Environ Monit Assess (2025) 197:1198
https://doi.org/10.1007/s10661-025-14644-9

RESEARCH

Spatio-temporal patterns of river water quality in the klang river basin, malaysia: a functional data analysis approach to detect pre- and post-pandemic shifts

Nur Fatihah Mohd Ali · Ibrahim Mohamed · Rossita Mohamad Yunus · Faridah Othman

Received: 30 June 2025 / Accepted: 26 September 2025

Abstract Understanding spatial and temporal patterns of river water quality over a multi-year period is crucial for effective basin management and pollution control. This study applies functional data analysis (FDA) to evaluate monthly water quality index (WQI) data from 16 monitoring stations across the Klang River Basin, Malaysia, covering the period from 2020 to 2023, which spans both pre- and post-pandemic conditions. By treating water quality index (WQI) measurements as smooth functions over time, FDA captures underlying trends and variations that are not readily detected using classical statistical techniques. Functional principal component analysis (FPCA) reveals that the first component accounts for 97% of the total variation, reflecting the dominant pattern in water quality over time, which is characterized by relatively stable upstream conditions and gradual deterioration downstream. The second and third components capture seasonal fluctuations and short-term disturbances, potentially linked to monsoonal cycles and shifts in human activities during the pandemic. Functional clustering based on FPCA scores groups stations according to their temporal behavior, distinguishing upstream areas with stable conditions from downstream areas experiencing greater variability. Spatial interpretation of these clusters offers additional insight into localized pollution sources and environmental stressors. Compared to classical PCA, FDA provides a more detailed, curve-based understanding of time-dependent and location-specific changes in water quality. The result underscore the value of FDA in environmental monitoring, particularly for detecting pre- and post-pandemic shifts, and support its application in guiding adaptive and spatially targeted management strategies for river basins.

Keywords Functional principal component analysis · Spatio-temporal clustering · Water quality trends · Monsoonal influence · River basin monitoring

Introduction

River water quality is vital for environmental health by sustaining biodiversity within aquatic ecosystems.

N. F. M. Ali (✉)
Department of Computational and Theoretical and Sciences, Kulliyyah of Science, International Islamic University Malaysia, 25200 Kuantan, Malaysia
e-mail: nurfatihahmohdali@iium.edu.my

I. Mohamed · R. M. Yunus
Institute of Mathematical Sciences, Faculty of Science, Universiti Malaya, 50603 Kuala Lumpur, Malaysia
e-mail: imohamed@um.edu.my

R. M. Yunus
e-mail: rossita@um.edu.my

F. Othman
Department of Civil Engineering, Faculty of Engineering, Universiti Malaya, 50603 Kuala Lumpur, Malaysia
e-mail: faridahothman@um.edu.my

Published online: 11 October 2025

However, rivers face significant threats from diverse pollution sources, including the food and beverage, chemical, semiconductor, and electronics industries (Hampel et al., 2015; United Nations Environment Programme 2008). Effective river system management requires continuous monitoring of water quality parameters to identify trends, assess degradation, and implement targeted conservation strategies (Azha et al., 2023). In Malaysia, the number of manual and automatic water quality monitoring stations increased from 1064 in 2006 to 1383 in 2024 enabling more comprehensive monitoring of river health across various areas (Loi et al., 2022). Continuous river monitoring is essential for identifying areas of water quality degradation and understanding their sources (Vega-Rodríguez et al. 2021).

Variations in river water quality from upstream to downstream areas are influenced by both natural processes and anthropogenic activities (Anh et al., 2023; Zhao et al., 2020). Upstream areas typically have cleaner water originating from forested catchments, while downstream sections often experience declining water quality due to runoff, industrial discharge, and wastewater effluents (Hamid et al., 2020). These spatial differences highlight the interconnectedness of river ecosystems and underscore the importance of location-specific monitoring. In rapidly urbanizing countries like Malaysia, residential, commercial, and industrial development significantly impacts water quality (Mokhtar, 2023), while agricultural runoff introduces pesticides, fertilizers, and sediments. Understanding spatial patterns helps identify pollution hotspots and guides tailored mitigation strategies. Meanwhile seasonal dynamics, which are often influenced by environmental conditions and hydrological cycles, significantly affect variations in water quality (Ling et al., 2017). In the Klang River basin, the Northeast Monsoon (November–March) results in heavy rainfall and increased river flow, while the Southwest Monsoon (May–September) is marked by drier conditions, reduced flow, higher pollutant concentrations, and stagnant conditions that deplete oxygen levels (Islam et al., 2015; Ismail et al., 2023; Nienie et al., 2017; Sidek et al., 2016). These monsoonal patterns introduce temporal variability that must be considered in water quality assessment.

The combination of spatial and seasonal variation highlights the complexity of river water quality dynamics. Classical methods for assessing water quality typically rely on statistical analysis of discrete measurements collected at fixed time points, often from a limited number of stations. Techniques such as correlation, regression, and principal component analysis (PCA) have been widely used to identify pollution sources and assess spatial variability (Barroso et al., 2024; Mohamed et al., 2015; Mohammed et al., 2022; Shan et al., 2021). While these methods provide valuable insights, they fall short in capturing the continuous and dynamic nature of water quality processes (Schreiber et al., 2022). For example, studies by Othman et al. (2012) and Nasir et al. (2011) used regression and cluster analysis to relate water quality to urban and industrial activities, but were unable to fully explain how pollutant levels fluctuate over time due to weather or seasonal shifts.

Functional data analysis (FDA) provides a more flexible alternative by treating water quality measurements as continuous functions over time. As demonstrated by Ramsay and Silverman (2005), FDA enables the construction of smooth functional representations from discrete observations, offering a more natural way to model environmental processes and uncover hidden patterns. FDA also allows for the comparison of temporal trends across variables and stations (Di Blasi et al., 2013; Muñiz et al., 2012; Yan et al., 2015). Most prior studies have applied classical statistical approaches such as PCA, regression models, or clustering to water quality data (Mohammed et al., 2022; Schreiber et al., 2022), but these approaches treat data points as isolated, often overlooking temporal dependence, seasonal variability, or irregular sampling. As a result, important features such as lag effects, cumulative changes, or seasonal dynamics may go undetected. In contrast, FDA accommodates site-specific trends and offers a more holistic view of water quality dynamics. Functional clustering, for example, groups stations with similar temporal profiles, revealing shared pollution patterns and facilitating targeted interventions (Gong et al., 2021; Mohd Ali et al. 2024). Studies by Ariza et al. (2023), Haggarty et al. (2015), and Kande et al. (2024) have shown that functional clustering improves the interpretation of complex environmental datasets by identifying consistent site characteristics over time.

Functional principal component analysis (FPCA) is a key tool within the FDA framework. FPCA decomposes functional data into orthogonal components that capture dominant modes of variation (Dona et al., 2009; Gong et al., 2015; Karuppusami et al., 2022). Several studies have used FPCA to analyze temporal patterns in river water quality. For instance, Henderson (2006) employed FPCA to process water quality data from multiple monitoring stations along a river, revealing spatial patterns and identifying areas with significant pollution. Similarly, Embling et al. (2012) applied FPCA to study seasonal water quality management strategies, highlighting the method's utility in environmental contexts. This study presents the first comprehensive application of FDA to river water quality data in the Klang River basin, Malaysia. By combining functional modeling with spatial interpretation, this study contributes a novel approach for understanding river water quality dynamics and supports adaptive monitoring and management strategies under changing environmental conditions. This study focuses on the Klang River basin in Malaysia, utilizing monthly water quality data collected from 16 monitoring stations between 2020 and 2023. The main aims are to transform discrete water quality data into smooth temporal functions using FDA, to apply FPCA to identify dominant temporal variation patterns and to cluster stations with similar water quality behavior and investigate seasonal and spatial influences on these patterns.

Materials and methods

Study area and dataset

The Klang River basin is located in the central Peninsular Malaysia and encompasses the states of Selangor and the Federal Territory of Kuala Lumpur. It drains from Ulu Gombak to the river mouth in Port Klang and covers about 1288 square kilometers of catchment area. The basin consists of the main Klang River and 11 tributaries, including Sg. Gombak, Sg. Kerayong, Sg. Penchala, and Sg. Damansara. Water quality in this basin is influenced by diverse land uses, ranging from protected upstream forests to densely urbanized downstream areas. Figure 1 shows the

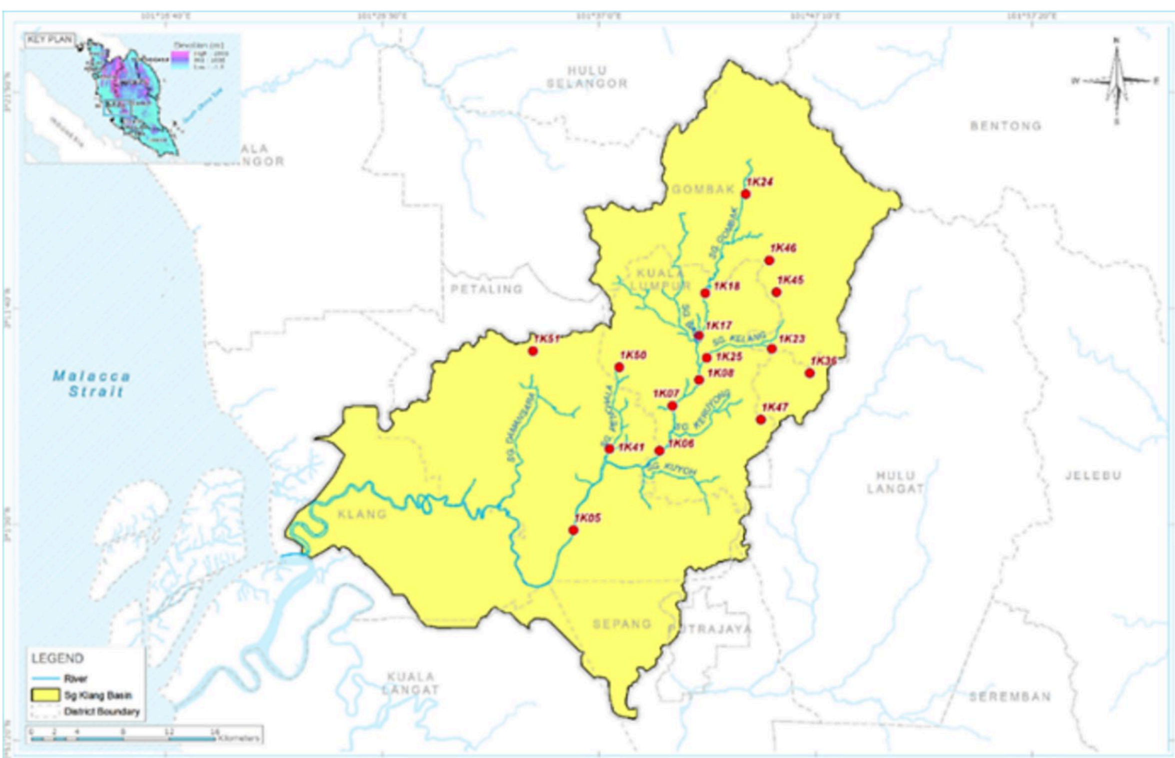

Fig. 1 Klang River basin map with water quality stations

water quality monitoring stations located within the Klang River basin. Most of the monitoring stations are located in the middle stream of the river basin. There are a total of 16 monitoring stations, which are Stations 1K05, 1K06, 1K07, 1K08, 1K25, 1K45, and 1K46 located along Sg. Klang. Stations 1K17, 1K18 and 1K24 are located along Sg. Gombak while Station 1K23 and 1K36 are situated along Sg. Ampang. Stations 1K41 and 1K50 are located along Sg. Penchala, while Stations 1K47 and 1K51 are located along Sg. Kerayong and Sg. Damansara respectively. The data considered consist of the monthly recorded water quality status along the Klang River basin from January 2020 to December 2023.

The water quality parameters measured at each station include dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), total suspended solids (TSS), ammoniacal nitrogen (NH3NL), temperature, and pH. These parameters are important to assess the quality status of river water and are combined into the Water Quality Index (WQI) model developed by the Department of Environment (DOE), Malaysia. The WQI is determined by calculating a weighted sum of six sub-indices for DO, BOD, COD, NH3NL, TSS, and pH, using the official formulation provided by DOE, as shown in Eq. 1.

Table 1 Sub-index calculations

Sub-index parameter	Value	Conditions
SIDO	0	$DO \leq 8$
	100	$DO \geq 92$
	$-0.395 + 0.030DO^2 - 0.00020DO^3$	$8 < DO < 92$
SIBOD	$100.4–4.23BOD$	$BOD \leq 5$
	$108e^{(-0.055BOD)} - 0.1BOD$	$BOD > 5$
SICOD	$-1.33COD + 99.1$	$COD \leq 20$
	$103e^{(-0.01575COD)} - 0.04COD$	$COD > 20$
SIAN	$100.5–105AN$	$AN \leq 0.3$
	$94e^{-0.573AN} - 5\ \lvert AN - 2\rvert$	$0.3 < AN < 4$
	0	$AN \geq 4$
SISS	$97.5e^{-0.00676SS} + 0.05SS$	$SS \leq 100$
	$71e^{-0.0016SS} - 0.015SS$	$100 < SS < 1000$
	0	$SS \geq 1000$
SIpH	$17.2–17.2pH + 5.02pH^2$	$pH < 5.5$
	$-242 + 95.5pH - 6.67pH^2$	$5.5 \leq pH < 7$
	$-181 + 82.4pH - 6.605pH^2$	$7 \leq pH < 8.75$
	$536–77.0pH + 2.76pH^2$	$pH \geq 8.75$

$$WQI = 0.22SIDO + 0.19SIBOD + 0.16SICOD + 0.16SISS + 0.15SIAN + 0.12SipH \tag{1}$$

Each sub-index is estimated using DOE's best-fit equations given in Table 1 and the final WQI ranges between 0 and 100. The determination of WQI for each location also allows for categorization based on the National Water Quality Standard (NWQS) (Department of Environment (DOE), 2019).

Based on DOE guidelines, WQI scores are grouped into three categories which are clean ($80 < WQI \leq 100$), slightly polluted ($60 \leq WQI \leq 80$), and polluted ($0 \leq WQI < 60$). The WQI is further classified into five major classes: Class I ($92.7 < WQI \leq 100.0$), Class II ($76.5 < WQI \leq 92.7$), Class III ($51.9 < WQI \leq 76.5$), Class IV ($31.0 < WQI \leq 51.9$), and Class V ($0 < WQI \leq 31.0$). Practically, no treatments are needed for Class I as the water is very clean and safe for direct drinking. Water in Class II requires only a conventional treatment, while Class III water can still be used for livestock drinking. However, water in Class IV is suitable only for irrigation, and Class V is considered polluted and cannot be used for any of the purposes listed in other classes. While widely used, this WQI has limitations as it does not include other important pollutants such as nutrients, heavy metals, or microbial indicators, and reducing water quality to a single index may overlook site-specific pollution issues and does not fully reflect ecological healths (Wong et al., 2020). All data used in this study were obtained from the Department of Environment, Malaysia.

Functional data transformation

The discrete water quality index data can be transformed into smooth functions or curves using smoothing techniques (Ramsay & Silverman, 2005). Let the discrete observations of water quality index data be denoted by $x_i(t_j)$, where $i = 1, \ldots, n$ represents each monitoring station located at spatial coordinates s_i and $j = 1, \ldots, T$ represents each observation time point. For each station i, the set of observation $x_i(t_j)$ is transformed into a smooth function or curve denoted by $X(s_i, t)$, defined on a continuous time domain $t \in [1, T]$, using a smoothing technique based on Fourier basis functions. To this end, suppose that the observations are fitted using the regression model:

$$x_i(t_j) = y_i(t_j) + \varepsilon_{ij}, i = 1, \ldots, n, j = 1, \ldots, T \tag{2}$$

where ε_{ij} denotes the random error. The smooth functions $y_i(t)$ are represented as linear combinations of K independent basis functions $\phi_k(t)$:

$$y_i(t) = \sum_{k=1}^{K} c_k^i \phi_k(t), t \in [1, T]. \tag{3}$$

In this study, Fourier basis functions were employed due to periodic nature of the water quality data, and the optimal number of basis functions (smoothing parameter) was determined using the Generalized Cross-Validation (GCV) criterion (Ramsay & Silverman, 2005; Ullah & Finch, 2013).The smoothed functional data $X_i(t)$ for each station i are thus given by:

$$X_i(t) = \hat{y}_i(t) = \sum_{k=1}^{K} \hat{c}_k^i \phi_k(t), t \in [1, T], \tag{4}$$

where the coefficients $\hat{c}_{ik}$ are estimated by minimizing the sum of squares errors

$$SSE(i) = \sum_{j=1}^{T} \left(x_i(t_j) - y_i(t_j)\right)^2. \tag{5}$$

This procedure yields a collection of smooth curves $\{X(s_i, t)\}$, each indexed by the spatial location s_i of the monitoring station. These spatially-referenced functional data are subsequently analyzed to investigate both temporal patterns and spatial variation in water quality across the region. The functional data processing was performed using the *fda* package in *R* (Ramsay et al., 2009).

Functional principal component analysis and spatial characterization

To investigate the main modes of variation in the smoothed water quality functions across stations and explore spatial patterns, a two-stage approach was employed. Let $X(s_i, t)$ denote the smoothed water quality function at monitoring station i, observed over a continuous time domain $t \in [1, T]$, and located at spatial location s_i. For notational simplicity in the FPCA formulation, we denote this as $X_i(t)$, where each index i corresponds to a unique station location s_i. In the first stage, FPCA was applied to the set of functions $\{X_i(t)\}_{i=1}^n$, where FPCA decomposes each function into a mean function and a series of orthogonal functional principal components (fPCs), capturing dominant temporal patterns in the data:

$$X_i(t) = \mu(t) + \sum_{m=1}^{M} \xi_{im} \phi_m(t), \tag{6}$$

where $\mu(t)$ is the overall mean function across all stations, $\phi_m(t)$ are the orthonormal eigenfunctions (fPCs), and ξ_{im} are the fPCs scores for station i, representing the contribution of the m-th component to the variation in $X_i(t)$. These scores are computed as the inner product:

$$\xi_{im} = \int X_i(t) \phi_m(t) dt \tag{7}$$

The eigenfunctions $\phi_m(t)$ often interpreted as weighted functions, define how different parts of the time domain contribute to the principal modes of variation in the data. The first fPC score for station i is then calculated as:

$$\xi_{i1} = \int (X_i(t) - \mu(t)) \phi_1(t) dt \tag{8}$$

The first fPC is determined by maximizing the variance of these scores under the constraint that the squared norm of the weighted function $\left|\xi_1^2\right|$ satisfying $\int \xi_1^2(t) dt = 1$, and all subsequent components orthogonal to the previous ones (Ramsay & Silverman, 2005). All computations were carried out using the *fda* package in *R*.

In the second stage, the leading fPC scores ξ_{im} were analyzed across space to explore spatial variation. Each score represents how strongly a station's WQI curve aligns with the temporal pattern captured by the corresponding fPC. Positive scores indicate expression of the pattern in the same direction as the fPC, while negative scores indicate the opposite tendency, and scores near zero suggest behavior close to the mean WQI profile. The variation of these scores across monitoring stations highlights regional similarities and contrasts in WQI dynamics, possibly reflecting environmental or anthropogenic influences.

Spatial functional clustering

To identify spatial clustering of monitoring stations with similar temporal water quality behaviors, spatial functional clustering was performed based on the functional principal component (fPC) scores derived in the previous stage. These scores, $\{\xi_{im}\}$, serve as low-dimensional representations of the smoothed water quality functions $X(s_i, t)$, effectively capturing the dominant temporal variation at each station. By clustering stations using their leading fPC scores which is the first two components, we aim to group locations that exhibit similar temporal profiles in water quality dynamics. *K*-means clustering method was applied to the fPC score matrix, and the optimal number of clusters was determined using elbow method. The resulting clusters were then mapped spatially to interpret regional patterns and investigate potential environmental or anthropogenic factors influencing the water quality profiles across stations.

Results and discussion

Functional representation of water quality index data

Monthly river water quality index (WQI) values were smoothed into continuous functional curves for each monitoring station using a Fourier basis representation. Figure 2 displays the resulting functional data, where each curve represents the temporal water quality trend of a station from 2020 to 2023. The WQI values range between 44 and 95, with the overall mean function fluctuating around a WQI of 72. Several stations consistently deviate from the mean curve, indicating local variability in seasonal trends or episodic events. Among these, stations 1K24 (Sg. Gombak), 1K45 (Sg. Klang), 1K50 (Sg. Penchala), and 1K51 (Sg. Damansara) recorded WQI values persistently above the mean. These stations specifically, Station 1K24 (Sg. Gombak) and 1K45 (Sg. Klang) are located.

near protected upstream catchments, including water supply dams such as Batu Dam and Klang Gates Dam, and appear less exposed to anthropogenic stressors. This suggests effective management in these areas, warranting continued protection

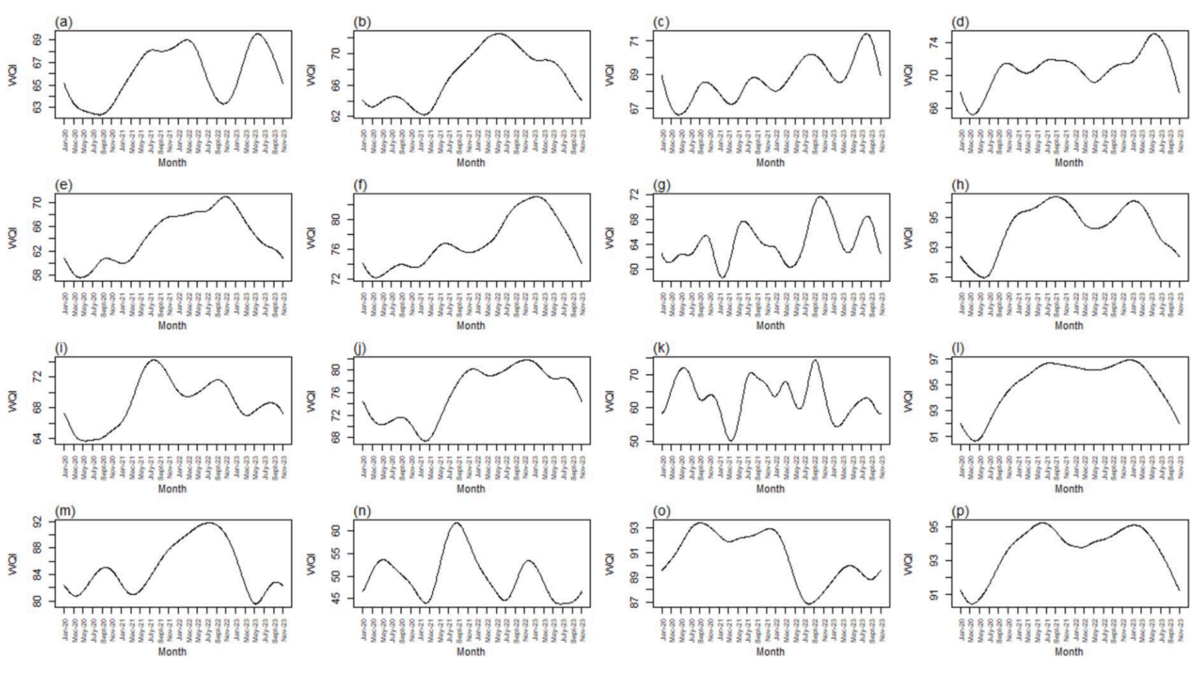

Fig. 2 Functional data plots for the 16 monitoring stations: (**a**) 1K05, (**b**) 1K06, (**c**) 1K07, (**d**) 1K08, (**e**) 1K17, (**f**) 1K18, (**g**) 1K23, (**h**) 1K24, (**i**) 1K25, (**j**) 1K36, (**k**) 1K41, (**l**) 1K45, (**m**) 1K46, (**n**) 1K47, (**o**) 1K50, (**p**) 1K51

and monitoring to preserve their good water quality status.

Conversely, most downstream stations recorded functional curves below the mean. Station 1K47 (Sg. Kerayong) displayed the lowest and most unstable WQI values throughout the study period. The repeated declines in water quality at this station occurred at irregular intervals, suggesting the influence of localized, possibly intermittent, pollution sources. These fluctuations warrant further investigation into surrounding land use and potential pollutant discharge events.

Spatial and temporal variation among water quality stations

Functional principal component analysis (FPCA) was used to decompose the smoothed WQI curves into major modes of temporal variation across station. Figure 3 presents the first three weight functions (fPCs), each capturing a different aspect of variation in WQI from January 2020 to December 2023. Table 2 reports the corresponding eigenvalues, variance explained, and cumulative variance. The first functional principal component (fPC1) explains 97.2% of the total variation, indicating it captures the dominant pattern of temporal differences in WQI across all stations. This large proportion arises

Table 2 The fPCs scores for the first three components (fPC1–fPC3), with eigenvalues, percentage of variance explained, and cumulative variance

Station	fPC1	fPC2	fPC3
1K05	−41.58	1.31	−5.81
1K06	−35.67	−7.54	3.33
1K07	−28.56	1.40	−3.77
1K08	−19.06	0.69	−7.97
1K17	−49.76	−10.33	1.48
1K18	11.40	−8.79	−4.30
1K23	−49.02	−0.81	−5.00
1K24	94.46	2.82	−1.42
1K25	−29.03	−0.14	−2.12
1K36	5.52	−12.82	3.09
1K41	−55.51	8.58	12.12
1K45	97.67	0.69	−0.21
1K46	49.69	−3.82	11.62
1K47	−118.46	12.61	1.19
1K50	76.92	13.48	−0.75
1K51	90.99	2.64	−1.47
Eigenvalues	3894.50	53.48	29.42
% Variance	97.18	1.33	0.73
Cumulative variance	97.18	98.52	99.25

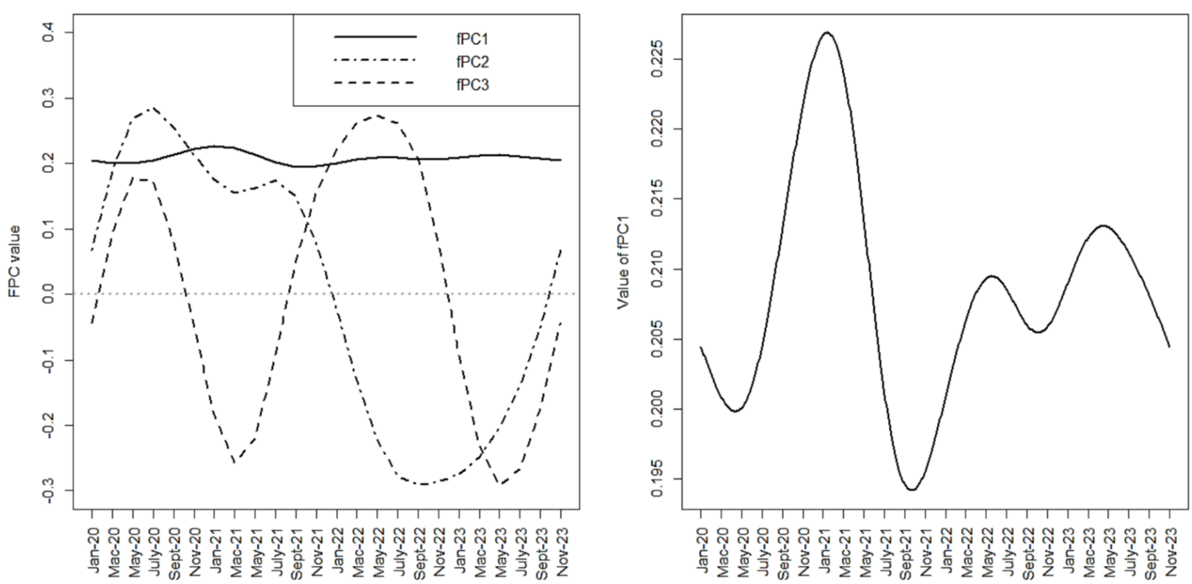

Fig. 3 The first three functional principal component curves of the WQI

because aggregating across stations emphasizes the common seasonal hydrological cycle, which strongly governs water quality fluctuations in the basin. So, fPC1 mainly reflects the monsoonal signal that underpins variation basin-wide, while local differences are expressed in higher-order components. The shape of fPC1 highlights that the biggest differences in water quality between stations occurred around January 2021. This period matches the peak of the Northeast Monsoon, when downstream stations especially 1K47 (Sg. Kerayong) showed strong changes in WQI. These changes are likely influenced by increased runoff from urban areas, contributing to episodic pollution. In contrast, upstream stations displayed more stable trends, likely due to their relative insulation from anthropogenic stressors. In addition, a notable trough in fPC1 appears in October 2021, suggesting minimal differences among stations during this period, possibly due to transitional weather conditions between monsoon phases. This uniformity may reflect a temporary hydrological balance, where pollution dispersion and dilution effects are more evenly distributed across the basin.

The second component (fPC2), accounts for approximately 2% of total variation and reveals subtler temporal dynamics. It shows positive values from January 2020 to early 2022, followed by negative values until the end of 2023. This trend suggests a shift in overall water quality conditions over time. The positive phase likely reflects improved water quality during the COVID-19 pandemic, when reduced industrial and human activity may have contributed to temporary reductions in pollutant loads. The subsequent decline could be linked to resumed development and anthropogenic activities, aligning with easing pandemic restrictions. The third component, (fPC3), explains about 0.7% of the variation and displays a periodic pattern with peaks in May 2020 and May 2022, and troughs in March 2021 and May 2023. These fluctuations are likely associated with seasonal hydrological shifts, including the inter-monsoon periods. The observed pattern may reflect increased pollutant concentrations during drier periods and dilution effects during wetter months. To support interpretation, perturbation plots in Fig. 4 visualize how each component alters the mean function. The perturbation refers to visualizing how the mean WQI curve changes when a multiple of an fPC is added (positive direction) or subtracted (negative direction). This helps illustrate the characteristic temporal pattern captured by each component. The fPC1 perturbation (Fig. 4a) shows a strong, consistent trend in water quality variation. fPC2 (Fig. 4b) highlights differences between pre- and post-2022 periods, while fPC3 (Fig. 4c) illustrates subtle, cyclic deviations potentially tied to seasonal rainfall patterns.

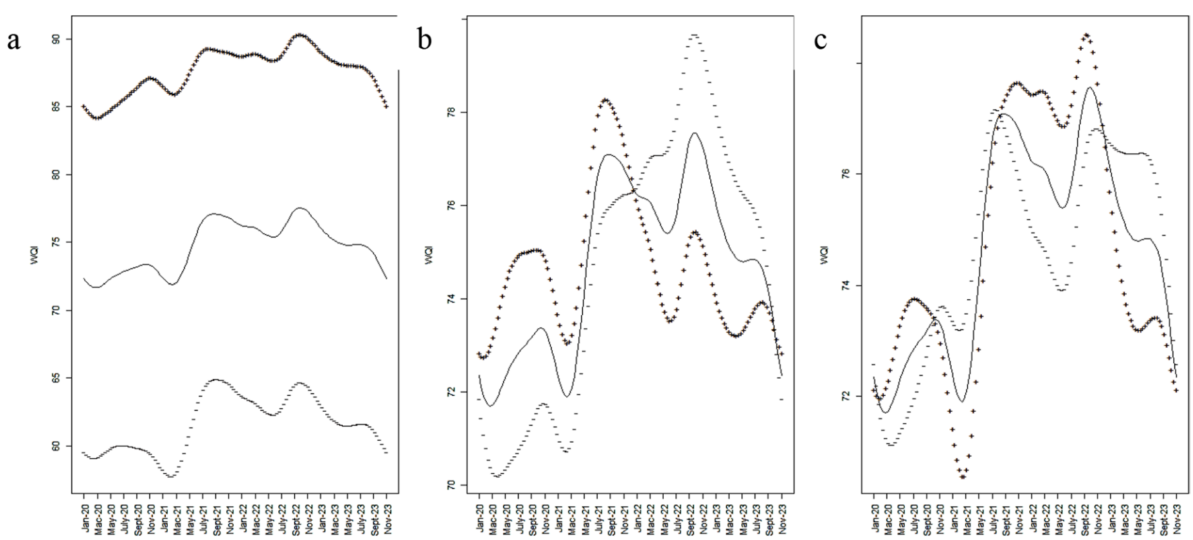

Fig. 4 Perturbation plots of the mean WQI curve by the first three functional principal components: (**a**) fPC1, (**b**) fPC2, (**c**) fPC3. Mean WQI (__ line), positive perturbation (+ + + line), negative perturbation (----- line)

Table 2 presents the fPC scores for each station. These scores quantify how strongly each station's water quality profile deviates from the basin-wide mean curve in the direction of a given fPC. Unlike the perturbation plots in Fig. 4, which visualize how the mean curve shifts when adding or subtracting multiples of an fPC, the station scores represent the actual magnitude and direction of each station's deviation from that mean pattern. For fPC1, stations 1K45, 1K24, 1K50, and 1K51 scored highest, aligning with overall higher WQI levels and more stable trends which shows the characteristics of upstream or well-managed catchments. Conversely, station 1K47 recorded the most negative score (−118.46), indicating a strong deviation in the opposite direction of fPC1, thereby reinforcing its status as the most degraded and variable station.

For fPC2, positive scores at stations 1K50, 1K47, and 1K41 suggest a pattern of improved water quality during the early part of the study, followed by a decline which was consistent with the COVID-19 timeline. This pattern highlights the temporary improvement in water quality conditions during the early lockdown period, followed by a rebound of degradation as economic and social activities resumed post-2020. Such findings align with other studies reporting the short-lived nature of water quality recovery during the pandemic, underscoring that reductions in pollution were largely incidental rather than the result of structural changes in catchment management (Najah et al., 2021). Negative scores at 1K36 and 1K17 suggest the inverse trend, potentially due to local differences in land use response or resilience. Regarding fPC3, stations 1K41 and 1K46 recorded the highest scores, indicating stronger seasonal or cyclical water quality variation, possibly due to recurring discharges or runoff events. Meanwhile, stations such as 1K08 and 1K05 recorded strongly negative scores, suggesting reduced or inverse seasonal effects.

The biplot in Fig. 5 displays the spatial distribution of stations based on their fPC1 and fPC2 scores. Stations located in the first quadrants (positive fPC1 and positive fPC2) such as 1K45, 1K24, 1K50 have both better and more stable water quality patterns over time (high fPC1) and a temporal trajectory of initial improvement followed by decline (positive fPC2). In contrast, station 1K47 appears in the third quadrant (negative fPC1 and negative fPC2), reflecting both poorer and unstable water quality and an opposite temporal trend, likely influenced by downstream urban pollution.

Spatial clustering based on functional scores

To further explore spatial patterns in river water quality, *k*-means clustering was applied to the first two functional principal component (fPC) scores. This approach grouped stations based on similarities in their temporal water quality trends, resulting in three distinct spatial clusters, as shown in Fig. 6. Cluster 1 includes upstream stations such as 1K24 (Sg. Gombak) and 1K45 (Sg. Klang), which had high positive fPC1 scores. These stations consistently exhibited higher WQI values and more stable patterns, suggesting minimal disturbance and better catchment management. Cluster 2 represents midstream stations with moderate fPC1 and fPC2 scores, indicating more variable water quality likely influenced by mixed land uses and transitional catchment conditions. Cluster 3 comprises downstream or urban-influenced stations such as 1K47 (Sg. Kerayong), which recorded strongly negative fPC1 scores, reflecting persistently degraded water quality and greater fluctuation over time.

Consistent with these patterns, classification of WQI scores using the National Water Quality Standards (Department of Environment (DOE), 2019) shows that most Cluster 1 stations fall into Class II (Clean), suitable for recreational use and aquatic life protection. Cluster 2 stations fluctuate between Class II and Class III (Slightly Polluted), reflecting transitional conditions shaped by land use and seasonal effects. Cluster 3 stations are predominantly classified as Class III (Slightly Polluted to Polluted), underscoring persistent anthropogenic pressures and reduced suitability for sensitive uses. The spatial distribution of clusters corresponds closely with known catchment characteristics. Stations in Cluster 3 are located in densely populated or industrial zones, where surface runoff and point-source pollution are more likely. In contrast, Cluster 1 stations are found in relatively undisturbed or forested regions, reinforcing the influence of land use and human activity on river health. By integrating FPCA with clustering, this analysis provides a clearer understanding of how how temporal water quality dynamics vary spatially across the Klang River Basin. The distinct grouping of stations

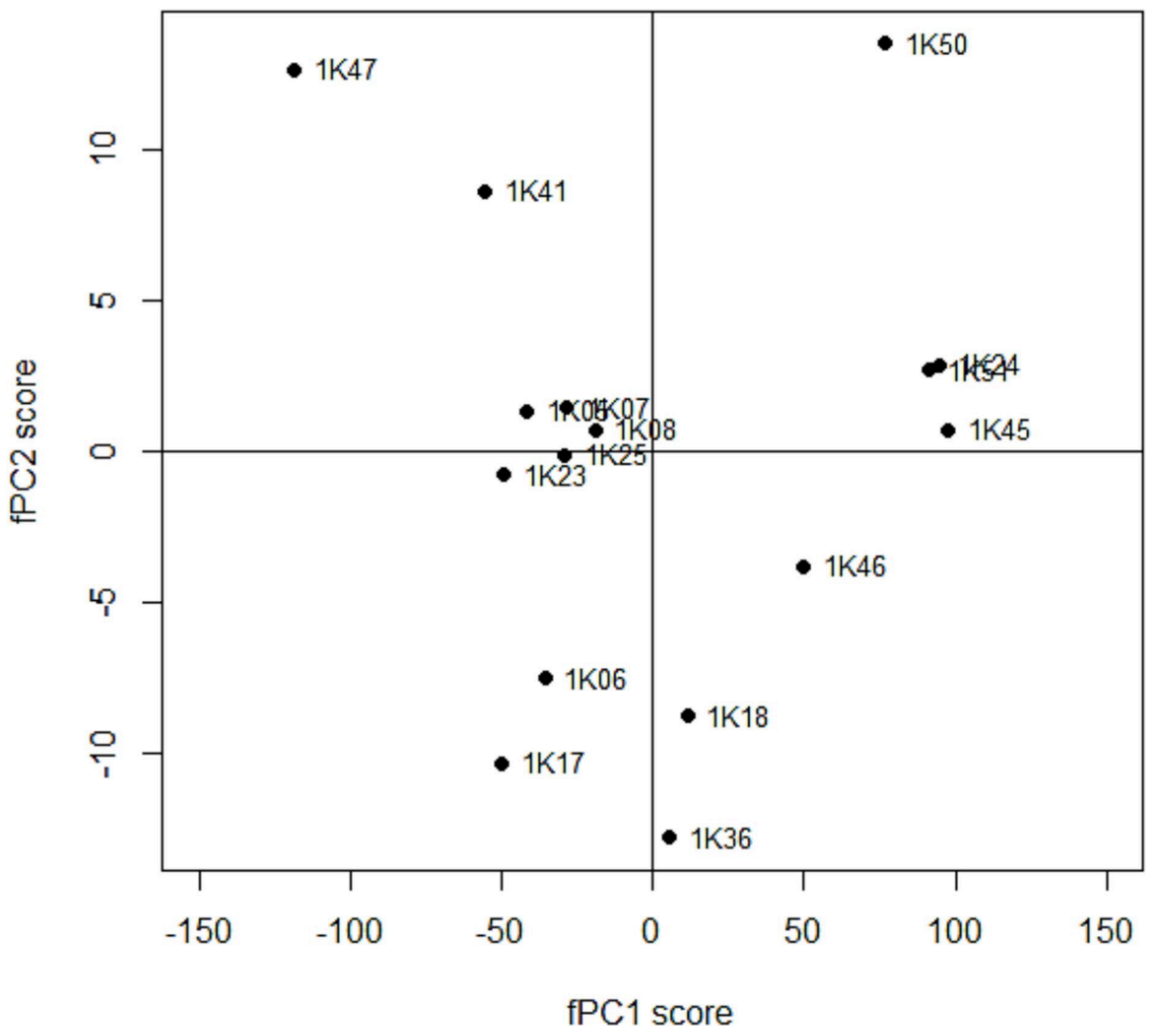

Fig. 5 The biplot of fPCs scores

emphasizes the importance of localized monitoring and targeted interventions, particularly in downstream areas where water quality remains vulnerable to urban pressures.

In this study, FDA demonstrates strong potential for capturing continuous temporal dynamics and identifying dominant modes of variation in river water quality, which may be less apparent with traditional statistical techniques (Ramsay & Silverman, 2005). While machine learning approaches such as random forests and neural networks offer predictive power for classification and forecasting (Costa et al., 2024), and remote sensing tools enable large-scale monitoring with high spatial coverage (Zhang et al., 2024), FDA provides a more interpretable framework for tracing the evolution of water quality curves over time. Instead of competing with emerging methods, this analysis complements them and underscores the FDA's role in generating spatio-temporal insights to aid in sustainable river basin management.

Conclusion

This study applied FDA to examine spatio-temporal pattern of river water quality in the Klang River basin

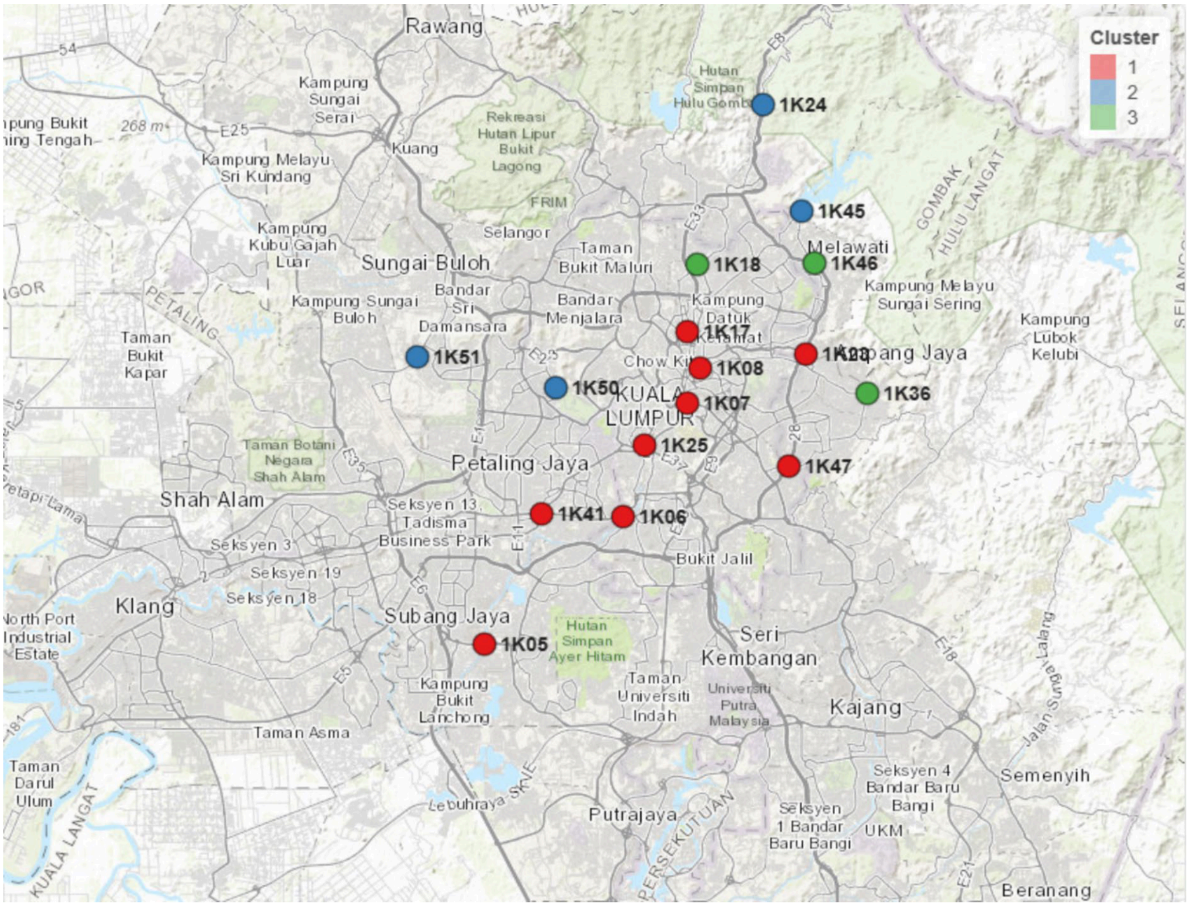

Fig. 6 Cluster Map of Functional WQI Data

from 2020 to 2023. By treating monthly WQI data as smooth temporal functions, FDA allowed a more comprehensive assessment of water quality dynamics compared to traditional methods. The results showed that upstream stations, particularly those located near water supply dams, maintained consistently good water quality. In contrast, downstream stations exhibited greater variability, particularly during monsoon periods, reflecting stronger anthropogenic and hydrological influences. The first functional principal component, fPC1 accounted for 97% of the total variation and highlighted key periods such as January 2021 when inter-station differences were most pronounced. The second and third components revealed subtler patterns, including shifts that aligned with pandemic-related activity changes, showing improvements during the early COVID-19 restrictions, followed by a decline as economic and social activities resumed alongside recurring seasonal fluctuations. Using the scores from the first three fPCs, the 16 stations were grouped into three spatial clusters based on similar water quality patterns. Cluster 1 included upstream stations with stable, good-quality water. Cluster 2 represented midstream stations with moderate but variable conditions, while Cluster 3 encompassed downstream stations characterized by poorer and more unstable water quality, often in urbanized areas.

While this study offers valuable insights into temporal and spatial patterns of river water quality, it is limited by the relatively small number of monitoring stations and its focus on a single river basin. In addition, the classification of water quality relied on the existing WQI criterion, which may not fully capture context-specific thresholds. Future studies could refine or validate this framework to enhance its robustness. Moreover, the clustering relied on scalar

FPCA scores, which do not fully capture the spatial structure or flow-connected dynamics of river networks. Future work could explore functional spatial clustering techniques, incorporate hydrological connectivity, or apply the approach at broader spatial scales to enhance generalizability and policy relevance.

Overall, this FDA-based approach provided clearer insight into both temporal and spatial variation in river water quality. It also offers a practical tool for environmental monitoring and management by supporting early detection of emerging pollution trends and helping prioritize interventions in vulnerable locations. The findings emphasize the importance of protecting upstream sources, mitigating downstream pollution, and incorporating seasonal variability into long-term river basin management strategies.

Author Contribution N.F.M.A. wrote the main manuscript text and editing the manuscript. I.M. contributed to reviewing and editing the manuscript, and participated in discussions to refine the analysis. R.M.Y. contributed to reviewing and editing the manuscript, and participated in discussions to refine the analysis. F.O. prepared the map of river basin and participated in discussions to refine the analysis. All authors reviewed the manuscript.

Funding Open access funding provided by The Ministry of Higher Education Malaysia and International Islamic University Malaysia.

Data availability The data that support the findings of this study are available from the authors but restrictions apply to the availability of these data, which were used under license from the Department of Environmental Malaysia for the current study, and so are not publicly available. Data are, however, available from the authors upon reasonable request and with permission from the Department of Environmental Malaysia.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Anh, N. T., Nhan, N. T., Schmalz, B., & Le Luu, T. (2023). Influences of key factors on river water quality in urban and rural areas: A review. *Case Studies in Chemical and Environmental Engineering*. https://doi.org/10.1016/j.cscee.2023.100424

Ariza, A., Lebourges-Dhaussy, A., Nerini, D., Pauthenet, E., Roudaut, G., Assunção, R., & Bertrand, A. (2023). Acoustic seascape partitioning through functional data analysis. *Journal of Biogeography, 50*(9), 1546–1560.

Azha, S. F., Sidek, L. M., Ahmad, Z., Zhang, J., Basri, H., Zawawi, M. H., & Ahmed, A. N. (2023). Enhancing river health monitoring: Developing a reliable predictive model and mitigation plan. *Ecological Indicators, 156*, 111190.

Barroso, G. R., Pinto, C. C., Gomes, L. N. L., & Oliveira, S. C. (2024). Assessment of water quality based on statistical analysis of physical-chemical, biomonitoring and land use data: Manso River supply reservoir. *Science of the Total Environment, 912*, 169554.

Costa, D., Bayissa, Y., Barbosa, K. V., Villas-Boas, M. D., Bawa, A., Junior, J. L., Neto, A. J. S., & Srinivasan, R. (2024). Water quality estimates using machine learning techniques in an experimental watershed. *Journal of Hydroinformatics, 26*(11), 2798–2814.

Department of Environment Malaysia (DOE). (2019). *Malaysia environmental quality report 2019 (Laporan Kualiti Alam Sekeliling Malaysia 2019)*. Ministry of Natural Resources and Environmental Sustainability: Department of Environment.

Di Blasi, J. P., Torres, J. M., Nieto, P. G., Fernández, J. A., Muñiz, C. D., & Taboada, J. (2013). Analysis and detection of outliers in water quality parameters from different automated monitoring stations in the Miño river basin (NW Spain). *Ecological Engineering, 60*, 60–66.

Dona, G., Preatoni, E., Cobelli, C., Rodano, R., & Harrison, A. J. (2009). Application of functional principal component analysis in race walking: An emerging methodology. *Sports Biomechanics, 8*(4), 284–301.

Embling, C. B., Illian, J., Armstrong, E., van der Kooij, J., Sharples, J., Camphuysen, K. C., & Scott, B. E. (2012). Investigating fine-scale spatio-temporal predator–prey patterns in dynamic marine ecosystems: A functional data analysis approach. *Journal of Applied Ecology, 49*(2), 481–492.

Gong, M., Miller, C., & Scott, M. (2015). Functional PCA for remotely sensed lake surface water temperature data. *Procedia Environmental Sciences, 26*, 127–130.

Gong, M., Miller, C., Scott, M., O'Donnell, R., Simis, S., Groom, S., & Spyrakos, E. (2021). State space functional principal component analysis to identify spatiotemporal patterns in remote sensing lake water quality. *Stochastic Environmental Research and Risk Assessment, 35*, 2521–2536.

Haggarty, R. A., Miller, C. A., & Scott, E. M. (2015). Spatially weighted functional clustering of river network data. *Journal of the Royal Statistical Society Series C: Applied Statistics, 64*(3), 491–506.

Hamid, A., Bhat, S. U., & Jehangir, A. (2020). Local determinants influencing stream water quality. *Applied Water Science, 10*(1), 1–16.

Hampel, M., Blasco, J., & Segner, H. (2015). Molecular and cellular effects of contamination in aquatic ecosystems. *Environmental Science and Pollution Research, 22*, 17261–17266.

Henderson, B. (2006). Exploring between site differences in water quality trends: A functional data analysis approach. *Environmetrics, 17*(1), 65–80.

Islam, M. S., Uddin, M. K., Tareq, S. M., Shammi, M., Kamal, A. K. I., Sugano, T., ..., & Kuramitz, H. (2015). Alteration of water pollution level with the seasonal changes in mean daily discharge in three main rivers around Dhaka City, Bangladesh. *Environments, 2*(3), 280-294.

Ismail, W. R., Omar, M. A., Rahaman, Z. A., Najib, S. A. M., Sah, A. S. R. M., & Razad, A. A. (2023). Hydrology, water quality and trophic state of Pergau Reservoir, Kelantan, Malaysia. *Water Science, 37*(1), 131–150.

Kande, Y., Diogoul, N., Brehmer, P., Dabo-Niang, S., Ngom, P., & Perrot, Y. (2024). Demonstrating the relevance of spatial-functional statistical analysis in marine ecological studies: The case of environmental variations in micronektonic layers. *Ecological Informatics, 81*, Article 102547.

Karuppusami, R., Antonisamy, B., & Premkumar, P. S. (2022). Functional principal component analysis for identifying the child growth pattern using longitudinal birth cohort data. *BMC Medical Research Methodology, 22*(1), 76.

Ling, T. Y., Gerunsin, N., Soo, C. L., Nyanti, L., Sim, S. F., & Grinang, J. (2017). Seasonal changes and spatial variation in water quality of a large young tropical reservoir and its downstream river. *Journal of Chemistry*. https://doi.org/10.1155/2017/8153246

Loi, J. X., Chua, A. S. M., Rabuni, M. F., Tan, C. K., Lai, S. H., Takemura, Y., & Syutsubo, K. (2022). Water quality assessment and pollution threat to safe water supply for three river basins in Malaysia. *Science of the Total Environment, 832*, 155067.

Mohamed, I., Othman, F., Ibrahim, A. I., Alaa-Eldin, M. E., & Yunus, R. M. (2015). Assessment of water quality parameters using multivariate analysis for Klang River basin, Malaysia. *Environmental Monitoring and Assessment, 187*(1), 1–12.

Mohammed, A., Samara, F., Alzaatreh, A., & Knuteson, S. L. (2022). Statistical analysis for water quality assessment: A case study of Al Wasit nature reserve. *Water (Basel), 14*(19), 3121.

Mohd Ali, N. F., Mohamad Yunus, R., Mohamed, I., & Othman, F. (2024). Spatial functional outlier detection in multivariate spatial functional data. *Sains Malaysiana, 53*(6), 1463–1476.

Mokhtar, Z. (2023). Review of Malaysia's environmental waterways' compliance with industrial effluent discharge. *ScienceOpen Preprints*. https://doi.org/10.14293/PR2199.000011.v1

Muñiz, C. D., Nieto, P. G., Fernández, J. A., Torres, J. M., & Taboada, J. (2012). Detection of outliers in water quality monitoring samples using functional data analysis in San Esteban estuary (Northern Spain). *Science of the Total Environment, 439*, 54–61.

Najah, A., Teo, F. Y., Chow, M. F., Huang, Y. F., Latif, S. D., Abdullah, S., & El-Shafie, A. (2021). Surface water quality status and prediction during movement control operation order under COVID-19 pandemic: Case studies in Malaysia. *International Journal of Environmental Science and Technology, 18*(4), 1009–1018.

Nasir, M. F. M., Samsudin, M. S., Mohamad, I., Awaluddin, M. R. A., Mansor, M. A., Juahir, H., & Ramli, N. (2011). River water quality modeling using combined principle component analysis (PCA) and multiple linear regressions (MLR): A case study at Klang River, Malaysia. *World Applied Sciences Journal, 14*(1), 73–82.

Nienie, A. B., Sivalingam, P., Laffite, A., Ngelinkoto, P., Otamonga, J. P., Matand, A., & Poté, J. (2017). Seasonal variability of water quality by physicochemical indexes and traceable metals in suburban area in Kikwit, Democratic Republic of the Congo. *International Soil and Water Conservation Research, 5*(2), 158–165.

Othman, F., Me, A. E., & Mohamed, I. (2012). Trend analysis of a tropical urban river water quality in Malaysia. *Journal of Environmental Monitoring, 14*(12), 3164–3173.

Ramsay, J., Hooker, G., & Graves, S. (2009). Introduction to functional data analysis. *In Functional data analysis with R and MATLAB* (pp. 1–19). Springer. https://doi.org/10.1007/978-0-387-98185-7_1

Ramsay, J. O., & Silverman, B. W. (2005). Principal components analysis for functional data. *In Functional data analysis* (2nd ed., pp. 147–172). Springer. https://doi.org/10.1007/b98888

Schreiber, S. G., Schreiber, S., Tanna, R. N., Roberts, D. R., & Arciszewski, T. J. (2022). Statistical tools for water quality assessment and monitoring in river ecosystems–A scoping review and recommendations for data analysis. *Water Quality Research Journal, 57*(1), 40–57.

Shan, V., Singh, S. K., & Haritash, A. K. (2021). Evaluation of water quality and potential metal contamination in ecologically important Bhindawas bird sanctuary, India. *Applied Water Science, 11*(1), 8.

Sidek, L., Basri, H., Lee, L. K., & Foo, K. Y. (2016). The performance of gross pollutant trap for water quality preservation: A real practical application at the Klang Valley, Malaysia. *Desalination and Water Treatment, 57*(52), 24733–24741.

Ullah, S., & Finch, C. F. (2013). Applications of functional data analysis: A systematic review. *BMC Medical Research Methodology, 13*(1), 1–12.

United Nations Environment Programme. (2008). W*ater quality for ecosystem and human health* (2nd ed.). Ottawa, Ontario, Canada: UNEP Global Environment Monitoring System (GEMS)/Water Programme. ISBN 92-95039-51-7.

Vega-Rodríguez, M. A., Perez, C. J., Reder, K., & Floerke, M. (2021). A stage-based approach to allocating water quality monitoring stations based on the WorldQual model: The Jubba River as a case study. *Science of the Total Environment, 762*, 144162.

Wong Y. J., Shimizu Y., He K. & Nik Sulaiman N. M. (2020). Comparison among different ASEAN water quality indices for the assessment of the spatial variation of surface water quality in the Selangor river basin, Malaysia. *Environmental Monitoring and Assessment 192*, 1–16.

Yan, F., Liu, L., Li, Y., Zhang, Y., Chen, M., & Xing, X. (2015). A dynamic water quality index model based on functional data analysis. *Ecological Indicators, 57*, 249–258.

Zhang, Y., Li, W., Wen, W., Zhuang, F., Yu, T., Zhang, L., & Zhuang, Y. (2024). Universal high-frequency monitoring methods of river water quality in China based on machine learning. *Science of the Total Environment, 947*, 174641.

Zhao, P., Li, Z., Zhang, R., Pan, J., & Liu, Y. (2020). Does water diversion project deteriorate the water quality of reservoir and downstream? A case-study in Danjiangkou reservoir. *Global Ecology and Conservation, 24*, e01235.