IJCMER
www.ijcmer.org

INTERNATIONAL JOURNAL OF CHEMICAL, MATERIAL AND ENVIRONMENTAL RESEARCH

An International Open Free Access, Peer Reviewed Research Journal

A Commentary Review on Data-Intensive Analysis of River Water Quality as a Nonlinear Dynamic System

[1,]Hongxing Song, [2,]Yun Yan , [3,]Meiyu Luo, [4,]Yang Gu, [5,]Rong He

[1,2]Hydrology and Water Quality Center of Shenzhen, Shenzhen 518055, China

[3,4] School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

[3,4,5]Shenzhen Zhishu Environmental Technology Co., Ltd, Shenzhen 518055, China

ABSTRACT： Investigating the nonlinear behaviors of river water quality changes are of significant for catchment science. The emerging of data-intensive pathway for scientific investigation opens a new era for environment and hydrology science. In this review paper, two kinds of river water quality systems are firstly defined. Their associated time domain chaos behavior, frequency domain fractal behavior, teleconnection behavior based on complex network, and information entropy characterization, are discussed under a unified framework. Knowledge gaps on this data-intensive pathway are identified and future studies relative to phenomenological theory are proposed. It is necessary to establish the model with high precision data to capture the chaos information of water quality time series to a greater extent. The question that whether the complex network exists widely in river quality system and how to use complex network to instruct water quality modeling and management need to be solved. Besides, the research of information domain is in the preliminary stage, it is a mainstream direction to directly learn the evolution characteristics of it for water quality system through deep learning technology and big data. This commentary wakes up an open area for many potential interesting findings in water environment science.

KEYWORDS: data-intensive; nonlinear system; chaos; fractal; complex network; information entropy; water quality system

I. INTRODUCTION

Since the beginning of human exploration of science, scientific research has entered the three stages: experimental observation, theoretical deduction and computational simulation, and is now entering the fourth paradigm data-intensive research (Chen and Zhang 2014; Gorton et al. 2008; Hey, Tansley, and Tolle 2009; Kouzes et al. 2009). Compared to computational simulation, which derives the result based on the understanding of the law of causation determined by known physics regulation, the data intensive paradigm establishes complex correlations between multidimensional parameters based on data analyzed by algorithms or phenomenology. The reason for this paradigm shift in science is the explosion of experimental observational data and the unique superiority of artificial intelligence techniques, providing a new way to analytically solve complex scientific problems in complex systems (Hochachka et al. 2012; Hogle 2016; Kelling et al. 2009; Michener and Jones 2012; Newman, Ellisman, and Orcutt 2003; Peters-Lidard et al. 2017).

Nonlinear science is a comprehensive basic science that reveal the law of complex movement in regional systems (Drazin and Drazin 1992), which is also an important frontier in the development of modern hydrological theory and applied basic research on water resources security under changing environments (Kundzewicz and Napiórkowski 1986). It has drawn a lot of attention from scholars and is currently focused on exploring the issue of water quantity change (Kirchner 2009).The surface water environmental system characterized by water quality indicators is likewise an open, complex, nonlinear system, as well as a dynamic non-equilibrium composite system (Kirchner and Neal 2013), as shown in Figure 1. Systematic analysis of spatio-temporal changes in river water quality could provide a significant scientific basis for watershed water quality management and water resources development and utilization.

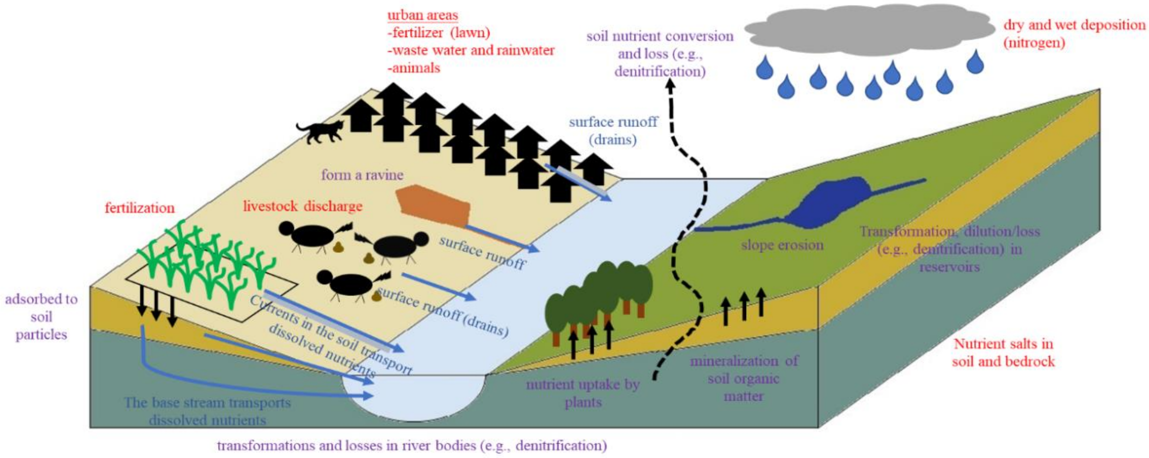

Figure 1. Diagram of water transport process

Nonlinear effects in water systems are generally studied using mechanistic models based on physical processes and systematical analysis based on data-driven techniques. Currently, in the field of hydrology and water environments, the study of water quality processes on the land surface or in-stream is more often analyzed using physical process models(Zhang, Zhang, and Liu 2007). However, it is difficult to exhaust all physical-chemical-biological processes in a physical process model, and even if all of them are described, the model will be too complex for practical application.

Therefore, in many cases, valuable information needs to be extracted from the growing wealth of data through suitable data-driven or phenomenological methods (Chin 2008; Powers et al. 1994). However, due to the high cost of water quality monitoring and interdisciplinary reasons, the research on river water quality change patterns from the system approach is extremely limited. Theoretically, phenomenological analysis based on time series could uncover the universal laws and essential features of the nonlinear behavior of water quality systems (Kirchner, Feng, and Neal 2000; Kirchner and Neal 2013). With the rapid development of online monitoring, big data, artificial intelligence, and other emerging technologies in recent years, the trend of expanding the integration of water science is becoming more obvious, which creates an important opportunity to re-explore the basic laws of water quality system changes from the perspective of data science and phenomenology. It is necessary to present a summary of the existing research and future development of data intensive analysis of river water quality as a nonlinear dynamic system (Govindaraju 2000; Kim et al. 2021; Lange and Sippel 2020).

II. DEFINITION OF TWO TYPES OF RIVER WATER QUALITY SYSTEM

Factors affecting river water quality come from a variety of sources including physics, chemistry, biology, hydrology, meteorology, and human activities. The spatial and temporal patterns of river water quality driven by these complex variables can be analyzed and studied from two different perspectives. From one perspective, they can be seen as a process of transport of material energy (dissolved matter, substrate, temperature, bacteria, etc.) in the river, which is usually described using a physical process model (Xu and Liao 2003). From the other perspective, they are regarded as a collection of parameters reflecting the quality state of the target water body, or as a system in the statistical–physical sense, characterized by a data-driven model or phenomenology model. In this paper, we define two types of water quality systems observed under different perspectives as water quality system-I (WQS-I) and water quality system-II (WQS-II). They are different in terms of analytical methods, purpose, and scope of application.

The spatial and temporal variations in water quality in rivers include temporal variation at a point, spatiotemporal variation in the diffusion of pollutants, and temporal variation in the overall water quality status of a watershed or region. In this article, we propose the following definitions: (1) A change in the water quality concentration at specific points at different time resolutions is called a fluctuation in water quality. Fluctuation in WQS-I is manifested in the form of a pollutant penetration curve and is monitored in time series of water quality in WQS-II. (2) The overall spatial and temporal variation in water quality in a watershed refers to evolution, which is expressed as the transport process of pollution masses in WQS-I and the trend of water quality change in WQS-II. In addition, due to topography, interactions between surface water and groundwater, meteorology, human activities, and other factors, fluctuations in water quality at different points in different watersheds must have a connection, and some of them are teleconnections that are difficult to observe.

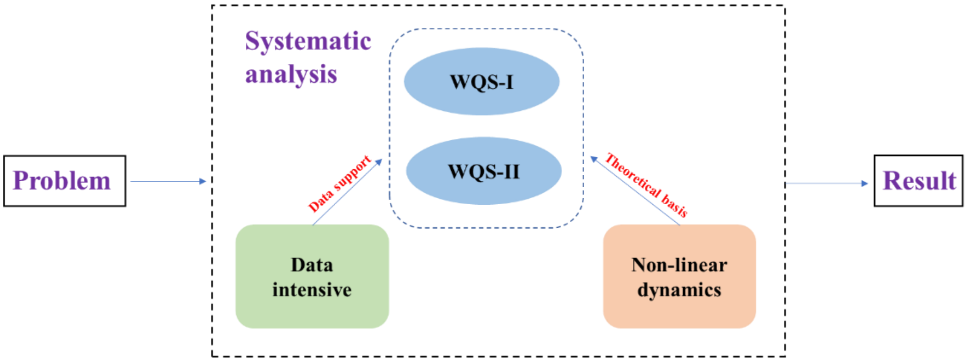

Figure 2. Conceptual framework for river water quality as a nonlinear dynamic system

Based on the viewpoint of systemic analysis, a system would be a nonlinear system if the relationship between input and output or the connection with internal state variables does not satisfy the linear superposition principle. Nonlinear science is a basic discipline that studies the commonality of nonlinear systems, and its research scope is wide, including fractals, periodic oscillations, bifurcation, complexity, chaos, and speckle patterns. After decades of development and application, nonlinear science has involved many fields of natural and social sciences. Complexity, uncertainty, and self-similarity are several important research problems of nonlinear systems. The nonlinear behaviors of evolution in WQS-I and fluctuation in WQS-II correspond to systematic macro-thermodynamics and system dynamics characteristics, respectively, as shown in Figure 2..

III. RESEARCH ON HYDROLOGICAL AND WATER QUALITY NONLINEAR DYNAMIC SYSTEM

The theory and methods of nonlinear hydrological systems, which mainly study the variation in water volume, have produced a lot of research results and played an important role in production practice. Xia gave a complete introduction to system identification theory and methods in 2002 (Xia 2002) and proposed a distributed time-variant gain model expressed by the generalized Volterra functional series (Xia et al. 2004). Sivakumar has conducted extensive research on the chaotic dynamics of hydrological systems (Sivakumar 2000; Sivakumar 2016). Zhang et al. generalized the mechanism of nonlinear hydrological effects (Zhang 2013), which included: (1) impact of watershed organization patterns; (2) complex combinations and interactions of factors within the basin system; (3) interaction between basin and external environment; (4) variation in key control factors at different temporal and spatial scales.

Compared to water quantity simulation research, the development of water quality process modeling has lagged behind. In addition, due to the interdisciplinary, expensive monitoring and analysis of chemical components, there have been relatively few nonlinear studies on water quality systems. The current research on the behavior of nonlinear water quality systems generally includes chaos in the time domain, fractal scaling in the frequency domain, and complex networks in the geography domain. These important nonlinear system behaviors in river water quality systems have a specific space–time span, as shown in Figure 3.

Chaotic characteristics of river water quality time domain "fluctuation" and knowledge gaps : Chaos is an extrinsic, complex, and seemingly irregular motion in a deterministic system due to intrinsic randomness, which is summarized by Lorenz as the so-called "butterfly effect" (Scott 2007; Bradley and Kantz 2015; He 2000; Xu, Liao, and You 2009; Quade et al. 2018). It is an inherent characteristic of nonlinear dynamical systems and is a common phenomenon in nonlinear systems (Oestreicher 2022; Thietart and Forgues 1995). Chaos is a stochastic motion in a deterministic nonlinear dynamical system, where the process is between deterministic and stochastic relations. Chaos theory started in the 1980s and has been widely applied to study various social and natural phenomena (Aguirre and Letellier 2009; Balazs and Voros 1986; Guegan 2009; Olsen and Degn 1985; Skinner et al. 1992).

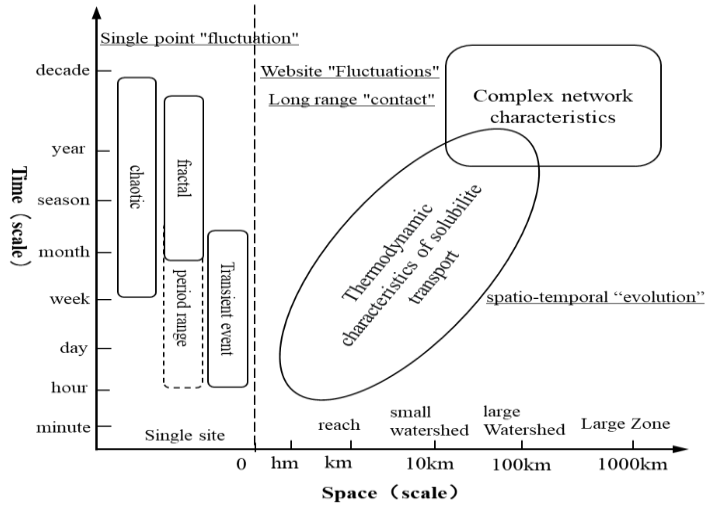

Figure 3. Spatial and temporal scale diagram of nonlinear phenomena in river water quality system

In hydrological studies, the chaotic dynamics behavior of nonlinear processes such as runoff, rainfall, rainfall/runoff ratio, bottom sediment transport, and groundwater dissolved matter transport have been studied more frequently (Sivakumar 2000; Sivakumar 2016). Chaos theory and methods for river water quality systems are rarely reported; at present, the relevant papers mainly focus on chaotic water quality prediction models, and only dissolved oxygen parameters have been studied (Xu et al. 2003; Heddam 2022; Wu, Lu, and Wang 2009). There is no systematic and in-depth analysis of the characteristics and mechanisms of chaotic behavior of river water quality systems based on a big dataset; furthermore, there is a lack of understanding of the manifestation, mechanism, and influencing factors of the chaotic phenomenon of river water quality systems, that is, there is a gray area of theory and knowledge, and there are many questions worth answering, such as the following: In what circumstances are there water quality fluctuations in the process of chaotic phenomena? What is the difference between the chaotic behavior of water quality fluctuations characterized by different water quality parameters and sampling frequency? Are there differences in chaotic characteristics at different stations, and what are the influencing factors?

Non-dynamic systems of a lower order are often reconstructed by the techniques of reconstructing the phase space in research methods to form multi-input, multi-output systems. When a dissipative system reaches an equilibrium state after a long time of evolution, its orbit is attributed to a finite region in the phase space, i.e., the attractor. The chaotic structure of the system can be determined by calculating the correlation dimensions D of the attractor, the Lyapunov exponents, and other characteristic quantities. Most of the studies of chaotic dynamics in hydrology are based on the above framework. The intrinsic determinism of chaotic phenomena allows many stochastic phenomena to be predicted (Sivakumar, Berndtsson, and Persson 2001) such as the Lyapunov exponent method (Zhang 2013). However, most of them are short-term forecast. Recent advances in machine learning and deep learning techniques have brought new approaches to chaotic forecasting and also improved the spatiotemporal scales of forecasting, but the basic forecasting framework is still within the process shown in Figure 4. In 2015, it was reported that a machine learning method based on empirical dynamics simulation established a salmon fishery resource model with chaotic characteristics (Ye et al. 2015), which effectively predicted and analyzed nonlinear phenomena. The mechanism relationship between environment and population biology ignored by the fishery resource model was found. In 2018, Pathak et al. found that the evolution of chaotic systems at a large space–time scale could be successfully predicted by a reservoir computing system (Pathak et al. 2018). After training with the evolution data of the Kuramoto–Sivashinsky equation, the neural network system approximately predicted the evolution of the flame system in the next eight Lyapunov time lengths, which is an eight-fold improvement over the traditional prediction methods. Moreover, the Lyapunov attractor was also successfully reconstructed (Pathak et al. 2017). This conclusion found a breakthrough for developing new models for water quality prediction that could be long-term forecast, which could combine a chaotic system phenomenology index and reservoir computing system to build new data-driven models. In turn, advanced machine learning methods can be used to discover new mechanistic relationships among influencing factors on water quality changes.

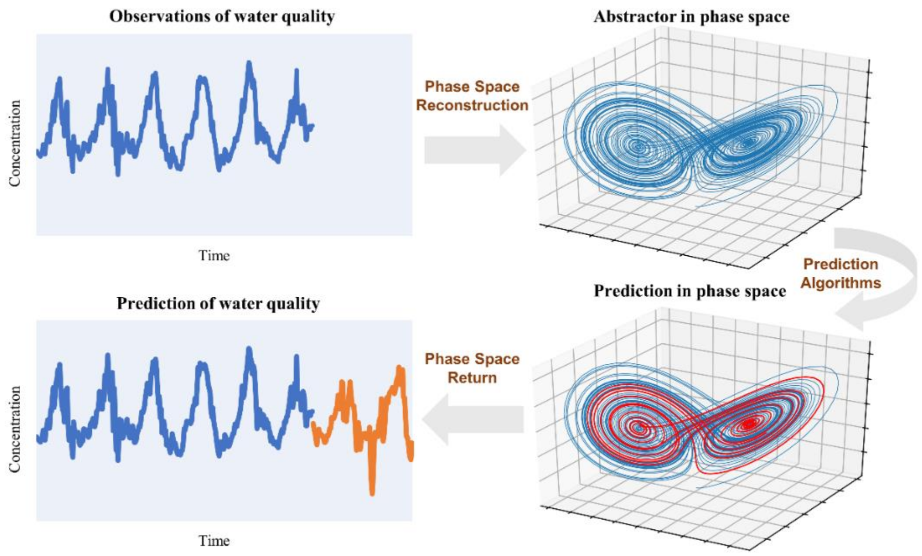

Figure 4. Phase space reconstruction-based traditional framework of water quality prediction in chaos system

Frequency domain analysis of river water quality system fluctuation and knowledge gaps : In the analysis of water quality system fluctuations in the frequency domain, the fractal characteristics of hydrological time series and the identification of transient events are the key points of research. Fractality is a common nonlinear system characteristic in hydrological simulation, which refers to the existence of some form of similarity between the local and the whole of the study object (Aguirre, Viana, and Sanjuán 2009). From the angle of hydrological analysis, the complexity and uncertainty of a hydrological system would be more obvious with an increasing time scale, which undoubtedly brings challenges to hydrological research. Fractal scaling provides a possibility that hydrological representation at different time scales could be inferred from it at a specific time scale. Hurst used rescale range analysis (R/S analysis) to calculate the Hurst index to study the storage capacity of the Aswan High Dam in Egypt in 1951 (Hurst 1951), which first introduced the concept of fractals into the field of hydrology. In 1982, Benoit Mandelbrot infused the self-similarity property of fractals with that of Hurst (Mandelbrot 1982). At present, spectral analysis (Li et al. 2005; Teknik and Ghods 2017), wavelet transform (Rehman and Siddiqi 2009; Sharma, Pachori, and Acharya 2017; Zhu, Jin, and Du 2012), and detrend fluctuation analysis (DFA) (David et al. 2020; Dewandaru et al. 2015) have been used to investigate fractal behavior, as shown in Figure 5. Evidently, river water quality has periodic and seasonal changes under the joint influence of hydrometeorological and human activities. Meanwhile, the fluctuations in some chemical components would show time fractal characteristics in the frequency domain. James and Colin delved into the possible causes of the 1/f time fractality and influence of it on water quality trend detection and solute transport (Kirchner, Feng, and Neal 2000; Kirchner and Neal 2013). Although this work was pioneering, the time resolution of the observation data only reached the accuracy of 7-hour sampling intervals, while other scenarios with large watersheds and tidal river flows were not analyzed.

Moreover, transient events such as precipitation and pollution leakage would lead to non-periodic abnormal fluctuations in the water quality system, as shown in Figure 6. It is of great significance to identify the occurrence and pattern of these transient events. Dohan and Whitfield established the identification of short-term transient events for electrical conductivity and temperature based on 15-minute high-frequency sensor monitoring and wavelet transform (Dohan and Whitfield 1997). The approximate location, duration, and intensity of transient events were identified by separating periodic signals. Jang et al. studied an anomaly detection and warning technology with high-time-resolution measurements, wavelet analysis, and machine learning, which greatly shortened the average detection time and improved the accuracy (Jiang et al. 2020).

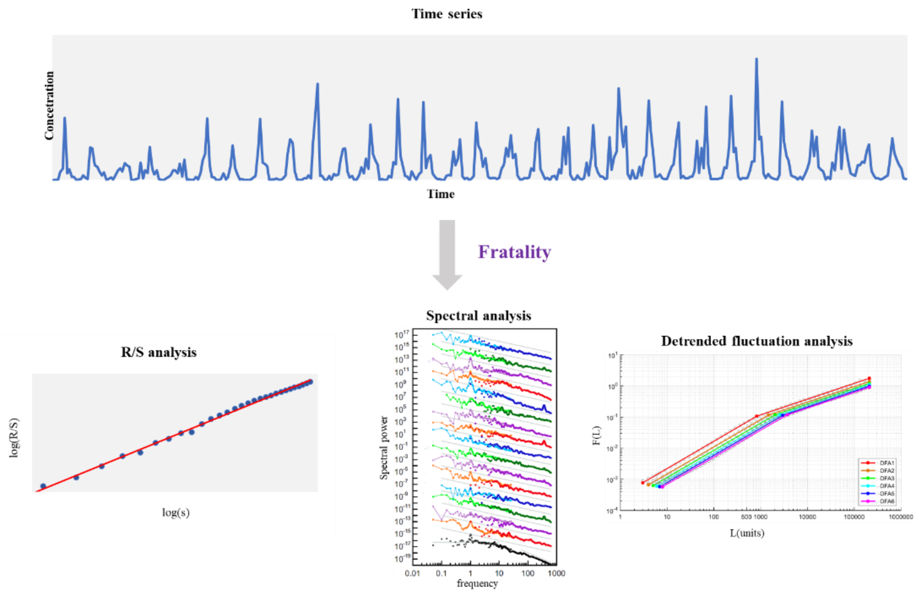

Figure 5. Research methods of time fractal behavior(Kirchner and Neal 2013; Habib et al. 2017)

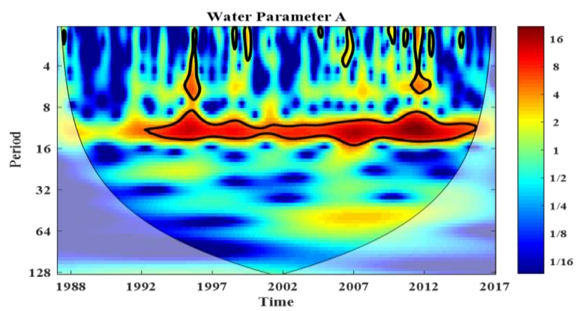

Figure 6. Detection of transient event by wavelet analysis

The fractal behavior and the corresponding driving mechanism of abnormal events are not clear, and future research needs to build models based on the support of high-precision data to capture the information of water quality time series to a greater extent.

The fractal behavior and the corresponding driving mechanism of abnormal events are not clear, and future research needs to build models based on the support of high-precision data to capture the information of water quality time series to a greater extent.

Complex network theory and knowledge gaps for the geographical linkage of river water quality fluctuations : Complexity is the essential characteristic of a nonlinear system of river water quality. In addition to the temporal complexity shown in time, there is also the complexity of spatial linkages. The topography and geomorphology, surface–groundwater interaction, regional meteorology, human activities, and other factors in the spatial complexity of the impact inevitably lead to the existence of a complex correlation (called geographical linkage) of water quality changes in the watershed or regional spatial points, some of which are far apart long-range links. These connections can be described and studied through effective modeling of complex network theory, which has been developed rapidly in recent years.

"Complexity" is relative to "simplicity"; there is no unified understanding of the precise definition of "complexity" in scientific theory and practical application. It can be said that complex things "consist of interconnected or intertwined parts" (Sivakumar and Singh 2012). The extent to which the interrelationships between the components of a system can be explained by existing knowledge, and the number of types of tools that can model these interconnections (Sivakumar 2015), can characterize system complexity. A "network" can be defined as a collection of things linked together, which can be linear or nonlinear, strong or weak, known or unknown, etc. The nature and strength of these connections are critical to the study of the system as a whole.

Since the 1950s, random graph theory has been used to study the statistical properties of some large-scale, complex systems. In 1998, Watts et al. triggered a worldwide wave of complex network research and marked the beginning of complex network research in the modern sense (Watts and Strogatz 1998). Since the 1990s, complex networks have gradually formed a more self-complete discipline, which is even called the "new science of networks" (Albert, Jeong, and Barabási 1999; Barabási and Albert 1999; Watts 2004). Bibliometric analysis showed that the "h-index" of the subject term "complex networks" reached 126.

However, equally surprising is the limited research on complex networks in the field of hydrology and water environments. In 2014, Sivakumar identified the lack of research on complex network phenomena in hydrology (Sivakumar and Singh 2012) and subsequently published several papers expanding the new field of hydrological complex network research (Kirchner, Feng, and Neal 2000; Sivakumar 2016). Correlations between the characteristic parameters of rainfall and runoff complex networks and geographic factors were found, as well as the preference of runoff complex networks for small-world and scale-free networks (Sivakumar 2015; Watts and Strogatz 1998). Some studies have analyzed complex networks of runoff time series (Li et al. 2012) and hydroclimate factors (Liu 2014), but not geographically complex networks. Xiang et al. optimized an urban water quality monitoring network using a Petri network (Xiang, Chen, and Li 2016), and Sitzenfrei et al. (Sitzenfrei et al. 2020) optimized an urban water supply and drainage network based on complex network science. Gayer et al. (Gayer et al. 2021) used complex network models to assess the impact of anthropogenic elements on river water quality. Zischg et al. (Zischg et al. 2019) used complex network science to evaluate the stability of a water distribution network. Simone et al. (Simone et al. 2022) used a complex network to accurately monitor the vulnerability and monitoring design of urban drainage networks, pollutants, and the spread of pathogens.

The research on complex networks is almost blank in the analysis of river water quality systems, and there are only reports on complex networks of water blooms in lakes and reservoirs, which are an urgent need of research. Jiang et al. presented preliminary results at the EGU2022 meeting (Jiang et al. 2022). Therefore, many questions are worth exploring and answering, including the following: Are complex network phenomena prevalent in river water quality systems? What kind of complex network patterns do they tend to have, and what are their driving mechanisms? Are there differences in complex network characteristics of water quality systems under different water quality parameters and time resolutions? How can network information be used to guide water quality modeling and management? (Figure 7) In addition, what is the intrinsic connection between chaos, fractals, and complex networks of water quality fluctuations? How can the phenomenological feature index be integrated for water quality prediction, monitoring network optimization, and other practical applications? These are deep questions that need to be studied.

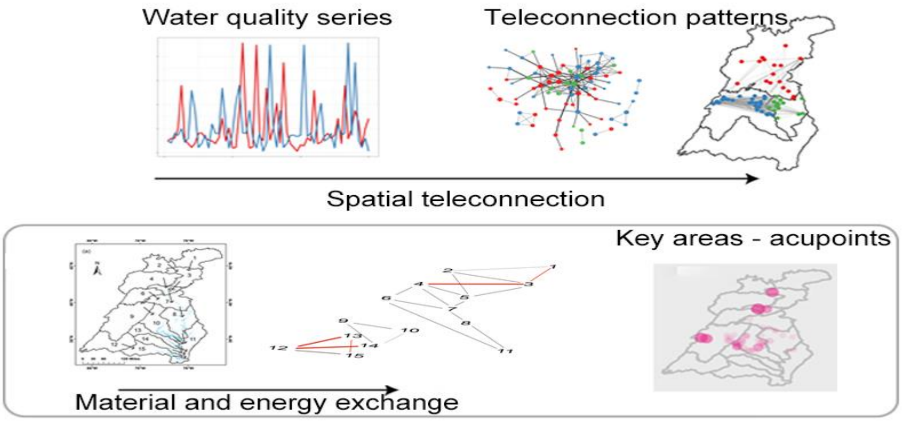

Figure 7. Schematic diagram of regional complex network

IV. INFORMATION DOMAIN CHARACTERIZATION IN THE "EVOLUTIONARY" THERMODYNAMICS OF RIVER WATER QUALITY SYSTEMS

Corresponding to the dynamics behavior of the time-series change of the water quality system is the thermodynamic law of the overall state evolution. As a water quality system is a nonlinear system far from the equilibrium state, the diffusion and reaction processes of material components are irreversible, and in the evolution process, they may appear in space and can be directly observed in the self-organization, pattern, chaos, and other critical phenomena. By analyzing the evolution law of the transport process of water quality components, the key time and space nodes can be extracted and the pollution process can be effectively monitored. Statistical information theory provides an effective tool for the characterization and analysis of this non-equilibrium thermodynamic behavior, as shown in Figure 8.

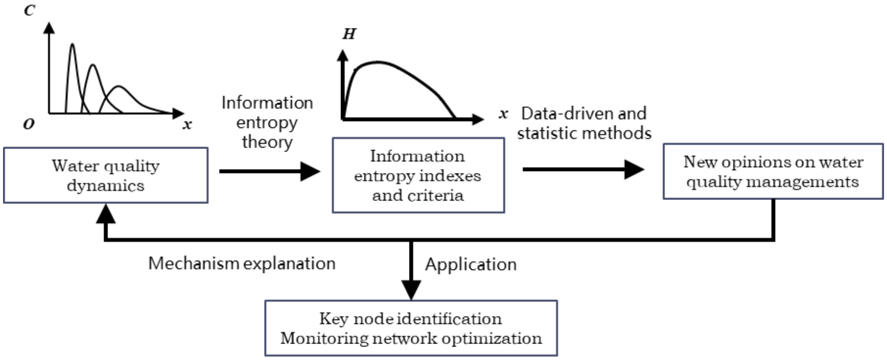

Figure 8. Schematic representation of the information domain

Information theory was first proposed by Shannon in 1948 in the study of communication systems (Shannon 1948), and one of the important concepts is "information entropy" as a measure of the degree of uncertainty. The term "information entropy" was borrowed from thermodynamics by Shannon, and it was the first time information was expressed in a quantitative way, which attracted widespread attention in academic circles. At present, information entropy theory has been widely used in communication, image processing, bioinformatics, hydrology, statistics, and other engineering and scientific fields (Wang and Zhu 2001). V.P. Singh conducted a more systematic study on the application of information entropy in hydrological and hydraulic engineering (Singh 2014).

The analysis of the convective diffusion process (ADR process, i.e., the first type of water quality system in this study) of pollutants in water bodies using information entropy started in 1983. Senf (1983) deduced that the "functional information" in a one-dimensional diffusion model increases to the maximum with time, that is, the "critical time" needed for an equilibrium state by using information entropy (Senf 1983). In 1994, Kitanidis of Stanford University proposed the concept of the dilution index "E" as a quantitative description of the concentration information entropy in the ARD process and analyzed the variation law of the dilution index in the Gaussian diffusion process (Kitanidis 1994). In 2017, Chiognan and Rolle of the Technical University of Munich combined the ideas of Senf and Kitanidis and further proposed the concept of the "critical reaction time" (Chiogna and Rolle 2017). The deformation, mixing, and reaction kinetics of contaminant clusters between spiral flows in three-dimensional porous media were analyzed by using the dilution index (Hazas et al. 2022; Ye et al. 2020). In recent years, Jiang et al. discovered the spatial variation in information entropy in ADR systems and proposed the phenomenological index of the critical location to monitor network design (Shi et al. 2018). The critical location index is used to monitor the network design. For the study of multiphase flow systems, some researchers adopted a multivariate multiscale increment entropy (MMIE)-based method (Wang and Jin 2020). The results show that MMIE can effectively reveal the dynamic complexity of different flow types and their evolutionary behavior with changes in the flow regime, indicating the advantage of information entropy tools in describing problems such as complex systems' characteristics and evolutionary processes.

At present, the research on the non-equilibrium-state thermodynamics of the evolution of the first type of water quality system under information entropy characterization (called the "information domain") is in the initial stage (Aghakouchak 2014; Alfonso, Lobbrecht, and Price 2010; Keum et al. 2017; Singh 1997); although some considerable research results have been obtained, there are still a series of issues that need to be elucidated. The characteristics of the critical reaction time and critical position of the ADR process under different flow regimes and different inputs and constraints of signal sources (pollution sources) have not been fully clarified. The research on the spatiotemporal variation in information entropy in the source, activation, and migration process of water quality components in complex rivers is still insufficient. Information flow characteristics, whether they depend on the material flow and energy flow, whether they are as constant as the material flow and energy flow, and whether they can be standardized or are the inverse of the source information, still require further research. Based on the diverse information entropy index, whether it can establish a bridge towards water environment management and provide corresponding criteria for water environment management decisions under various situations is a topic worthy of research. Can the possible pattern phenomenon of the river water quality system be observed under the information entropy perspective (Kirchner thinks it is unlikely to appear as a critical phenomenon(Kirchner and Neal 2013))? The question of how to directly learn the information entropy evolutionary features of water quality systems through big data and the latest physics-informed deep learning techniques (Raissi, Perdikaris, and Karniadakis 2019) is also worth exploring.

V. FUTURE RESEARCH

Data-intensive science is closely related to phenomenological theory in physics. Famous physical scientist Yang Z (Yang 1998) divides physics into three paths: experiment, phenomenological theory, and theoretical framework. Phenomenological theory lies between experimental observation and theoretical framework. Phenomenology states that the physical law can be obtained by generalizing experimental facts rather than their internal causes when explaining physical phenomena. The relation of phenomenological theory to experiment and theory is shown in Figure 9.

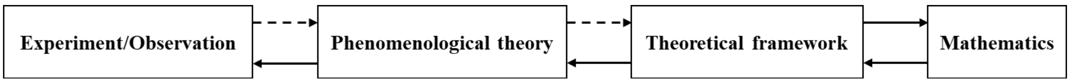

Figure 9. The relation of phenomenological theory to experiment and theory

Drawing on the above methodological system, future research on this topic could be explored by relying on the emerging research path: data-intensive analysis–phenomenological law presentation–mechanism interpretation and application. Starting from the analytical methods of phenomenology and system theory, a new framework for water quality system analysis should be developed and established. There are many valuable studies to be carried out in the future from the four aspects of dynamics, thermodynamics, mechanism analysis, and management method development.

Law of dynamics: Nonlinear behavior of river water quality "fluctuation": For normal fluctuations in water quality at a single site, the chaos phenomenon that may appear in the time series of typical water quality parameters such as dissolved oxygen, nitrate nitrogen, turbidity, conductance, and chloride ions under different time resolutions could be revealed through phase space reconstruction, and the law of characteristic quantities such as the phase space structure, correlation only, and Lyapunov index could be explored. It is important to explore the periodic characteristics and scale-free phenomena that typical water quality parameters may present in the frequency domain under a high time resolution and find the fractal scale index characteristics of typical parameters in a typical watershed.

For abnormal water quality fluctuations at a single station caused by short-term events such as rainfall, point source discharge, and blooms, the frequency-domain characteristics of typical water quality parameters under abnormal events could be studied by wavelet transform to identify different abnormal modes.For the potential "long-range" connections between water quality stations in the basin and regions, the complex network characteristics of river water quality systems in the region could be analyzed using the network topological feature index, and the possible "small-world model" and other complex network models in the transport of dissolved matter could be explored based on the correlation of water quality fluctuations at the stations.

Laws of thermodynamics: Nonlinear behavior of river water quality "evolution": The basic framework of information-domain river water quality system analysis needs to be investigated, including the probability space definition, statistical sampling method, and system boundary definition. For the transport process of a single component, the spatial and temporal evolution of information entropy under different input characteristics of information sources could be analyzed, and the analytical expressions and numerical solutions of thermodynamic state indices such as the dilution index, critical migration time, and critical migration distance defined by information entropy need to be derived.

For the complex water chemical reactions of nitrogen and phosphorus nutrients in rivers, the spatiotemporal variation in information entropy in the water component source and activation and migration processes needs to be investigated. Meanwhile, the data of river tracer tests need to be collected on a large scale. After that, the rule of information entropy of tracer tests could be calculated directly through data statistics to verify the theoretical solution.

AI: Driving Mechanism analysis of nonlinear features and feature reconstruction based on Machine learning : For the characteristics of chaotic nonlinear dynamics and scale-free phenomena of single stations, the possible mechanisms of chaotic behavior and scale-free phenomena need to be analyzed from human activities and natural factors, and the conditions and thresholds of the above nonlinear characteristics need to be explored qualitatively and quantitatively by combining the mechanism model of a typical watershed and big data analysis.

For the complex network characteristics of regional water quality connection in a typical watershed, the physical transmission process of the complex network model of typical water quality needs to be analyzed, and the evolutionary dynamic mechanism of the network needs to be analyzed in combination with the hydrological and water quality model of the basin. The sources, transport routes, and fluxes of river nutrients could be analyzed based on high-frequency observation. Moreover, it is necessary to explore the possible correlation among chaotic features in the time domain, fractal features in the frequency domain, and regional complex network features of each station in the basin, and use information entropy to explore the causal relationship of nonlinear features.

It is worth noting that the nonlinear dynamic characteristic indices such as the chaotic Lyapunov index and scale-free index need to be learned from water quality time series by using the reserve pool calculation system. The evolution of information entropy under the control of nonlinear partial differential equations needs to be studied using physics-informed deep learning technology.

New management methodology: Development of water environment modeling and management based on phenomenological feature index : Based on the characteristic rule and mechanism of the "time–frequency–geo-information domain" of the above water quality system, innovative application research on water environment monitoring, modeling, and management could be carried out around typical river basins. This includes new models of water quality prediction and parameter estimation based on a chaotic characteristic index and advanced deep learning technology, new methods of quantitative design of monitoring networks based on information entropy and a complex network index, new classification systems of river water environment systems based on a nonlinear characteristic index, and new models of nutrient migration law analysis based on complex networks.

VI. CONCLUSION

In the field of hydrology, nonlinear system theory mainly takes the variation in water quantity as the research object; investigations of river water quality are relatively few. This review summarizes the research progress of chaotic behavior in the time domain, fractal behavior in the frequency domain, complex network behavior in the geographic space, and information-domain characterization of river water quality as a nonlinear system. For chaotic behavior, most studies tend to predict future behavior with big data and models, while the characteristic and mechanism of chaos are not deeply analyzed. Since the driving mechanism of time fractals is still unclear, it is necessary to establish a model with high-precision data to capture the information of water quality time series to a greater extent. Similarly, many gaps remain in complex networks. The question of whether complex networks widely exist in river quality systems, and of how to use complex networks to instruct water quality modeling and management, needs to be solved. Moreover, the research of the information domain is in the preliminary stage; it is a mainstream direction to directly learn the evolution characteristics of it for water quality systems through deep learning technology and big data.

ACKNOWLEDGEMENT

This research was supported by funding provided by the Science, Technology and Innovation Commission of Shenzhen Municipality (grant no. JSGG20201103094600001), and the Research Project of Hydrology and Water Quality Center of Shenzhen (grant no. SWSZ-2022-0017).We are grateful to the constructive comments by Prof. Jiping Jiang.

REFERENCE

[1] Aghakouchak, Amir. 2014. 'Entropy–copula in hydrology and climatology', *J. Hydrometeorol.*, 15: 2176-89.

[2] Aguirre, Jacobo, Ricardo L Viana, and Miguel AF Sanjuán. 2009. 'Fractal structures in nonlinear dynamics', *Rev. Mod. Phys.*, 81: 333.

[3] Aguirre, Luis A, and Christophe Letellier. 2009. 'Modeling nonlinear dynamics and chaos: a review', *Math. Probl. Eng.*, 2009.

[4] Albert, Réka, Hawoong Jeong, and Albert-László Barabási. 1999. 'Diameter of the world-wide web', *Nature*, 401: 130-31.

[5] Alfonso, Leonardo, Arnold Lobbrecht, and Roland Price. 2010. 'Optimization of water level monitoring network in polder systems using information theory', *Water Resour. Res.*, 46.

[6] Balazs, NL, and A Voros. 1986. 'Chaos on the pseudosphere', *Phys. Rep.*, 143: 109-240.

[7] Barabási, Albert-László, and Réka Albert. 1999. 'Emergence of scaling in random networks', *science*, 286: 509-12.

[8] Bradley, Elizabeth, and Holger Kantz. 2015. 'Nonlinear time-series analysis revisited', *Chaos*, 25:

097610.
[9] Chen, CL Philip, and Chun-Yang Zhang. 2014. 'Data-intensive applications, challenges, techniques and technologies: A survey on Big Data', *Inf. Sci.*, 275: 314-47.
[10] Chin, David A. 2008. 'Phenomenological models of hydrologic processes in south Florida', *J. Hydrol.*, 349: 230-43.
[11] Chiogna, Gabriele, and Massimo Rolle. 2017. 'Entropy-based critical reaction time for mixing-controlled reactive transport', *Water Resour. Res.*, 53: 7488-98.
[12] David, Sérgio Adriani, CMC Inácio Jr, Derick David Quintino, and José António Tenreiro Machado. 2020. 'Measuring the Brazilian ethanol and gasoline market efficiency using DFA-Hurst and fractal dimension', *Energy Econ.*, 85: 104614.
[13] Dewandaru, Ginanjar, Rumi Masih, Obiyathulla Ismath Bacha, and A Mansur M Masih. 2015. 'Developing trading strategies based on fractal finance: An application of MF-DFA in the context of Islamic equities', *Physica A.*, 438: 223-35.
[14] Dohan, Kathleen, and Paul H Whitfield. 1997. 'Identification and characterization of water quality transients using wavelet analysis. I. Wavelet analysis methodology', *Water Sci. Technol.*, 36: 325-35.
[15] Drazin, Philip G, and Philip Drazin Drazin. 1992. *Nonlinear systems* (Cambridge University Press: Cambridge, UK).
[16] Gayer, Fernanda Almeida Marchini, Dejanira de Franceschi de Angelis, Andre Franceschi de Angelis, and Elaine Cristina Catapani Poletti. 2021. 'Use of Complex Network Modelling to Assess the Influence of the Parameters on Water Quality of Rivers', *Water Air Soil Pollut.*, 232: 1-11.
[17] Gorton, Ian, Paul Greenfield, Alex Szalay, and Roy Williams. 2008. 'Data-intensive computing in the 21st century', *Computer*, 41: 30-32.
[18] Govindaraju, Rao S. 2000. 'Artificial neural networks in hydrology. I: Preliminary concepts', *J. Hydrol. Eng.*, 5: 115-23.
[19] Guegan, Dominique. 2009. 'Chaos in economics and finance', *Annu Rev. Control.*, 33: 89-93.
[20] Habib, Abrar, James PR Sorensen, John P Bloomfield, Katie Muchan, Andrew J Newell, and Adrian P Butler. 2017. 'Temporal scaling phenomena in groundwater-floodplain systems using robust detrended fluctuation analysis', *J. Hydrol.*, 549: 715-30.
[21] Hazas, Mónica Basilio, Francesca Ziliotto, Massimo Rolle, and Gabriele Chiogna. 2022. 'Linking mixing and flow topology in porous media: An experimental proof', *Phys. Rev.* , 105: 035105.
[22] He, Ji-Huan. 2000. 'A review on some new recently developed nonlinear analytical techniques', *Int. J. Nonlinear Sci. Numer. Simul.*, 1: 51-70.
[23] Heddam, Salim. 2022. 'Parallel Chaos Search Based Incremental Extreme Learning Machine Based Empirical Wavelet Transform: A New Hybrid Machine Learning Model for River Dissolved Oxygen Forecasting.' in, *Computational Intelligence for Water and Environmental Sciences* (Springer: Singapore).
[24] Hey, Anthony JG, Stewart Tansley, and Kristin Michele Tolle. 2009. *The fourth paradigm: data-intensive scientific discovery* (Microsoft research: Redmond, USA).
[25] Hochachka, Wesley M, Daniel Fink, Rebecca A Hutchinson, Daniel Sheldon, Weng-Keen Wong, and Steve Kelling. 2012. 'Data-intensive science applied to broad-scale citizen science', *Trends Ecol. Evol.*, 27: 130-37.
[26] Hogle, Linda F. 2016. 'Data-intensive resourcing in healthcare', *BioSocieties*, 11: 372-93.
[27] Hurst, Harold Edwin. 1951. 'Long-term storage capacity of reservoirs', *Transactions of the American society of civil engineers*, 116: 770-99.
[28] Jiang, jiping, Sijie Tang, Bellie Sivakumar, Tianrui Pang, Na Wu, and Yi Zheng. 2022. 'Teleconnection Patterns of River Water Quality Dynamics Based on Complex Network Analysis'.
[29] Jiang, Jiping, Yi Zheng, Tianrui Pang, Baoyu Wang, Ritik Chachan, and Yu Tian. 2020. 'A comprehensive study on spectral analysis and anomaly detection of river water quality dynamics with high time resolution measurements', *J. Hydrol.*, 589: 125175.
[30] Kelling, Steve, Wesley M Hochachka, Daniel Fink, Mirek Riedewald, Rich Caruana, Grant Ballard, and Giles Hooker. 2009. 'Data-intensive science: a new paradigm for biodiversity studies', *BioScience*, 59: 613-20.
[31] Keum, Jongho, Kurt C Kornelsen, James M Leach, and Paulin Coulibaly. 2017. 'Entropy applications to water monitoring network design: a review', *Entropy*, 19: 613.
[32] Kim, Taereem, Tiantian Yang, Shang Gao, Lujun Zhang, Ziyu Ding, Xin Wen, Jonathan J Gourley, and Yang Hong. 2021. 'Can artificial intelligence and data-driven machine learning models match or even replace process-driven hydrologic models for streamflow simulation?: A case study of four watersheds with different hydro-climatic regions across the CONUS', *J.Hydrol.*, 598: 126423.
[33] Kirchner, James W. 2009. 'Catchments as simple dynamical systems: Catchment characterization,

rainfall-runoff modeling, and doing hydrology backward', *Water Resour. Res.*, 45.
[34] Kirchner, James W, Xiahong Feng, and Colin Neal. 2000. 'Fractal stream chemistry and its implications for contaminant transport in catchments', *Nature*, 403: 524-27.
[35] Kirchner, James W, and Colin Neal. 2013. 'Universal fractal scaling in stream chemistry and its implications for solute transport and water quality trend detection', *Proc. Natl. Acad. Sci.*, 110: 12213-18.
[36] Kitanidis, Peter K. 1994. 'The concept of the dilution index', *Water Resour. Res.*, 30: 2011-26.
[37] Kouzes, Richard T, Gordon A Anderson, Stephen T Elbert, Ian Gorton, and Deborah K Gracio. 2009. 'The changing paradigm of data-intensive computing', *Computer*, 42: 26-34.
[38] Kundzewicz, Zbigniew W, and JAROSŁAW J Napiórkowski. 1986. 'Nonlinear models of dynamic hydrology', *Hydrolog. Sci. J.*, 31: 163-85.
[39] Lange, H, and S Sippel. 2020. 'Machine learning applications in hydrology.' in, *Forest-water interactions* (Springer: Cham, Switzerland).
[40] Li, Xiaoli, J Polygiannakis, P Kapiris, A Peratzakis, K Eftaxias, and Xin Yao. 2005. 'Fractal spectral analysis of pre-epileptic seizures in terms of criticality', *J. Neural. Eng.*, 2: 11.
[41] Li, Xinjie;, Tiesong; Hu, Xuning; Guo, Xiang; Zeng, and Tao Zhang. 2012. 'Application of 0-1 test algorithm to chaos to runoff time series', *Adv. Water Sci.*, 23: 861-68.
[42] Liu, Zuhan. 2014. 'Complexity and nonlinearity of climate-hydrological processes in Tarim River Basin', East China Normal University.
[43] Mandelbrot, Benoit B. 1982. *The fractal geometry of nature* (W. H. Freeman and Company: USA).
[44] Michener, William K, and Matthew B Jones. 2012. 'Ecoinformatics: supporting ecology as a data-intensive science', *Trends Ecol. Evol.*, 27: 85-93.
[45] Newman, Harvey B, Mark H Ellisman, and John A Orcutt. 2003. 'Data-intensive e-science frontier research', *Commun. ACM.*, 46: 68-77.
[46] Oestreicher, Christian. 2022. 'A history of chaos theory', *Dialogues Clin. Neurosci.*
[47] Olsen, Lars Folke, and Hans Degn. 1985. 'Chaos in biological systems', *Q. Rev. Biophys.*, 18: 165-225.
[48] Pathak, Jaideep, Brian Hunt, Michelle Girvan, Zhixin Lu, and Edward Ott. 2018. 'Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach', *Phys. Rev. Lett.*, 120: 024102.
[49] Pathak, Jaideep, Zhixin Lu, Brian R Hunt, Michelle Girvan, and Edward Ott. 2017. 'Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data', *Chaos*, 27: 121102.
[50] Peters-Lidard, Christa D, Martyn Clark, Luis Samaniego, Niko EC Verhoest, Tim Van Emmerik, Remko Uijlenhoet, Kevin Achieng, Trenton E Franz, and Ross Woods. 2017. 'Scaling, similarity, and the fourth paradigm for hydrology', *Hydrol. Earth Syst. Sci.*, 21: 3701-13.
[51] Powers, Susan E, Linda M Abriola, Joyce S Dunkin, and Walter J Weber Jr. 1994. 'Phenomenological models for transient NAPL-water mass-transfer processes', *J. Contam. Hydrol.*, 16: 1-33.
[52] Quade, Markus, Markus Abel, J Nathan Kutz, and Steven L Brunton. 2018. 'Sparse identification of nonlinear dynamics for rapid model recovery', *Chaos*, 28: 063116.
[53] Raissi, Maziar, Paris Perdikaris, and George E Karniadakis. 2019. 'Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations', *J. Comput. Phys.*, 378: 686-707.
[54] Rehman, S, and AH Siddiqi. 2009. 'Wavelet based Hurst exponent and fractal dimensional analysis of Saudi climatic dynamics', *Chaos*, 40: 1081-90.
[55] Scott, Alwyn. 2007. *The nonlinear universe: chaos, emergence, life* (Springer: Berlin, German).
[56] Senf, Lothar. 1983. 'Diffusion and information', *Int. J. Heat Mass Transfer.*, 26: 1710-12.
[57] Shannon, Claude Elwood. 1948. 'A mathematical theory of communication', *Bell Syst. Tech. J.*, 27: 379-423.
[58] Sharma, Manish, Ram Bilas Pachori, and U Rajendra Acharya. 2017. 'A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension', *Pattern. Recognit. Lett.*, 94: 172-79.
[59] Shi, Bin, Jiping Jiang, Bellie Sivakumar, Yi Zheng, and Peng Wang. 2018. 'Quantitative design of emergency monitoring network for river chemical spills based on discrete entropy theory', *Water Res.*, 134: 140-52.
[60] Simone, A, A Cesaro, G Del Giudice, C Di Cristo, and O Fecarotta. 2022. 'Potentialities of Complex Network Theory Tools for Urban Drainage Networks Analysis', *Water Resour. Res.*, 58: e2022WR032277.
[61] Singh, Vijay P. 2014. "Entropy theory in hydraulic engineering: an introduction." In.: American Society of Civil Engineers.
[62] Singh, VP. 1997. 'The use of entropy in hydrology and water resources', *Hydrol. Process.*, 11: 587-626.
[63] Sitzenfrei, Robert, Qi Wang, Zoran Kapelan, and Dragan Savić. 2020. 'Using complex network analysis

for optimization of water distribution networks', *Water Resour. Res.*, 56: e2020WR027929.

[64] Sivakumar, B. 2000. 'Chaos theory in hydrology: important issues and interpretations', *J. Hydrol.*, 227: 1-20.

[65] Sivakumar, B, and VP Singh. 2012. 'Hydrologic system complexity and nonlinear dynamic concepts for a catchment classification framework', *Hydrol. Earth Syst. Sci.*, 16: 4119-31.

[66] Sivakumar, Bellie. 2015. 'Networks: a generic theory for hydrology?', *Stoch Env Res. Risk A.*, 29: 761-71.

[67] ———. 2016. *Chaos in hydrology: bridging determinism and stochasticity* (Springer: Dordrecht, Netherlands).

[68] Sivakumar, Bellie, Ronny Berndtsson, and Magnus Persson. 2001. 'Monthly runoff prediction using phase space reconstruction', *Hydrol. Sci. J.*, 46: 377-87.

[69] Skinner, James E, Mark Molnar, Tomas Vybiral, and Mirna Mitra. 1992. 'Application of chaos theory to biology and medicine', *Integr Psychol. Behav Sci.*, 27: 39-53.

[70] Teknik, Vahid, and Abdolreza Ghods. 2017. 'Depth of magnetic basement in Iran based on fractal spectral analysis of aeromagnetic data', *Geophys. J. Int.*, 209: 1878-91.

[71] Thietart, Raymond Alain, and Bernard Forgues. 1995. 'Chaos theory and organization', *Organ. Sci.*, 6: 19-31.

[72] Wang, Dayang, and Ningde Jin. 2020. 'Multivariate multiscale increment entropy: a complexity measure for detecting flow pattern transition in multiphase flows', *Nonlinear Dyn.*, 100: 3853-65.

[73] Wang, Dong, and Yuanshen Zhu. 2001. 'A review of the application of information entropy to water systems', *Hydrology*, 21: 9-14.

[74] Watts, Duncan J. 2004. 'The" new" science of networks', *Annu Rev. Polit Sci.*: 243-70.

[75] Watts, Duncan J, and Steven H Strogatz. 1998. 'Collective dynamics of 'small-world'networks', *Nature*, 393: 440-42.

[76] Wu, Jun, Jian Lu, and Jiaquan Wang. 2009. 'Application of chaos and fractal models to water quality time series prediction', *Environ Modell. Softw.*, 24: 632-36.

[77] Xia, Jun. 2002. *Theory and methods of hydrologic nonlinear systems* (Wuhan University Press: Wuhan, China).

[78] Xia, Jun, Gangsheng Wang, Ge Tan, Aizhong Ye, and Guohe Huang. 2004. 'Hydrologic nonlinear systems and distributed time variant gain models', *Sci. China.*, 34: 1062-71.

[79] Xiang, Yangxue;, Xuesong; Chen, and Jia Li. 2016. 'Urban water environmental quality monitoring: a study based on complex network model', *China Pop. Resour. Env.*: 328-32.

[80] Xu, Hang, Shi-Jun Liao, and Xiang-Cheng You. 2009. 'Analysis of nonlinear fractional partial differential equations with the homotopy analysis method', *Commun. Nonlinear Sci. Numer Simul.*, 14: 1152-56.

[81] Xu, Min, Guangming Zeng, Gengxin Xie, Guohe Huang, and Xiaokang Su. 2003. 'Preliminary research on the application of chaos theory to dissolved oxygen prediction', *Acta Sci. Circumst.*, 23: 776-80.

[82] Xu, Zuxin, and Zhenliang Liao. 2003. 'Development stages and spatial levels of water quality mathematical model research', *Environ Sci. Shanghai.*, 22: 79-85.

[83] Yang, Zhenning. 1998. 'Beauty and Physics', *J. Zhejiang Univ.(Humanit. soc. sci.)*: 1-8.

[84] Ye, Hao, Richard J Beamish, Sarah M Glaser, Sue CH Grant, Chih-hao Hsieh, Laura J Richards, Jon T Schnute, and George Sugihara. 2015. 'Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling', *Proc. Natl. Acad. Sci.*, 112: E1569-E76.

[85] Ye, Yu, Gabriele Chiogna, Chunhui Lu, and Massimo Rolle. 2020. 'Plume deformation, mixing, and reaction kinetics: An analysis of interacting helical flows in three-dimensional porous media', *Phys. Rev.*, 102: 013110.

[86] Zhang, Donghui, Jincun Zhang, and Fanggui Liu. 2007. 'Some comments on nonlinear effect in catchment hydrology', *Adv. Water Sci.*, 18: 776-84.

[87] Zhang, Yong. 2013. 'New prediction of chaotic time series based on local Lyapunov exponent', *Chinese. Phys. B.*, 22: 050502.

[88] Zhu, Yonghua, Xinqiao Jin, and Zhimin Du. 2012. 'Fault diagnosis for sensors in air handling unit based on neural network pre-processed by wavelet and fractal', *Energy. Build.*, 44: 7-16.

[89] Zischg, Jonatan, Julian D Reyes-Silva, Christopher Klinkhamer, Elisabeth Krueger, Peter Krebs, P Suresh C Rao, and Robert Sitzenfrei. 2019. "Complex network analysis of water distribution systems in their dual representation using isolation valve information." In *World Environmental and Water Resources Congress 2019: Hydraulics, Waterways, and Water Distribution Systems Analysis*, 484-97. American Society of Civil Engineers Reston, VA.