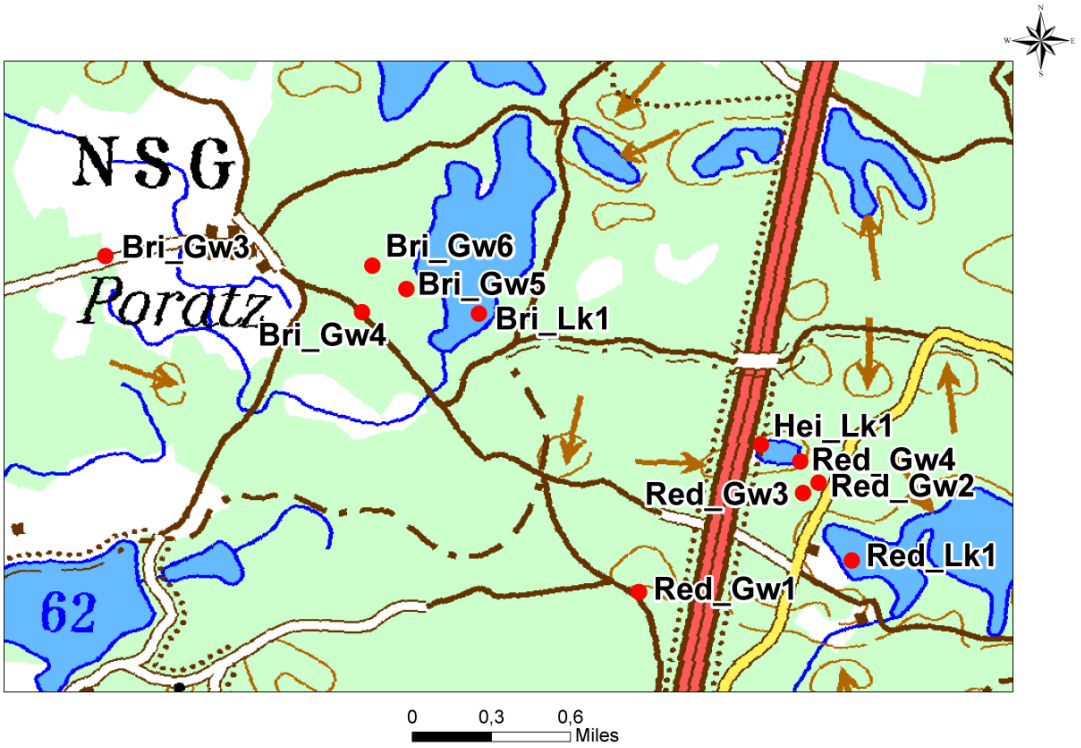

Figure 6.1 Study region map with 7 groundwater wells (red points), 3 lake level recording sites (red points), lakes (blue polygon), forests (green background), cities (white polygon) and roads (red and brown lines). This map is based on the digital topological map (1:100) of Brandenburg State.

Groundwater levels from all the wells in Redernswalder See decreased from March to August, where only Gw1 increased after July. Those in Briesensee exhibited similar changing trends and seasonal characteristics. The short-term fluctuations of the groundwater level data were stronger in Redernswalder See than in Briesensee. The screening intervals of groundwater wells in Redernswalder See range from 14 to 24 m, deeper than those in Briesensee with screening depths between 7 and 15 m (Table 6.1).

All the studied lakes levels showed similar behavior with more fluctuations from May to August. This is probably due to the occurrence of summer rainfall.

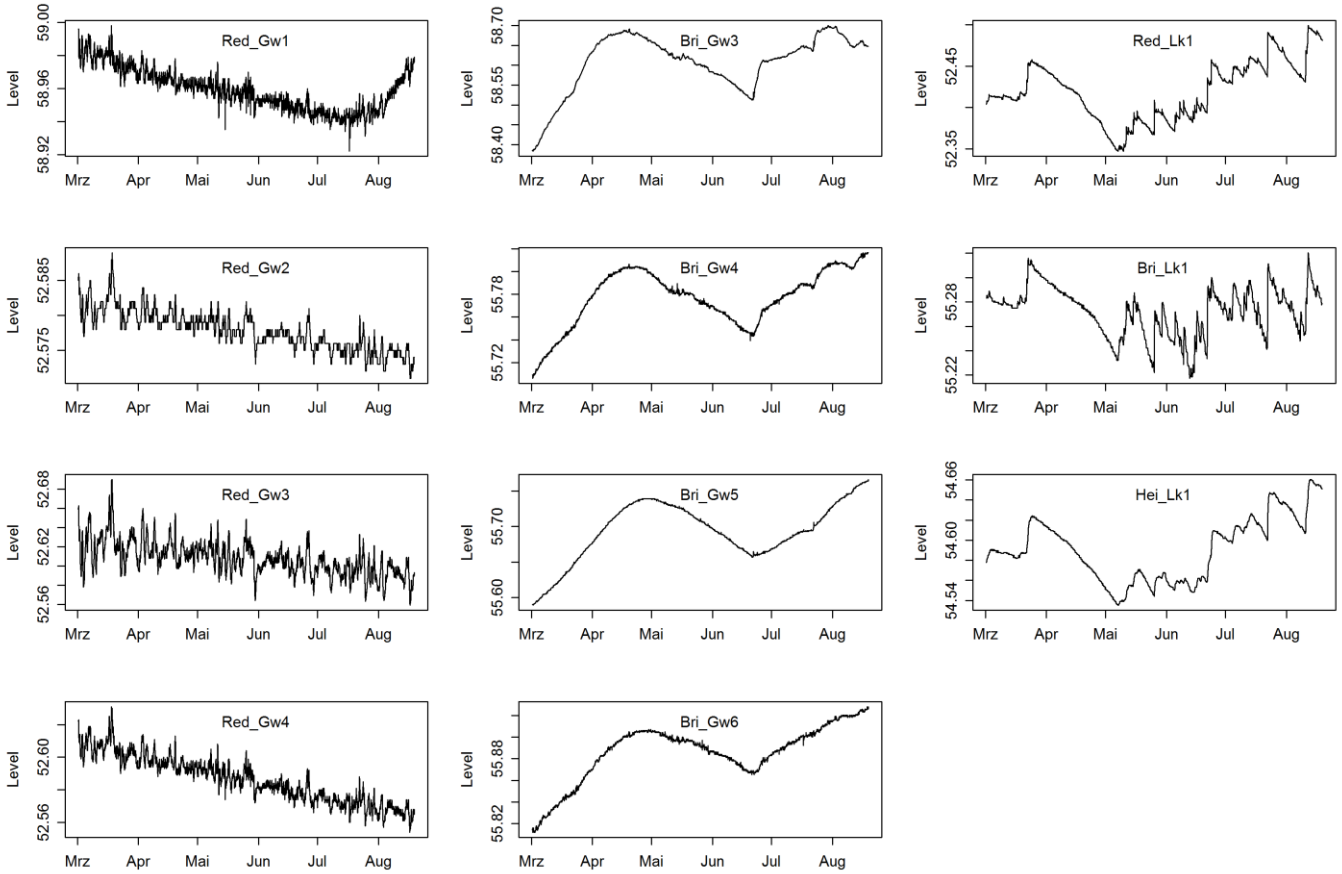

Figure 6.2 Studied groundwater and lake level time series, where “Red”, “Bri” and “Hei” are Redernswalder See, Briesensee and Heilsee, respectively. “Gw” denotes the groundwater level of recording wells, while “Lk” denotes the lake water level from recording sites.

6.2 Results

All of the three lakes levels time series showed almost similar CD values, around 1.5 to 2 (Table 6.1). In contrast, the CD values of 7 groundwater levels time series exhibited obvious spatial pattern (Figure 6.3), with higher CD values ranging from 3.8 to 5 in the region of Redernswalder See and lower values in the region of Briesensee (1.8 for Bri_Gw3 and 0.9 for the other wells). There was no CD value obtained for Red_Gw2. In addition, the CD values displayed uncorrelated to their corresponding screening depth of groundwater wells.

Table 6.1 The CD values of groundwater and lake level data. The screening interval is the distance below upper end of gauge [m].

ID	Site	Data type	Screening interval	CD
Red_Gw1	Redernswalder See	Groundwater level	15-17	5
Red_Gw2	Redernswalder See	Groundwater level	23-24	None
Red_Gw3	Redernswalder See	Groundwater level	23-24	3.825
Red_Gw4	Redernswalder See	Groundwater level	14-15	4.1
Bri_Gw3	Briesensee	Groundwater level	7-9	1.8
Bri_Gw4	Briesensee	Groundwater level	13-15	0.9
Bri_Gw5	Briesensee	Groundwater level	11-13	0.9
Bri_Gw6	Briesensee	Groundwater level	11-13	0.935
Red_Lk1	Redernswalder See	Lake level		1.57
Bri_Lk1	Briesensee	Lake level		2
Hei_Lk1	Heilsee	Lake level		1.63

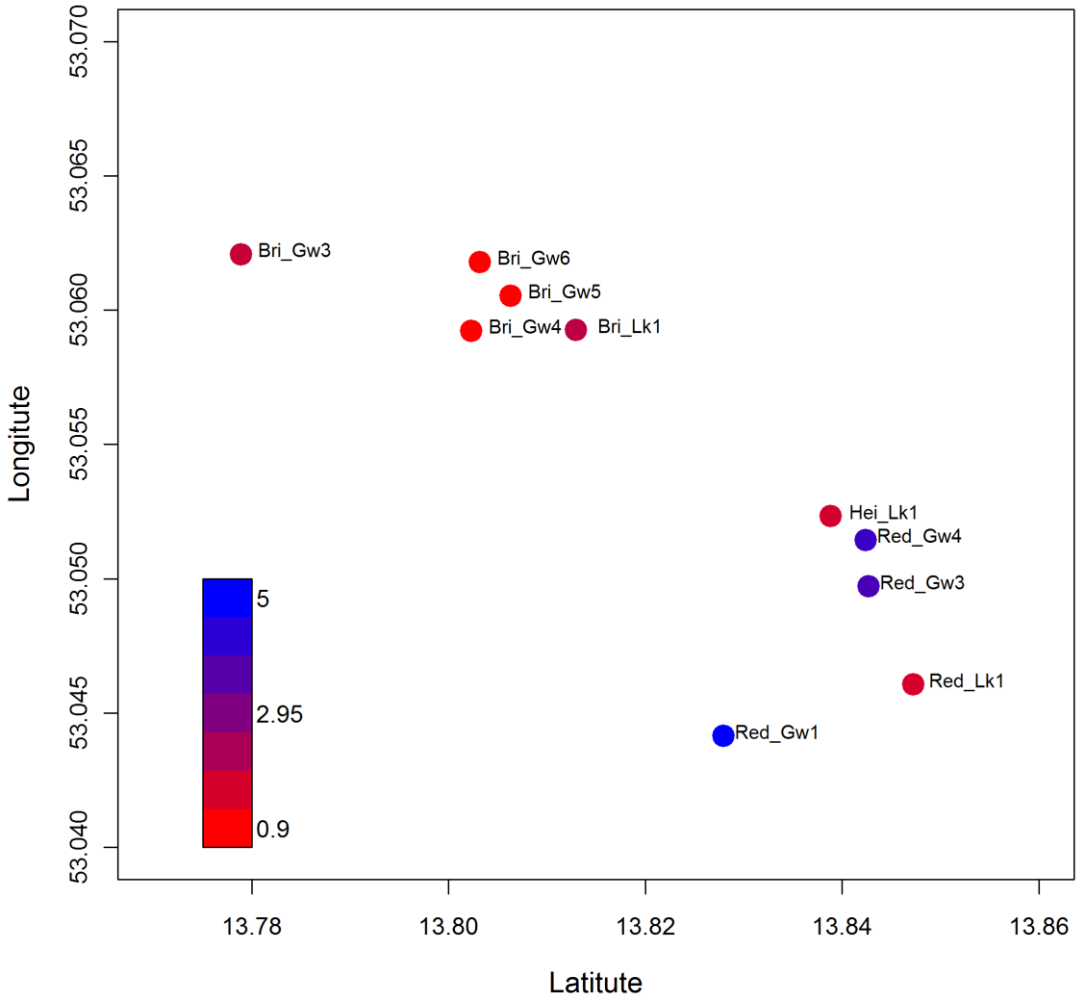

Figure 6.3 Map of the sampling sites (7 groundwater wells and 3 lakes recording sites) with CD values (colour bar).

6.3 Discussion

No CD value could be obtained from Red_Gw2, probably due to insufficient resolution of the measurements and a high noise level of time series, which can clearly be identified in the empirical cumulative distribution function (ECDF) (Figure 6.4). The insufficient resolution of measurements usually display the high stepwise in the time series and clear breakup points in the ECDF (Figure 6.4), which induce the severe problem to get the

reliable CD value. On the other hand, the presence of noise influences the CD estimation primarily from the identification of scaling region. Noise may corrupt the scaling behavior at all length scales, and its effects are significant especially at smaller length scales. Therefore, noise reduction methods are needed. However, in this case, even when noise reduction methods were employed to filter the data of Red_Gw2, we still cannot get the reliable CD values since those methods were not efficient with respect to the high noise level.

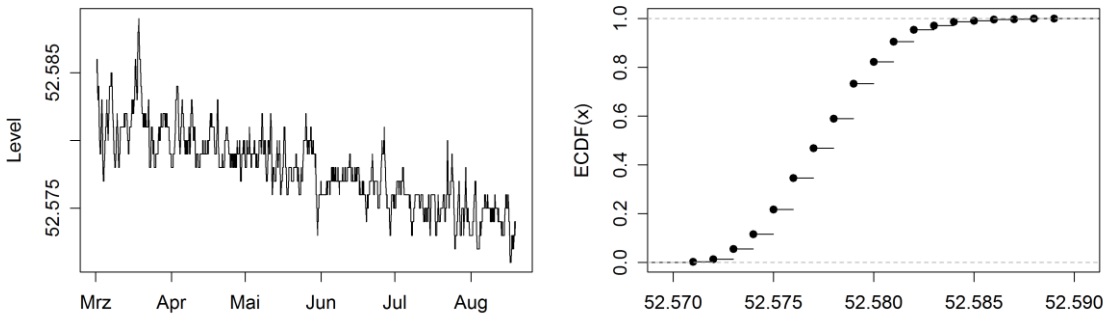

Figure 6.4 Left: Red_Gw2 groundwater level time series; Right: Its empirical cumulative distributed function (ECDF)

The correlation exponents of Bri_Gw3 decreases rather markedly with increasing embedding dimension (Figure 6.5). They might be influenced by the seasonal components of time series. The saturation value of the correlation exponents tends to decrease with increasing seasonality (Khan et al., 2005). Therefore, the CD value of Bri_Gw3 is not reliable. From the CD results of the other wells in Briesensee (Table 6.1), it seems that there is only one process dominating the groundwater level of unconfined aquifer, less processes than for the lake level (i.e. 2). This process might relate to precipitation minus evaporation. Since the studied lakes do not exhibit any inlet or outlet stream, the lake level likely depends on groundwater dynamics as well as on local hydro-climatic conditions (Lischeid et al., 2010).

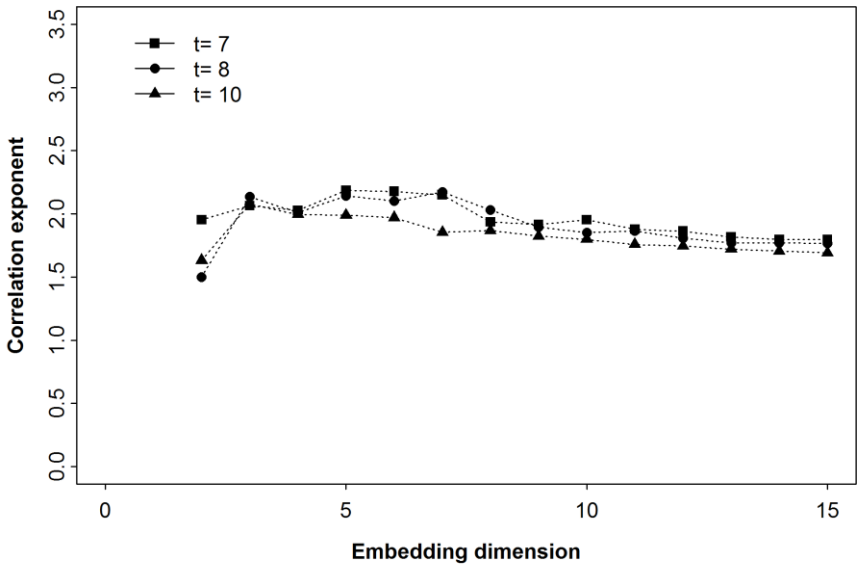

Figure 6.5 Correlation exponents versus embedding dimension plot for different time lags (i.e. 7, 8 and 10).

From the CD analysis of groundwater level data, the corresponding numbers of dominant processes are 4 to 5 in the aquifer of the Redernswalder See and 1 in the aquifer of Briesensee. That is, the system dynamics of groundwater level in the region of Redernswalder See is more complex than that of Briesensee. The reason for the higher CD values in Redernswalder See aquifer might be that these groundwater wells are screened in a confined aquifer. In a confined aquifer, groundwater head data does not only reflect hydrological processes, but also fluctuates with atmospheric pressure which would yield larger CD values. In general, barometric effects and water level changes are greater and more pronounced in confined aquifers (e.g. Rasmussen and Crawford, 1997). Lischeid et al. (2010) showed that Red_Gw3 and Red_Gw4 were screened in a confined aquifer and their high-passed groundwater levels were strongly negatively correlated with barometric pressure. They also gave some evidences that Red_Gw1 was affected by atmospheric pressure fluctuations. It seems that high CD values indicate (partly) confined conditions which would make the CD method very promising. While the groundwater head within the open well is instantly affected by a barometric pressure change, the total head within the aquifer may or may not be affected by those changes (Rasmussen and Crawford, 1997). Therefore, these CD values may represent the number of dominant

processes of groundwater head measured in the groundwater wells instead of representing the complexity of confined aquifer dynamics.

Furthermore, even in the same aquifer system, the CD method can also distinguish different complexities of underlying dynamics considering the variation of the CD values. The higher the CD values, the more complex the system dynamics turn to be. For example, the Red_Gw1 system dynamics is more complex than that of Red_Gw3, since the CD value of Red_Gw1 is 5, higher than the values of Red_Gw3 (Table 6.1).

7 Applicability and paractibility of the Correlation Dimension method

7.1 The robustness of the Correlation Dimension method

The CD analysis of physical and mathematical systems offers an alternative way to investigate natural systems. In this thesis, the CD method was used for assessing the number of dominant processes. This information is especially valuable when physical mechanisms governing hydrologic system are studied from the data themselves in an inverse manner (Sivakumar, 2008).

The CD method provides independent information compared with standard time series analysis approaches. For the first CD method application (Chapter 4), the relationships between CD values, the variance of observed discharge time series, autocorrelation for different time lags, the slope of power spectrum, the Hurst coefficient, the variance of PCA scores and PCA loadings were not significant for hydrographs from 35 catchments in Federal State of Brandenburg, Germany. That is, the CD method provides information independent from that of standard time series analysis approaches. Moreover, the second CD method application (Chapter 5) also showed that the CD largely gave independent information compared with autocorrelation and power spectrum analysis.

For the observed time series, the CD method yields reliable values to assess the complexity (i.e. intrinsic dimensionality, degree of freedom or the number of dominant processes) of observed data, which is useful for catchment classification and system dynamics exploration. The classification based on the dominant processes has more solid physical and mathematical foundations than the common traditional methods. The underlying system dynamics of groundwater head and lake level, explored by the CD method (Chapter 6), indicated that the confined aquifer involved more dominant processes governing the groundwater head dynamics than unconfined aquifer. These evidences prove that the CD method offers inherent further information to understand the underlying system dynamics of data.

For hydrological models, the CD method helps the modeller to exclude models with insufficient complexity, thus reducing the equifinality problem and model structure uncertainties. It proved to be more powerful than the traditional NSc efficiency criterion (Chapter 5). The combination of the CD method and NSc efficiency criterion seems to detect the number of dominant processes and improve the prediction accuracy, serving as a promising evaluation approach of model performance. In addition, the CD method gives a new model validation approach besides the simple comparison of observed and simulated hydrographs. It assesses the variations of the number of dominant processes accounting for the respective observations and simulations. This helps to check how the simulations represent the observed system dynamics in a physical and mathematical way.

In summary, all these evidences reveal that the CD method is able to assess the number of dominant processes that can represent the complexity of hydrological systems and provides a lot of valuable information about the hydrological system dynamics. The method can be applied to the other systems, such as geology, hydro-chemistry, atmospheric science, climate and global change, ocean science etc.

7.2 Methodological problems of the Correlation Dimension method

Since Taken's theorem (1981) presupposes series of infinite length and being completely avoid of measurement errors, any practical application of the CD method obviously proves heuristic. Some problems had been addressed for CD analysis, including the identification of proper time lags and embedding dimension, the minimum data size, noise level determination and reduction, the effect of intermittency, the presence of zero values in rainfall data, high autocorrelation. A small data size may result in a significant underestimation of the CD value, whereas the presence of noise may overestimate the CD value. A large time lag may overestimate the CD value, while a small time lag may underestimate the CD value. In Chapter 2, we have already discussed the identification of proper time lags and embedding dimensions (Section 2.1.1.1 and 2.1.1.2), the minimum data size (Section 2.1.3) and noise reduction (Section 2.1.4). Consequently different strategies are suggested to handle these issues. In this section, we will discuss other methodological problems.

The CD method employs the scaling range chosen procedure. No automatic algorithm is powerful to accomplish this task. Therefore, the determination of the CD value remains a semi-automatic program based on the expert's knowledge. In common, the CD analysis of one time series may cost a day for beginners, but only takes an hour for experienced applicators. The main reasons to consume so much time are the visual inspection to identify the scaling range and lacking of powerful methods to identify the proper time lag τ and embedding dimension m. It is a serious problem to choose the right scaling range in the $\log C(r)$ versus $\log r$ plot, when the data are polluted by noise. For example, noise in the data may lead to two different short linear ranges, resulting in a wrong decision of the applicator for choosing the right scaling range. We proposed the distinct window techniques to tract the scaling range. However, visual inspection and the CD estimation remain difficult. As we checked many previous studies of the CD analysis, no clear universal principle was found for the scaling range identification. Therefore, achieving the robust automatic CD algorithm depends on the development of a powerful method to identify the correct scaling range, proper time lag τ and embedding dimension m.

An important limitation of the CD analysis of rainfall time series is the presence of a large numbers of zero values. Some papers (e.g. Tsonis et al., 1993; Sivakumar et al., 2001) pointed out that amounts of zero rainfall might underestimate the CD values because the reconstructed hyper-surface in phase space would tend to a point when a certain value dominates in a time series, whereas Sivakumar (2005a) offered evidences that the presence of zero rainfall may not always result in an underestimation of the CD values. In fact, zero rainfall is also indicative of and equally important to understand how the dynamics of the system evolved, and the elimination of zero rainfall could lead to unrealistic estimates in the CD estimation of the rainfall time series. Therefore, the CD analysis of rainfall time series should consider zero values of rainfall.

Even including the amounts of zero values, Koutsoyiannis (2006) showed that the CD was simply zero for any embedding dimension for a given example. This phenomenon is called intermittency. Intermittency is not unique to rainfall series, but also exists in discharge time series which exhibit intermittency without zero values. Intermittency in discharge series performs a J-shape distribution (i.e. that is defined for positive values of

the variable and has a high coefficients of skewness), and produce random points whose largest percentage are close to zero and a small number of points take large values. Graf von Hardenberg et al. (1997) demonstrated that this equivalent of intermittency makes the estimated CD value very small, although the actual dimension of the system is infinite. Finally, they proposed ways to refine the algorithm in order to obtain correct results. The simplest of them is to filter the data by excluding all the delay vectors x having at least one component $x_i < c$, where c is typically a small percentage (e.g. 5%) of the average of the data series. The appropriate cut-off value c leaves out all off data points of the intermittent time series. This simple algorithm was proven very effective (Koutsoyiannis, 2006). From our investigations, none of the studied time series exhibited the intermittency problem in the $\log C(r)$ against $\log r$ plot.

Hydrological time series, especially on fine time scales (e.g. hourly or daily), are characterized by high autocorrelation coefficients. The effects of autocorrelation may act synergistically with the effects of an asymmetric distribution function and the effect of sample size (Koutsoyiannis, 2006). In auto-correlated series, a larger number of data points may not suffice to avoid misleading results. Therefore, the Theiler window is needed to exclude temporal correlated points from the pair counting. It is usually identified by the space time separation plot (Provenzale et al., 1992; Kantz and Schreiber, 2004). In this plot, the parameter "time lag τ" is employed, and hence the appropriate selection of the time lag in constructing delay vectors should be cautious (Section 2.1.1.1).

It is useful to know that the CD as a popular fractal dimension tends to underestimate the true dimensionality and noise may pollute the estimation (Lee and Verleysen, 2006). Since each method has its limitations, it would be unwise to think that a given problem can be solved using one particular method. Given that each method often possesses different advantages and limitations, it would be better to attempt to maximize their advantages and minimize their disadvantages.

8 Conclusions

Hydrological models are often developed for specific situations and thus their extensions and generations to other situations are difficult. Therefore, it is necessary to classify catchments regarding their behavior or catchment properties. In the first study, the observed hydrographs of 35 small catchments in the Federal State of Brandenburg, Germany, were analyzed by the CD method. The majority of catchments exhibited CD values between 3 and 6. The CD values did not exhibit any pronounced spatial pattern, but they displayed significant correlations with the heterogeneity of the catchment, which seems to be the main factor explaining high CD values of the observed hydrographs. Forest area percentage of land use is positively correlated with CD values, while agricultural land area percentage is negatively correlated. It seems that the underlying system dynamics of forest area are more complex than agricultural land with respect to the observed runoff behavior. Moreover, the CD method captures different information independent from that of standard time series analysis methods, including variance, autocorrelation analysis, power spectrum analysis, Hurst analysis and PCA. It can estimate the intrinsic dimensionality of the data and provide information on the minimum number of processes underlying the signal dynamics. This information makes the CD method as an ideal candidate to reduce model uncertainties.

In the second study, the model equifinality problem was addressed by comparing three hydrological models with different model complexities and different parameterizations, applied to the same two catchments. Equifinality arose in any of the three models. Results from models with similar and widely differencing parameter values were analysed by the CD method. Simulations with similar NSc values differed considerably with respect to CD values, but the CD values converged for the best simulation in terms of NSc values towards that of the observed discharge. Therefore, the CD method is proposed to tackle the equifinality problem in hydrology. The new suggested approach that combined the NSc efficiency criterion and the CD method seems to be more powerful compared to the usual approach. It can save the time, but also detect the intrinsic properties of system dynamics. In addition, the CD analyses of model rainfall, evapotranspiration and discharge time series suggest that hydrological models likely act as intrinsic

dimensionality reducing filters for the high-dimensional model input. If the model complexity increases, hydrological models reduce the intrinsic dimensionality of simulations. Correspondingly, a minimum model complexity seems to be required to fulfill this task rather than raising the intrinsic dimensionality of the input data.

In the third study, the CD method was applied to groundwater head and lake level data in the biosphere reserve Schorfheide-Chorin region. The intrinsic dimensionality of groundwater level ranged from 0.9 to 5, while lake level performed small variations, around 1.57 to 2. The CD values of groundwater level exhibited no correlation with the screening depth of groundwater wells, but displayed spatial patterns for confined and unconfined aquifer, respectively. It seems that high CD values indicate partly confined conditions. Furthermore, the CD method can also recognize the different complexities of underlying dynamics within the same aquifer system. The higher the CD values, the more complex system turns to be.

In summary, the CD method can estimate the reliable number of dominant processes and it seems to be a powerful method for system dynamics exploration, catchment classification and model evaluation. However, it cannot identify the actual processes occurred in the system. Therefore, it is a great challenge to achieve the larger goal of identifying the dominant processes since none of the existing methods seems capable of fulfilling this task.

Referrences

Abebe NA, Ogden FL, Pradhan NR. 2010. Sensitivity and uncertainty analysis of the conceptual HBV rainfall-runoff model: Implications for parameter estimation. *Journal of Hydrology* **389**: 301-310.

Akselrod S, Gordon D, Ubel FA, Shannon DC, Barger AC, Cohen RJ. 1981. Power spectrum analysis of heart rate fluctuation: a quantitative probe of Beat-To-Beat Cardiovascular control. *Science* **213**: 220-222.

Bao W, Wang C. 1997. Structure and application of Vertical-mixed Runoff Model. *Journal of Hydrology China* **3**: 18-21.

Beven K. 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources* **16**: 41-51.

Beven K, Freer J. 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology* **249**: 11-29.

Bossard M, Feranec J, Otahel J. 2000. CORINE land cover technical guide - Addendum 2000. European Environment Agency.

Buchala S, Davey N, Gale TM, Frank RJ. 2005. Analysis of linear and nonlinear dimensionality reduction methods for Gender classification of face images. *International Journal of Systems Science* **36**: 931-942.

Casdagli M. 1989. Nonlinear prediction of chaotic time series. *Physica D* **35**: 335-356.

Central Basic Geodata Service for Germany, VG250, 2012. Geo Information: http:// www.geodatenzentrum.de.

Cheng CT, Ou CP, Chau KW. 2002. Combing a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall-runoff model calibration. *Journal of Hydrology* **268**: 72-86.

Cheng CT, Zhao MY, Chau KW, Wu XY. 2006. Using genetic algorithm and TOPSIS for Xinanjiang model calibration with a single procedure. *Journal of Hydrology* **316**: 129-140.

Das A, Das P, Roy AB. 2002. Nonlinear data analysis of experimental (EEG) data and comparison with theoretical (ANN) data. *Complexity* **7**: 30-40.

Decoster GP, Mitchell D. 1991. The efficacy of the correlation dimension technique in detecting determinism in small samples. *Journal of Statistical Computation and Simulation* **39**: 221-229.

Ding M, Grebogi C, Ott E, Sauer T, Yorke JA. 1993. Estimating correlation dimension from a chaotic time series: when does plateau onset occur? *Physica D* **69**: 404-424.

Draper NR, Smith H. 1981. Applied Regression Analysis, second edition. John Wiley & Sons.

Elshorbagy A, Simonovic SP, Panu US. 2002a. Noise reduction in chaotic hydrologic time series: facts and doubts. *Journal of Hydrology* **256**: 147-165.

Elshorbagy A, Simonovic SP, Panu US. 2002b. Estimation of missing streamflow data using principles of chaos theory. *Journal of Hydrology* **255**: 123-133.

Farmer JD, Sidorowich JJ. 1987. Predicting chaotic time series. *Phys. Rev. Lett.* **59**: 845-848.

Federal Ministry for the Environment, Nature Conservation and Nuclear Sefety, 2003. Hydrological Atlas of Germany 1961-1991 (Hydrologischer Atlas von Deutschland). www.bmu.de, Bonn.

Fu G, Liu C, Chen S, Hong J. 2004. Investigating the conversion coefficients for free water surface evaporation of different evaporation pans. *Hydrological processes* **18**: 2247-2262.

Fiering MB. 1967. Streamflow Synthesis. Harvard University Press. Online source: http://www.google.de/url?sa=t&rct=j&q=Fiering%2Babc%2Bmodel&source=web&cd=1&ved=0CC8QFjAA&url=http%3A%2F%2Fwww.geogr.uni-jena.de%2Ffileadmin%2FGeoinformatik%2FLehre%2FSoSe_2006%2FModul241%2FabcModel.pdf&ei=XOzTUd3dHcLtswaTyIGwAQ&usg=AFQjCNH7SSfgAJFrVgXfWu_7b4BIWKcQ9A&bvm=bv.48705608,d.Yms&cad=rjt

Fraedrich K. 1986. Estimating the dimensions of weather and climate attractors. *Journal of Atmospheric Sciences* **43**: 419-432.

Fraser AM, Swinney HL. 1986. Independent coordinates for strange attractors from mutual information. *Physical Review A* **33**: 1134-1140.

Gaume E, Sivakumar B, Kolasinski M, Hazoume L. 2006. Identification of chaos in rainfall temporal disaggregation: Application of the correlation dimension method to 5-minute point rainfall series measured with a tipping bucket and an optical raingage. Journal of Hydrology 328: 56-64.

Germer S, Kaiser K, Bens O, Hüttl RF. 2011. Water balance changes and responses of Ecosystems and society in the Berlin-Brandenburg region--a review. In: Global change: Challenges for regional water resources, DIE ERDE, pp: 65-95.

Gomez-Plaza A, Martinez-Mena M, Albaladejo J, Castillo VM. 2001. Factors regulating spatial distribution of soil water content in small semiarid catchments. *Journal of Hydrology* **253**: 211-226.

Graf von Hardenberg J, Paparella F, Platt N, Provenzale A, Spiegel EA, Tesser C. 1997. Missing motor of on-off intermittency. *Physical Review E* **55**: 58-64.

Grassberger P, Procaccia I. 1983a. Measuring the strangeness of strange attractors. *Physica D* **9**: 189-208.

Grassberger P, Procaccia I. 1983b. Estimation of the Kolmogorov entropy from a chaotic signal. *Physical Review A* **28**: 2591-2593.

Grayson RB, Blöschl G. 2000. Summary of pattern comparison and concluding remarks. Cambridge University Press.

Harlin J, Kung CS. 1992. Parameter uncertainty and simulation of design floods in Sweden. *Journal of Hydrology* **137**: 209-230.

Hegger R, Kantz H, Schreiber T. 1999. Practical implementation of nonlinear time series methods: The TISEAN package. *Chaos* **9**: 413-435.

Holzfuss J, G.Mayer-Kress. 1986. An approach to error-estimation in the application of dimension algorithms. In: Dimensions and Entropies in Chaotic System, Mayer-Kress G (ed.) Springer, pp: 114-122.

Hossian F, Sivakumar B. 2006. Spatial pattern of arsenic contamination in shadow wells of Bangladesh: regional geology and nonlinear dynamics. *Stochastic Environmental Research Risk Assessment* **20**: 66-76.

Jayawardena AW, Lai F. 1994. Analysis and prediction of chaos in rainfall and stream flow time series. *Journal of Hydrology* **153**: 23-52.

Jolliffe IT. 2002. Principle Component Analysis, second edition. Springer. http://scholar.google.de/scholar_url?hl=zh-CN&q=http://hbanaszak.mjr.uw.edu.pl/MarketingoweZastosowania/PCA/Jolliffe_2002_PrincipalComponentAnalysis.pdf&sa=X&scisig=AAGBfm0br1cLJi6yHc41PSU24cY-aFL4bQ&oi=scholarr&ei=z-TTUc4XhouzBtvDgMAF&ved=0CCwQgAMoADAA

Kantz H, Schreiber T. 2004. Nonlinear time series analysis. 2nd Edn., Cambridge University Press.

Kennel MB, Brown R, Abarbanel HDI. 1992. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A* **45**: 3403-3411.

Kim HS, Yoon YN, Kim JH, Kim JH. 2001. Searching for strange attactor in wastewater flow. *Stochastic Environmental Research Risk Assessment* **15**: 399-413.

Khan S, Ganguly AR, Saigal S. 2005. Detection and predictive modeling of chaos in finite hydrological time series. *Nonlinear Processes in Geophysics* **12**: 41-53.

Kirchner JW, Feng X, Neal C. 2000. Fractal stream chemistry and its implications for contaminant transport in catchments. *Nature* **403**: 524-526.

Koutsoyiannis D. 2006. On the quest for chaotic attractors in hydrological processes. *Hydrological Sciences Journal* **51**: 1065-1091.

Krause P, Boyle DP, Bäse F. 2005. Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences* **5**: 89-97.

Lange H. 1999. Time series analysis of Ecosystem variables with complexity measures. *Interjournal for complex systems* **250**.

Lee JA, Verleysen M. 2006. Nonlinear dimensionality reduction. Springer. Chapter 3: Estimation of intrinsic dimensionality. Page 48-55.

Lindström G, Johansson B, Persson M, Gardelin M, Bergström S. 1997. Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology* **201**: 272-288.

Lischeid G. 2009. Non-linear visualization and analysis of large water quality data sets: a model-free basis for efficient monitoring and risk assessment. *Stoch Environ Res Risk Assess* **23**: 977-990.

Lischeid G, Bittersohl J. 2008. Tracing biogeochemical processes in stream water and groundwater using non-linear statistics. *Journal of Hydrology* **357**: 11-28.

Lischeid G, Natkhin M. 2011. The potential of Land-Use change to mitigate water scarcity in Northeast Germany--a review. In: Global change: Challenges for regional water resources, DIE ERDE, pp: 97-113.

Lischeid G, Natkhin M, Steidl J, Dietrich O, Dannowski R, Merz C. 2010. Assessing coupling between lakes and layered aquifers in a complex Pleistocene landscape based on water level dynamics. *Advances in Water Resources* **33**: 1331-1339.

Lorenz EN. 1963. Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences* **20**: 130-141.

Lorenz EN. 1991. Dimension of weather and climate attractors. *Nature* **353**: 241-244.

Madsen H. 2000. Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *Journal of Hydrology* **235**: 276-288.

Mandelbrot B. 1977. Fractals: Form, chance and dimensions. W. H. Freeman & Company.

Mandelbrot BB, Wallis JR. 1969. Robustness of the rescaled range R/S and the measurement of noncyclic long run statistical dependence. *Water Resources Research* **5**: 967-988.

Marwan N, Romano MC, Thiel M, Kurths J. 2007. Recurrence plots for the analysis of complex systems. *Physics Reports* **438**: 237-239.

Merz C, Pekdeger A. 2011. Anthropogenic changes in the Landscape Hydrology of the Berlin-Brandenburg region. In: Global change; Challenges for regional water resources, DIE ERDE, pp: 21-39.

Moon Y, Rajagopalan B, Lall U. 1995. Estimation of mutual information using kernel density estimators. *Physical review E* **52**: 2318-2321.

Nützmann G, Wolter C, Venohr M, Pusch M. 2011. Historical Patterns of Anthropogenic impacts on freshwaters in the Berlin-Brandenburg region. In: Global change: Challenges for regional water resources, DIE ERDE, pp: 41-64.

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology* **10**: 282-290.

Nerenberg MAH, Essex C. 1990. Correlation dimension and systematic geometric effects. *Physical Review A* **42**: 7065-7074.

Osborne AR, Provenzale A. 1989. Finite correlation dimension for stochastic systems with power-law spectra. *Physica D* **35**: 357-381.

Ott E. 1993. Chaos in dynamical systems. Cambridge University Press. Page 6.

Porporato A, Ridolfi L. 1997. Nonlinear analysis of river flow time sequences. *Water Resources Research* **33**: 1353-1367.

Provenzale A, Smith LA, Vio R, Murante G. 1992. Distinguishing between low-dimensional dynamics and randomness in measured time series. *Physica D* **58**: 31-49.

Provenzale A, Vio R, Cristiani S. 1994. Luminosity variations of 3C 345: Is there any evidence of low-dimensional chaos? *Astrophysical Journal* **428**: 591-598.

Qu S, Bao W, Shi P. 2007. A comparative study of the Xinanjiang model and the Vertical-mixed runoff model. IAHS Publ. 311, pp: 1-8.

Rasmussen TC, A.Crawford L. 1997. Indentifying and Removing Barometric Pressure Effects in Confined and Unconfined Aquifers. *Ground Water* **35**: 502-511.

R Development Core Team, 2006. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN: 3-900051-07-0, <http://www.Rproject.org>.

Renyi A. 1970. Probability theory. American Elsevier Publishing Company, New York. Online source: http://www.google.de/url?sa=t&rct=j&q=Renyi%2BA.%2B1970.%2BProbability%2Btheory.&source=web&cd=4&ved=0CD8QFjAD&url=http%3A%2F%2Fwww.ams.org%2Fbull%2F1973-79-02%2FS0002-9904-1973-13147-7%2FS0002-9904-1973-13147-7.pdf&ei=_vDTUcLKLojLtAbi-YGwCQ&usg=AFQjCNH0zQ0mkHYZr1rjxGHLtVKkt9fkww&bvm=bv.48705608,d.Yms&cad=rjt

Rhodes C, Morari M. 1997. The false nearest neighbors algorithm: an overview. *Comput. Chem. Eng.* **21**: 1149-1154.

Rodriguez-Iturbe I, De-Power BF, Sharifi MB, Georgakakos KP. 1989. Chaos in Rainfall. *Water Resources Research* **25**: 1667-1675.

Schertzer D, Tchiguirinskaia I, Lovejoy S, P.Hubert, Bendjoudi H. 2002. Which chaos in the rainfall-runoff process? A discussion on'Evidence of chaos in the rainfall-runoff process' by Sivakumar et al. *Hydrological Sciences Journal* **47**: 139–147.

Schouten JC, Takens F, Bleek CMvd. 1994. Estimation of the dimension of a noisy attractor. *Physical Review E* **50**: 1851-1861.

Schreiber T. 1993. Extremely simple nonlinear noise reduction method. *Physical Review E* **47**: 2401-2404.

Schreiber T, Grassberger P. 1991. A simple noise reduction method for real data. *Physics letters A* **160**: 411-418.

Schreiber T, Kantz H. 1996. Observing and predicting chaotic signals: Is 2% noise too much? , Springer: 1-22. Online source: http://www.google.de/url?sa=t&rct=j&q=Fiering%2Babc%2Bmodel&source=web&cd=1&ved=0CC8QFjAA&url=http%3A%2F%2Fwww.geogr.uni-jena.de%2Ffileadmin%2FGeoinformatik%2FLehre%2FSoSe_2006%2FModul241%2FabcModel.pdf&ei=XOzTUd3dHcLtswaTyIGwAQ&usg=AFQjCNH7SSfgAJFrVgXfWu_7b4BIWKcQ9A&bvm=bv.48705608,d.Yms&cad=rjt

Schuster HG. 1988. Deterministic Chaos. VCH. Weinheim. Online source: http://freepdfdb.org/pdf/deterministic-chaos-schuster

Seibert J. 2005. HBV light version 2: User's mnual. Stockholm University, Department of physical geography and quaternary geology.

Sivakumar B. 2001. Is a chaotic multi-fractal approach for rainfall possible? *Hydrological processes* **15**: 943-955.

Sivakumar B. 2002. Is correlation dimension a reliable indicator of low-dimensional chaos in short hydrological time series? *Water Resources Research* **38**: 1011. DOI: 10.1029/2001WR000333.

Sivakumar B. 2004a. Chaos theory in geophysics: past, present and future. *Chaos, Solitons and Fractals* **19**: 441-462.

Sivakumar B. 2004b. Dominant processes concept in hydrology: moving forward. *Hydrological processes* **18**: 2349-2353.

Sivakumar B. 2005a. Chaos in rainfall: variability, temporal scale and zeros. *Journal of Hydroinformatics* **7.3**: 175-184.

Sivakumar B. 2005b. Correlation dimension estimation of hydrological series and data size requirement: myth and reality. *Hydrological Sciences Journal* **50**: 591-603.

Sivakumar B. 2007. Nonlinear determinism in river flow: prediction as a possible indicator. *Earth Surface Processes and Landforms* **32**: 969-979.

Sivakumar B. 2008. Dominant Processes Concept, Model Simplification and Classification Framework in Catchment Hydrology. *Stochastic Environmental Research Risk Assessment* **22**: 737-748.

Sivakumar B. 2009. Nonlinear dynamics and chaos in hydrologic systems: latest developments and a look forward. *Stochastic Environmental Research Risk Assessment* **23**: 1027-1036.

Sivakumar B, Jayawardena AW, Li WK. 2007. Hydrologic complexity and classification: a simple data reconstruction approach. *Hydrological processes* **21**: 2713-2728.

Sivakumar B, Phoon KK, Liong SY, Liaw CY. 1999. A systematic approach to noise reduction in observed chaotic time series. *Journal of Hydrology* **219**: 103-135.

Sivakumar B, Singh VP. 2012. Hydrologic system complexity and nonlinear dynamic concepts for a catchment classification framework. *Hydrology and Earth System Sciences* **16**: 4119-4131.

Sivakumar B, Sorooshian S, Gupta HV, Gao X. 2001. A chaotic approach to rainfall disaggregation. *Water Resources Research* **37**: 61-72.

Sivakumar B, Wallender WW, Puente CE, Islam MN. 2004. Streamflow disaggregation: a nonlinear deterministic approach. *Nonlinear Processes in Geophysics* **11**: 383-392.

Sivalumar B. 2000. Chaos theory in hydrology: important issues and interpretations. *Journal of Hydrology* **227**: 1-20.

Smith LA. 1988. Intrinsic limits on dimension calculations. *Physics letters A* **133**: 283-288.

Sonin AA. 2001. The physical basis of dimensional analysis, second edition. Page 6. Online source: http://www.google.de/url?sa=t&rct=j&q=Sonin%2BAA.%2B2001.%2BThe%2Bphysical%2Bbasis%2Bof%2Bdimensional%2Banalysis&source=web&cd=1&ved=0CC8QFjAA&url=http%3A%2F%2Fweb.mit.edu%2F2.25%2Fwww%2Fpdf%2FDA_unified.pdf&ei=GevTUY79LomRswbf0ICQAQ&usg=AFQjCNH0Zupq6nh1jxaH-tV5uqG_Syz1Zg&bvm=bv.48705608,d.Yms&cad=rjt

State Office for Mining, Geology and Raw Material of Brandenburg, 2012a. Hydrogeologic Map of Brandenburg: http://www.geo.brandenburg.de/hyk50.

State Office for Mining, Geology and Raw Material of Brandenburg, 2012b. Geo Informations: http://www.mugv.brandenburg.de/cms/detail.php/lbm1.c.200103.de.

Takens F. 1981. Detecting strange attractors in turbulence. In: Dynamical Systems and Turbulence, Lecture Notes in Mathematics, Rand DA, Young, L.S. (ed.) Springer, pp: 366-381.

Theiler J. 1986. Spurious dimension from correlation algorithms applied to limited time-series data. *Physical Review A* **34**: 2427-2432.

Thomas B, Lischeid G, Steidl J, Dannowski R. 2012. Regional catchment classification with respect to low flow risk in a Pleistocene landscape. *Journal of Hydrology* **475**: 392-402.

Togal H, Demirel MC, Booij MJ. 2012. Seasonality of low flows and dominant processes in the Rhine River. *Stochastic Environmental Research Risk Assessment*. DOI: 10.1007/s00477-012-0594-9.

Tsonis AA, Elsner JB. 1988. The weather attractor over very short timescales. *Nature* **333**: 545-547.

Tsonis AA, Elsner JB, Georgakakos KP. 1993. Estimating the dimension of weather and climate attractors: important issues about the procedure and interpretation. *Journal of the Atmospheric Sciences* **50**: 2549-2555.

Tsonis AA, Triantafyllou GN, Elsner JB. 1994. Searching for determinism in observed data: a review of the issue involved. *Nonlinear Processes in Geophysics* **1**: 12-25.

Wang Q, Gan TY. 1998. Biases of correlation dimension estimates of streamflow data in the Canadian prairies. *Water Resources Research* **34**: 2329–2339.

Wolf A, Swift JB, Swinney HL, Vastano JA. 1985. Determining Lyapunov Exponent from a time series. *Physica 16D*: 285-317.

Yapo PO, Gupta HV, Sorooshian S. 1996. Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. *Journal of Hydrology* **181**: 23-48.

Yapo PO, Gupta HV, Sorooshian S. 1998. Multi-objective global optimization for hydrologic models. *Journal of Hydrology* **204**: 83-97.

Yu PS, Yang TC. 2000. Fuzzy multi-objective function for rainfall-runoff model calibration. *Journal of Hydrology* **238**: 1-14.

Zhang X, Lindström G. 1996. A comparative study of a Swedish and a Chinese hydrological model. *Water Resources Bulletin* **32**.

Zhao RJ. 1992. The Xinanjiang model applied in China. *Journal of Hydrology* **135**: 371-381.

Appendix I - List of publications

Ma M, Lischeid G, Merz C, Thomas B., *under review*. "Correlation dimension analysis of observed discharge time series in small catchments: what information can the method provide?", *Journal of Hydrology*.

Ma M, Lischeid G, Merz C., *under review*. "Using the Correlation Dimension analysis to evaluate model performance", *Environmental modeling & software*.

For reasons of data protection, the Curriculum Vitae is not published in the online version.

For reasons of data protection, the Curriculum Vitae is not published in the online version.