26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Complexity of river water quality parameters using a nonlinear dynamic approach

Sakshi Dhumale[a], **Bellie Sivakumar**[a]

[a]*Indian Institute of Technology Bombay, Powai, Mumbai, Maharashtra, India*
Email: 23d0305@iitb.ac.in

Abstract: Riverine water quality parameters exhibit highly variable spatiotemporal behaviour due to the nonlinear interplay of watershed processes and interdependencies among parameters. Understanding this complexity is crucial, as it governs the predictability of water quality and influences the choice of modelling approaches. This study examines the dynamics of multiple river water quality parameters using a nonlinear dynamic approach. Specifically, the False Nearest Neighbour (FNN) method is applied to determine the dimensionality of water quality time series and assess its complexity. Single-variable time series of water quality parameters are reconstructed into a multi-dimensional phase space using delay embedding. Monthly data of nine water quality parameters from 56 monitoring sites across rivers in New Zealand are analysed. These parameters include dissolved oxygen (DO), dissolved reactive phosphorus (DRP), electrical conductivity (EC), pH, instantaneous discharge (Q-Inst), temperature (Temp), total nitrogen (TN), total phosphorus (TP), and turbidity (Turb). The results indicate that the dimensionality of the parameters ranges from 3 to 10, with pH and electrical conductivity showing lower complexity, while total phosphorus and turbidity exhibit higher complexity. Parameters such as temperature, DO, DRP, Q-Inst, and TN generally display low dimensionality but exhibit higher dimensionality at certain sites. This parameter-specific complexity assessment at the 56 stations helps optimise monitoring design, choose appropriate modelling approaches, and prioritise management efforts.

Keywords: River water quality, Nonlinear dynamics, Dimensionality, False Nearest Neighbours, New Zealand

1. INTRODUCTION

Water quality plays a vital role in sustaining ecosystem health, supporting human livelihoods, and enabling effective water resources management. River water quality is shaped by hydrologic, anthropogenic, and climatic processes (Van Vliet et al. 2013, 2023), and is widely recognised as a key indicator of watershed health. Water quality exhibits substantial variability, driven by nonlinear interactions among physical, chemical, and biological processes, as well as complex feedback mechanisms within watershed systems.

The variability and dynamics of water quality parameters in rivers vary both in space and in time. Spatial variability is driven by landscape heterogeneity, including land use, land cover, soils, topography, geology, climate, and catchment hydrology, as well as the spatial arrangement of source areas across the landscape (Lintern et al. 2018). Temporal variability, on the other hand, is strongly governed by hydrological and climatic dynamics, with same-day streamflow having been identified as the dominant driver of short-term fluctuations, while soil moisture, antecedent flow, vegetation, and water temperature also play important roles (Guo et al. 2019). Given these diverse influences, understanding water quality variability is essential for sustainable catchment management, early-warning systems, and pollution control.

Analysis of water quality dynamics in aquatic systems has mainly been approached from two modelling paradigms: mechanistic models and data-driven models. Mechanistic models, such as SWAT, WASP, MIKE 11, and QUAL2E, simulate water quality processes based on established physical, chemical, and biological principles and are widely applied in river basin management (Bai et al. 2022). In contrast, data-driven models identify statistical patterns directly from observed data. Traditional data-driven approaches, such as linear regression, multiple linear regression and ARIMA, are suited for capturing linear dependencies, whereas more advanced machine learning techniques, including neural networks, support vector regression, and random forests, are capable of representing nonlinear relationships in complex systems (Cojbasic et al. 2023).

The nonlinear and complex behaviour of water quality parameters in rivers can also be examined from the perspective of chaos theory. In recent decades, chaos theory has also emerged as a promising data-driven approach for investigating nonlinear and sensitive behaviours in environmental systems, offering an alternative means of analysing complexity beyond conventional modelling approaches. In has already found extensive applications to rainfall, river flow, and sediment transport (see Sivakumar 2017 for details), along with some notable recent advancements in pollutant dynamics in wastewater treatment plants (Ramkumar and Jothiprakash 2024), rainfall downscaling from General Circulation Model (GCM) outputs (Deepthi and Sivakumar 2024), and catchment classification (Vignesh et al. 2015; Khan and Sivakumar 2024). These studies have applied a variety of chaos identification and prediction methods, including phase space reconstruction, correlation dimension, false nearest neighbour algorithm, Lyapunov exponent method, Kolmogorov entropy method, surrogate data method, Poincaré maps, close returns plots, and nonlinear local approximation prediction. Despite the significant progress in the application of chaos theory in environmental systems, only very limited research has focused on their application to water quality in rivers (Wu et al. 2009; Jiang et al. 2021).

The present study addresses this gap. Specifically, the study addresses the complexity of water quality parameters in rivers from the perspective of dimensionality. The false nearest neighbour (FNN) method is applied to time series of each of nine riverine water quality parameters at 56 monitoring stations across New Zealand. These parameters include dissolved oxygen (DO), dissolved reactive phosphorus (DRP), electrical conductivity (EC), pH, instantaneous discharge (Q-Inst), temperature (Temp), total nitrogen (TN), total phosphorus (TP), and turbidity (Turb). For each station, the single-variable water quality time series is reconstructed in a multi-dimensional phase space, and the complexity is identified based on the optimal embedding dimension obtained using the FNN method.

2. METHODOLOGY

2.1. False Nearest Neighbour Method

The dimensionality of a time series represents its variability and, hence, the complexity; in general, the higher the dimension, the higher the variability and complexity. The False Nearest Neighbour (FNN) method, proposed by (Kennel et al. 1992), is a widely used nonlinear dynamic technique for determining the dimensionality of time series. The FNN method identifies points that appear close in lower-dimensional embeddings but separate significantly when the embedding dimension is increased. By eliminating such “false” neighbours, the method ensures that the reconstructed phase space more accurately represents the true dynamics of the system. The methodology involves the following steps:

1. **Phase Space Reconstruction:** A phase space is a multidimensional space where each axis represents a state variable, allowing visualisation of the system's evolution. Following Takens' theorem of delay embedding (Takens 1981), the phase space is reconstructed from a univariate time series using delay embedding, as follows:

$$\boldsymbol{Y}_j = \left(X_j, X_{\mathrm{j}+\tau}, X_{\mathrm{j}+2\tau}, \dots, X_{\mathrm{j}+(m-1)\tau}\right) \tag{1}$$

   where τ is the time delay and m are the embedding dimension.
2. **Nearest Neighbour Identification:** For each state vector $\boldsymbol{Y}_j$, its nearest neighbour $\mathbf{Y}_j^{\mathrm{NN}}$ is identified in the reconstructed space of dimension m using the Euclidean distance.
3. **False Neighbour Criteria:** The next step is to assess whether the nearest neighbours in dimension m remain close when moving to the higher dimension m+1. If a neighbour moves significantly farther away in the higher dimension, it is classified as a "false nearest neighbour". This indicates that the current dimension is insufficient to capture the true dynamics. A nearest neighbour is considered as false if it fails to satisfy either the distance criterion or the loneliness criterion, as detailed in Kennel et al. (1992).
4. The percentage of false neighbours (%FNN) is computed for different values of m. At lower dimensions, %FNN is generally high but gradually declines at the higher dimensions. The optimal embedding dimension (m_{opt}) is defined as the value of m at which %FNN falls to (or close to) zero. In real-world hydrological and environmental datasets, the presence of noise often prevents %FNN from reaching exactly zero. In such cases, the m corresponding to the minimum %FNN is adopted, and when multiple m yields the same minimum, the smallest is chosen as m_{opt}.

3. STUDY AREA AND DATA

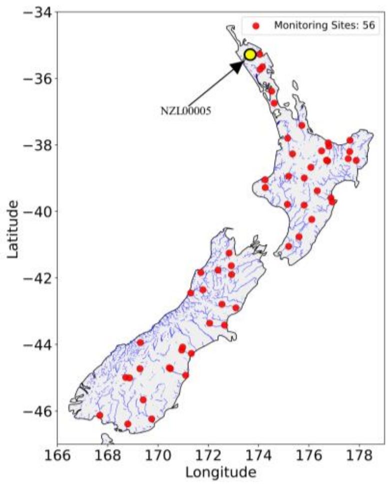

Figure 1 Map of the study area showing all 56 monitoring sites, with representative site highlighted in yellow

This study focuses on river water quality dynamics across 56 monitoring sites distributed throughout New Zealand. The locations of the sites are shown in Figure 1. A total of nine key water quality parameters are considered, encompassing physiochemical, nutrient, and hydrological attributes, namely DO, DRP, EC, pH, Q-Inst, Temp, TN, TP, and Turb. Together, these parameters provide a comprehensive representation of the processes governing riverine systems.

The dataset consists of monthly time series spanning a 21-year period (2000–2020), thereby capturing both seasonal variability and long-term trends in water quality. All data are obtained from the Global Environment Monitoring System for Water (GEMStat) portal (https://gemstat.org/data-gemstat/data-portal/, Accessed: 27 June 2025), which compiles quality-controlled monitoring records from global freshwater systems. Figure 2(a) to (i) presents the time series plots for all the nine parameters for a representative station (GEMStat Station Number: NZL00005) (see Figure 1), which is considered for illustration of the analysis and results. As seen, the nine parameters exhibit distinct temporal variability, with some parameters (e.g., DO and Temp) showing clear seasonal cycles, while others (e.g., DRP, TN, TP, Turb, and Q-Inst) display irregular fluctuations and abrupt peaks, reflecting significant variability among themselves.

4. RESULTS

For the phase space reconstruction and subsequent FNN analysis, the embedding dimension is varied from 1 to 10 and a delay time of 1 is used. Figure 3(a) to (i) presents the two-dimensional (i.e., m = 2) phase space plots for the nine water quality parameters from Station NZL00005. As shown in Figure 3, for most of the parameters (excluding DO and temperature), the reconstructed trajectories appear broad and widely scattered across the phase space. This reflects a more complex and potentially random nature of the dynamics. In contrast, the phase space projections of DO (Figure 3(a)) and temperature (Figure 3(f)) reveal relatively structured and well-defined attractors. The trajectories remain confined within specific regions, indicating a comparatively simple and deterministic structure. While phase space reconstruction provides qualitative insights into the dynamic patterns and complexity of the system, a quantitative measure is required to determine the optimal dimension necessary to adequately describe the complexity.

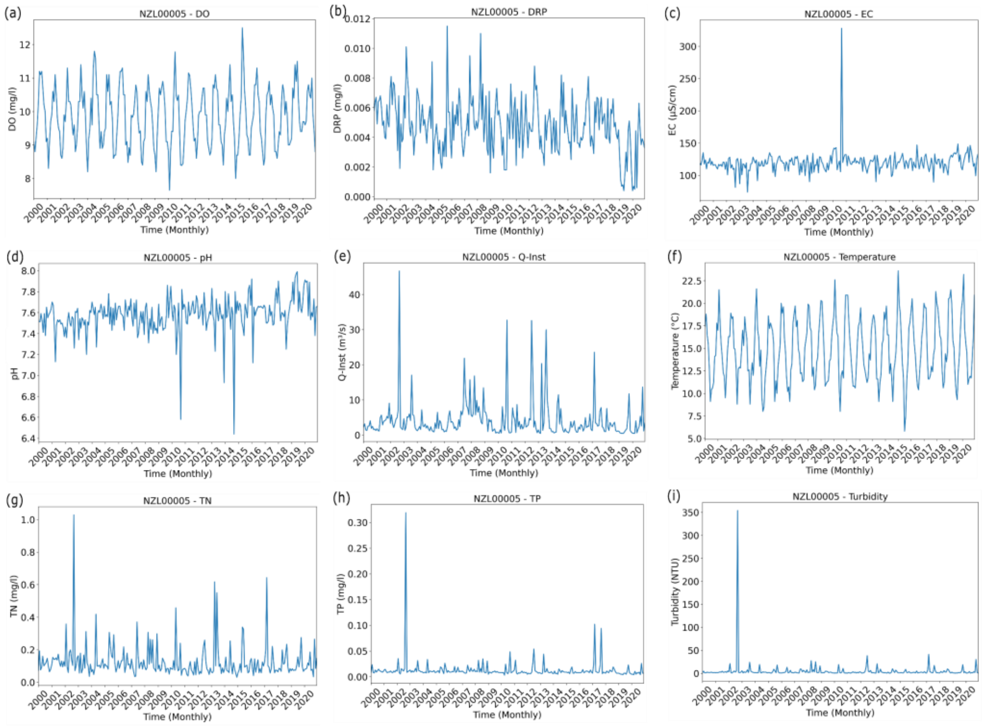

Figure 2 Time series plots of nine water quality parameters for a representative site (No. NZL00005)

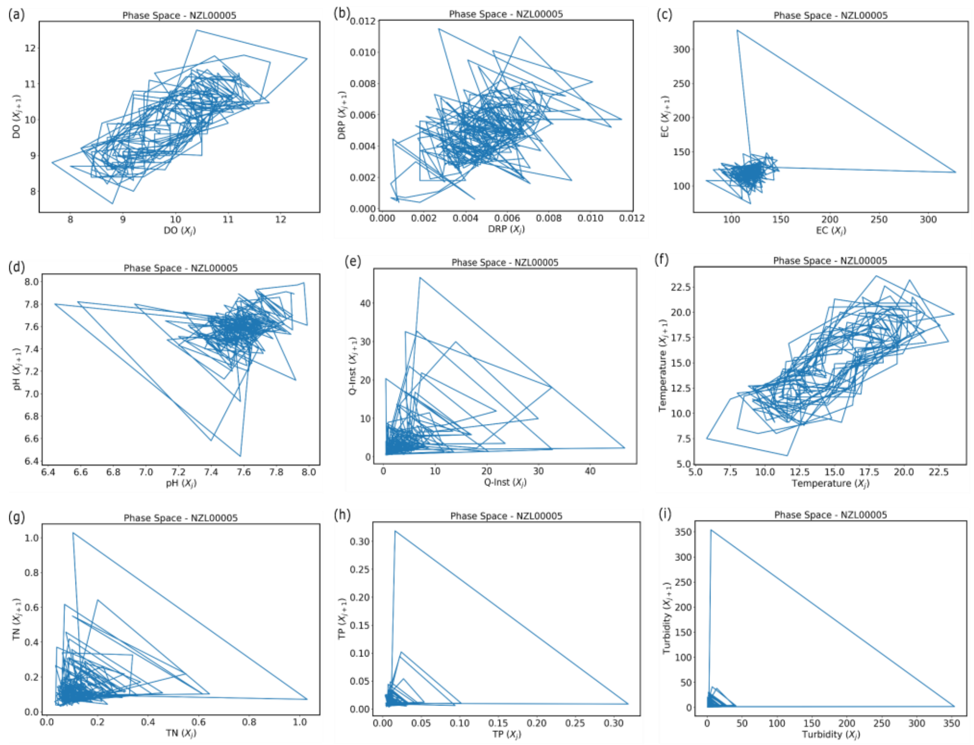

Figure 3 Phase space plots of nine water quality parameters for a representative site (No. NZL00005).

Figure 4 presents the variation of %FNN with embedding dimension for the nine water quality parameters from Station NZL0005. As seen in the figure, for each of the water quality time series, with an increase in embedding dimension, the %FNN gradually decreases and eventually attains a minimum value at certain embedding dimension (with either no increase or some increase beyond that point). For DO and temperature, the %FNN drops sharply and reaches zero at $m = 4$ and $m = 5$, respectively, after which it remains close to zero for higher dimensions.

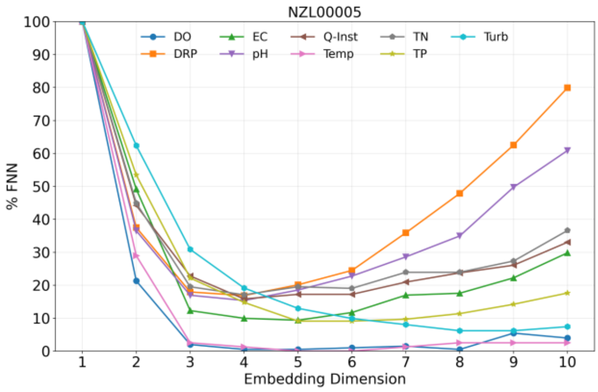

Figure 4 FNN plot for a representative site showing multiple water quality parameters, with each parameter depicted in a different colour

In contrast, for the remaining parameters (e.g., EC, DRP, pH, Q-Inst, TN, TP, Turb), the %FNN decreases with increasing embedding dimension but does not attain zero. Instead, it stabilises at a nonzero minimum before showing a gradual increase at higher embedding dimensions.

While the above results show useful insights into the qualitative and quantitative behaviour of individual parameters, it is equally important to investigate how these dynamics vary at each site across the study area. To this end, the optimal dimensions obtained from the FNN analysis are mapped for all 56 sites for each parameter. A total of nine spatial maps are generated, corresponding to the nine water quality parameters, as shown in Figure 5(a) to (i).

Overall, most of the stations exhibit relatively low embedding dimensions (predominantly m_{opt} in the range 3–5), indicative of low-dimensional complexity. The ranges of each parameter for all 56 stations are as follows: DO (3–10), DRP (3–6), EC (3–5), pH (3–5), Q-Inst (3–10), Temperature (3–8), TN (3–8), TP (3–10), and Turbidity (4–9). In particular, EC (Figure 5(b)) and pH (Figure 5(d)) demonstrate consistently low m_{opt} values across the entire monitoring network, with all sites falling within the range of m_{opt} = 3–5. This suggests that their complexity is dominantly governed by a small number of variables. However, DO, DRP, Q-Inst, and TN also display low embedding dimensions at the majority of sites, but a small subset of sites reveals markedly higher values ($m_{opt} > 5$), indicating more complex system dynamics in those specific locations. Notably, TP and turbidity stand out, with a large number of sites exhibiting higher m_{opt} values, suggesting the influence of a greater number of governing variables on their temporal behaviour.

5. DISCUSSION AND CONCLUSIONS

The nonlinear dynamic analyses of water quality parameters across multiple sites provide valuable insights into the complexity and underlying dynamics of river systems. Phase space reconstruction revealed marked differences among the parameters, with DO and Temp exhibiting well-structured trajectories, whereas TP and Turb showed broader and more scattered trajectories. The FNN analysis further quantified these findings by identifying the optimal embedding dimensions, showing that parameters such as pH and EC are governed by relatively few variables. Also, DO, DRP, Q-Inst, and TN revealed low embedding dimensions at most sites but with a few sites exhibiting higher values, reflecting spatial heterogeneity and localised complexity in their governing processes. Notably, TP and turbidity consistently required higher embedding dimensions across a larger number of sites, pointing to the dominance of a higher number of dominantly governing variables in shaping their temporal variability. Such insights are critical for improving predictive modelling, as they emphasise the need for parameter-specific and site-specific approaches that integrate nonlinear dynamics and complex system behaviours in river water quality assessment and management.

Although the FNN method provides insights into system dimensionality and complexity, it does not identify the specific parameters driving dynamics. Traditional approaches, such as correlation analysis, factor analysis, and PCA, focus on linear associations and may overlook nonlinear or causal influences. Multivariate phase space reconstruction and other nonlinear dynamic techniques offer a more rigorous alternative by jointly embedding multiple parameters to capture their nonlinear and potentially chaotic interactions. Therefore, future research should incorporate multivariate phase space reconstruction together with nonlinear techniques, such as mutual information, transfer entropy, or causal discovery methods, to identify the true drivers of water quality dynamics. By combining insights on system dimensionality with causal variable selection, it will be possible to develop more robust predictive models that enhance understanding and support improved management of river water quality systems.

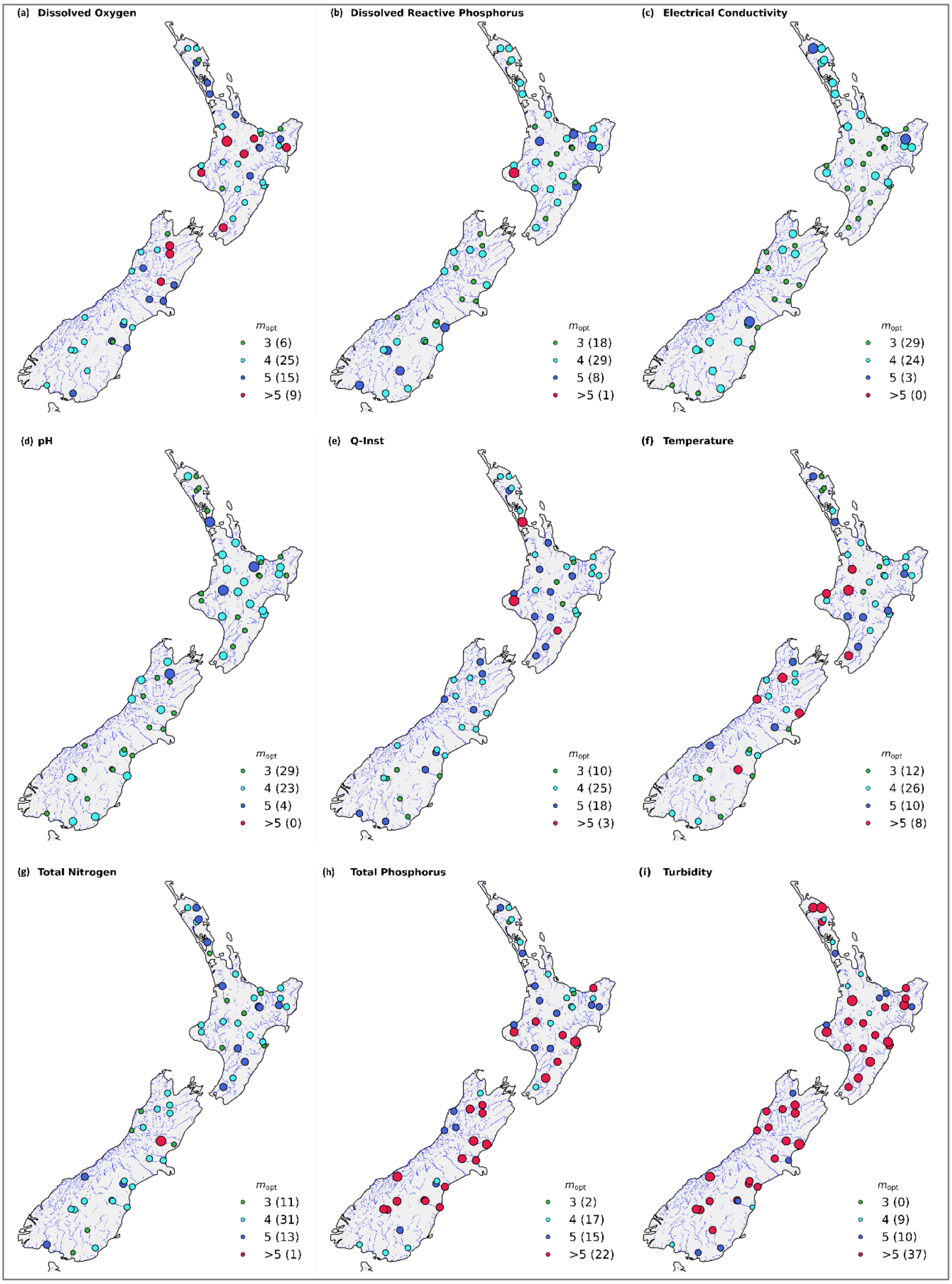

Figure 5 Optimal embedding dimension for nine water quality parameters across monitoring sites in New Zealand: (a) dissolved oxygen, (b) dissolved reactive phosphorus, (c) electrical conductivity, (d) pH, (e) discharge (Q-Inst), (f) temperature, (g) total nitrogen, (h) total phosphorus, and (i) turbidity. Coloured circles represent the m_{opt} values at each site, with counts of sites in each category shown in the bracket.

REFERENCES

Bai J, Zhao J, Zhang Z and Tian Z (2022) Assessment and a review of research on surface water quality modelling. Ecological Modelling, 466:109888.

Cojbasic S, Dmitrasinovic S, Kostic M, Turk Sekulic M, Radonic J, Dodig A and Stojkovic M (2023) Application of machine learning in river water quality management: a review. Water Science & Technology, 88(9):2297–2308.Deepthi B and Sivakumar B (2024) Spatial downscaling of rainfall from general circulation models: A chaos theory-based framework. Atmospheric Research, 308:107531.

Guo D, Lintern A, Webb JA, Ryu D, Liu S, Bende-Michl U, Leahy P, Wilson P and Western AW (2019) Key Factors Affecting Temporal Variability in Stream Water Quality. Water Resources Research, 55(1):112–129.

Jiang J, Tang S, Liu R, Sivakumar B, Wu X and Pang T (2021) A hybrid wavelet-Lyapunov exponent model for river water quality forecast. Journal of Hydroinformatics, 23(4):864–878.

Kennel MB, Brown R and Abarbanel HDI (1992) Determining embedding dimension for phase space reconstruction using a geometrical construction. Physical Review A, 45(6):3403–3411.

Khan SK and Sivakumar B (2024) A Nonlinear Local Approximation Approach for Catchment Classification. Entropy, 26(3):218.

Lintern A, Webb JA, Ryu D, Liu S, Bende-Michl U, Waters D, Leahy P, Wilson P and Western AW (2018) Key factors influencing differences in stream water quality across space. WIREs Water, 5(1): e1260.

Ramkumar D and Jothiprakash V (2024) A chaotic investigation on pollutant parameters of a wastewater treatment facility using false nearest neighbour algorithm. Stochastic Environmental Research and Risk Assessment, 38(1):1–16.

Sivakumar B (2017) Chaos in hydrology: Bridging Determinism and Stochasticity. Springer.

Takens F (1981) Detecting strange attractors in turbulence. In : Rand and L-S Young (Eds), Dynamical Systems and Turbulence, Warwick 1980, Lecture Notes in Mathematics, Springer Berlin Heidelberg, Berlin, Heidelberg.

Van Vliet MTH, Franssen WHP, Yearsley JR, Ludwig F, Haddeland I, Lettenmaier DP and Kabat P (2013) Global river discharge and water temperature under climate change. Global Environmental Change, 23(2):450–464.

Van Vliet MTH, Thorslund J, Strokal M, Hofstra N, Flörke M, Ehalt Macedo H, Nkwasa A, Tang T, Kaushal SS, Kumar R, Van Griensven A, Bouwman L and Mosley LM (2023) Global river water quality under climate change and hydroclimatic extremes. Nature Reviews Earth & Environment, 4(10):687–702.

Vignesh R, Jothiprakash V and Sivakumar B (2015) Streamflow variability and classification using false nearest neighbour method. Journal of Hydrology, 531:706–715.

Wu J, Lu J and Wang J (2009) Application of chaos and fractal models to water quality time series prediction. Environmental Modelling & Software, 24(5):632–636.

