Volume 11, Issue 3, March – 2026 International Journal of Innovative Science and Research Technology
ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar1692

Multivariate Analysis of the Irwell River Water Quality Drivers Bolton WWTW Effluent

Nwogbu Peter[1]; Dr. Yassin Osman[2]; Dr. Stephen Ikporo[3]; Mkpumah Emmanuel[4]

[1,2,3,4]Civil Engineering, Department of Environmental Management, University of Greater Manchester, Deane Road, Bolton in the United Kingdom.

Publication Date: 2026/04/07

Abstract: This paper evaluates how Bolton Wastewater Treatment Works (WWTW) influences the physicochemical water quality of the River Irwell using an integrated multivariate statistical approach. Weekly grab samples were collected for four weeks at three sites: upstream control, effluent discharge, and downstream recovery. Parameters measured includes BOD5, dissolved oxygen, electrical conductivity, turbidity, pH, temperature, and flow conditions following APHA standard methods with strict QA/QC procedures. Primary data were complemented by Environment Agency records, operator discharge data, and CSO events. Data were standardized and analysed using descriptive statistics and multivariate methods including.

The multivariate analysis showed distinct and statistically significant division of the upstream, effluent, and downstream areas of sampling (PERMANOVA pseudo-F = 4.19, p = 0.005). Factor Analysis revealed two latent factors of the most prominent result that explained 59.28 percent of the overall variance. Factor 1 was a great chemical gradient with high positive loadings of EC (+0.82) and negative loadings of DO (-0.96) and pH (-0.81), which is in line with the effects of treated wastewater effluent. A rather low loading of BOD5 (−0.26) demonstrated positive removal of biodegradable organic matter during the treatment process and a minor contribution of turbidity to this main gradient. Factor 2 was an independent process related to sediments and turbidity expressed the greatest positive loading to oxygen and organic pollution dynamics (+0.55).

Biplot analysis showed clear zonal clustering: upstream samples were linked to oxygen-rich, low-conductivity conditions; and effluent samples to oxygen-depleted, high-conductivity conditions; and downstream samples occupied an intermediate position, indicating partial chemical recovery and sediment influence. Dissolved oxygen was the most sensitive indicator of effluent impact, with mean depletion at the effluent site exceeding 55%, while electrical conductivity showed a consistent downstream gradient. A weighted effluent impact index (EII) classified baseline, impacted, and recovering zones without upstream false positives or effluent false negatives. Comparison of primary and secondary data showed differences in BOD5 and pH but strong agreement in temperature, highlighting the need for harmonized monitoring standards.

Keywords: *River Irwell; Bolton Wastewater Treatment Works; Urban River Pollution; Wastewater Effluent Impacts; Multivariate Statistical Analysis; Dissolved Oxygen; Electrical Conductivity.*

How to Cite: Nwogbu Peter; Dr. Yassin Osman; Dr. Stephen Ikporo; Mkpumah Emmanuel (2026) Multivariate Analysis of the Irwell River Water Quality Drivers Bolton WWTW Effluent. *International Journal of Innovative Science and Research Technology,* 11(3), 3437-3454. https://doi.org/10.38124/ijisrt/26mar1692

I. INTRODUCTION

➢ *Background and Historical Context*

The main source of treated wastewater is the Bolton Wastewater Treatment Works (WWTW) that is located close to the territory of the Outwood area. Most recently United Utilities has spent PS110m on increasing capacity, altering storm overflow outfalls and placing modern treatment technologies to reduce nutrient and organic loads (United Utilities, 2024). Despite these kinds of investments, water-quality pressures are significant. It has been recorded that there have been continuous activities of storm overflow under the works that have recorded 88 discharge events in 2024 which translates to more than 620 hours of intermittent discharge to the receiving watercourse.

Such occurrences occur as a rule when the combined sewer systems are overstretched and thus the partially treated or screened waste is discharged into the river. This is one of the most prominent effects of hydrology and infrastructure on the situation (poor ecology) in urban catchments (Rosenqvist, et al., 2021). In addition to overflows events, diffuse urban runoff in roads, business and residential places introduces hydrocarbons, metals, nutrients and suspended solids that further complicate the problems of the water quality management in the Irwell. Recent assessments of catchments indicate that the Irwell catchment has almost 86 percent of its water bodies in a moderate ecological state and biochemical oxygen demand, turbidity, metals and bacterial contaminations are of interest (Ecosystems Knowledge Network, 2025). These figures indicate the need to have analytical processes that do not just capture singular parameter measures but continue to establish the multivariate tendencies of water quality to have a better understanding of the combined and interrelating effects of effluents and other causes of contamination.

This kind of analysis is needed because the single-point analysis provides low levels of temporal area and can often fail to capture episodic contamination which in some cases is of crucial constituent to aquatic biota. Multivariate procedures, in comparison, can determine correlated reactions of a limited number of parameters and, therefore, can better assign effects to wastewater, runoff or legacy sediments (Ioele, et al., 2020; Ma, et al., 2020). Furthermore, Irwell catchment lies in the socio-environmental location, and this factor predisposes the need to develop a high-quality assessment.

The river traverses mixed land use, consisting of peri-urban moorland, through densely populated urban centres, having different load of pollution and flow regimes. The urban development has increased the size of hardsurface, accelerates the velocity of runoff and reduces the infiltration capacity that further increases the intensity and frequency of storm-overflow activations. It is an anthropogenic impact on the climatic response of variations in the intensity of rainfall, which promoted episodic discharges, which are immeasurable using conventional sampling programmes (Zhao, et al., 2022). Thus, there exists an opportunity of a special study around Bolton WWTW which may be employed in exploring effluent influence in a typical urban catchment environment. Map shows the River Irwell and the position of Bolton WWTW which can give a perspective of the layout of sampling areas and illustrate the position of upstream control points, position of effluent discharge and position of downstream monitoring points.

➢ *Study Area and Site Selection*

- *River Irwell Context*

The River Irwell is an urban catchment, which has historical industrial contributions, modern wastewater discharges, and a high number of CSOs (FloodMapper, 2024). It experiences change in its hydrology and pollutants signatures depending on rainfall, seasonality and WWTW operational cycles.

Table 1 First Day Sample Log

Sample ID	**Site**	**Date**	**Time**	**DO1(mg/l)**	**DO2 (mg/l)**	**Turbidity (NTU)**	**pH lab**	**EC (µS cm^{-1})**	**Notes**
S2025_001	US	2025-11-19	12:00	7.24 10.3°C	5.92 17.9°C	14.08	7.08 12.2°C	384 12.2°C	Dry weather
S2025_002	EF	2025-11-19	12:15	4.12 12.3°C	4.31 17.6°C	12.22	7.03 11.6°C	354 11.3°C	Visible plume
S2025_003	DS	2025-11-19	15:05	7.32 9.7°C	5.33 17.7°C	29.09	7.31 9.9°C	339 9.5°C	cloudy

Table 2 Second Day Sample Log

Sample ID	**Site**	**Date**	**Time**	**DO Site (mg/L)**	**DO1 lab (mg/L)**	**DO2 (5) (mg/L)**	**Turbidity (NTU)**	**PH lab**	**EC (µS/cm**	**Notes**
S2025-001	US	2025-11-27	10:00	9.2 7.5°C	7.1 13.5°C	5.28 18.8°C	6.02	7.34 14.1°C	383 14.2°C	Showering
S2025-002	EF	2025-11-27	10:15	3.73 12.3°C	6.02 13.7°C	4.92 19.1°C	2.49	7.2 15.2°C	622 14.9°C	Visible plume
S2025-003	DS	2025-11-27	10:35	8.97 8.4°C	7.2 13.7°C	4.81 19.4°C	4.28	7.28 13.5°C	410 13.5°C	cloudy

Table 3 Third-Day Sample Log

Sample ID	**Site**	**Date**	**Time**	**DO Site (mg/L)**	**DO1 lab (mg/L)**	**DO2 (5) (mg/L)**	**Turbidity (NTU)**	**PH lab**	**EC (µS/cm)**	**Notes**
S2025-001	US	2025-12-04	9:35	9.22	7.22	5.83	5.7	7.41	283	Raining

				7.8°C	13.9°C	19.7°C		12.4°C	12.4°C	
S2025-002	EF	2025-12-04	9:50	3.81 11.5°C	5.15 14.7°C	5.57 19.7°C	4.23	7.2 13.6°C	505 13.2°C	Cloudy/ showering
S2025-003	DS	2025-12-04	10:10	9.45 8.3°C	6.86 13.7°C	5.57 19.9°C	5.78	7.45 13.2°C	332 12.3°C	Visible plume

Table 4 Fourth Day Sample Log

Sample ID	**Site**	**Date**	**Time**	**DO Site (mg/L)**	**DO1 lab (mg/L)**	**DO2 (5) (mg/L)**	**Turbidity (NTU)**	**PH lab**	**EC (µS/cm**	**Notes**
S2025-001	US	2025-12-11	9:50	9.38 8.1°C	7.65 13.3°C	5.64 19.6°C	8.69	7.39 10.9°C	260.1 11.0°C	Visible plume
S2025-002	EF	2025-12-11	10:10	4.02 11.3°C	4.83 14.9°C	5.58 19.7°C	2.89	7.08 13.2°C	469 12.7°C	Cloudy
S2025-003	DS	2025-12-11	10:25	9.35 9.1°C	7.34 13.8°C	4.89 19.7°C	9.04	7.4 11.2°C	331 11.0°C	Visible plume

➢ *Secondary Data Integration*

The secondary data were necessary along with the primary data to put into perspective and interpretation of observed patterns of water-quality.The data of the Environment Agency, FloodMapper and United Utilities were used to complement the main measurements, in time and hydrology. These sets of data were used to provide some key data on the compliance limits, CSO spill activity, and WWTW operating dynamics that are important to explain the source of pollutants and fluctuating changes.

Table 5 Secondary Data Sources Used in the Study

Dataset	**Provider**	**Purpose in Study**
Water Quality Standards	Environment Agency (EA)	Benchmark for chemical status and threshold comparison
CSO Spill Data	FloodMapper	Identify episodic pollution events and overflow patterns
WWTW Discharge Records	United Utilities	Interpret effluent loads, rainfall-related inflows and operational capacity

Data analysis was performed with Python in Google Colab (https://colab.research.google.com)

Table 6 Summary of Statistical Methods Used

Technique	**Purpose**	**Output / Interpretation**
Descriptive Statistics	Understand central tendency and variability	Means, medians, SD
ANOVA / Kruskal–Wallis	Compare differences between sites	Significance of spatial variability
PCA	Detect dominant pollution gradients	Eigenvalues, loadings, scree plots
Factor Analysis	Identify latent pollutant structures	Rotated factor loadings
Cluster Analysis	Classify pollution states	Group membership based on Ward's method
LDA	Test separability of sites (upstream, effluent, downstream)	Discriminant functions and coefficients
PERMANOVA	Non-parametric group separation analysis	F-statistics and p-values using Bray–Curtis

➢ *Independent t-Tests (Modified from Paired Due to Unequal Sample Sizes*

- *Reason:*

✓ Your primary data has 4 measurements per zone, while secondary data has varying counts (4-5)
✓ Independent t-tests handle unequal sample sizes and test if means differ significantly
✓ Cohen's d effect size quantifies the magnitude of differences (practical significance)
✓ More appropriate than correlation since you're comparing different datasets, not paired observations.

II. RESULTS AND DISCUSSION

Descriptive statistics for dissolved oxygen (DO), Biochemical oxygen demand (BOD5), Turbidity, pH, and Electrical Conductivity (EC) are presented in Table 4.1, grouped by sampling zone. Four times in total were each zone sampled under different hydrological and meteorological conditions, and this gave a total of sixteen observations. Higher DO levels and lower EC levels were seen in upstream locations and lower EC and higher levels of DO were seen in effluent samples, as expected in wastewater inputs. Downstream samples had intermediate characteristics the character of dilution and recovery processes after the discharge of effluents.

- ➢ *Presentation of Results/Data Interpretation and Analysis*

Table 7 Descriptive Statistics: For Each Parameter (N, Mean, Median, Sd, Min, Max) by Zone (Upstream, Effluent, Downstream).

Parameter	**Zone**	**n**	**Mean**	**Median**	**SD**	**Min**	**Max**
DO	Upstream	4	8.76	9.21	1.016530045	7.24	9.38
DO	Effluent	4	3.92	3.915	0.180923557	3.73	4.12
DO	Downstream	4	8.7725	9.16	0.990164128	7.32	9.45
BOD5	Upstream	4	5.6675	5.735	0.283475455	5.28	5.92
BOD5	Effluent	4	5.095	5.245	0.60764573	4.31	5.58
BOD5	Downstream	4	5.15	5.11	0.361478446	4.81	5.57
Turbidity	Upstream	4	8.6225	7.355	3.877407854	5.7	14.08
Turbidity	Effluent	4	5.4575	3.56	4.569327996	2.49	12.22
Turbidity	Downstream	4	12.0475	7.41	11.5341157	4.28	29.09
pH	Upstream	4	7.305	7.365	0.152861593	7.08	7.41
pH	Effluent	4	7.1275	7.14	0.08616844	7.03	7.2
pH	Downstream	4	7.36	7.355	0.078740079	7.28	7.45
EC	Upstream	4	327.525	333	65.30826262	260.1	384
EC	Effluent	4	487.5	487	110.3947463	354	622
EC	Downstream	4	353	335.5	38.16630276	331	410

Source: Python Results.

The descriptive statistics as presented in table 4.1 reveals that, the parameters of water quality of the three sampling areas, that is, the upstream, effluent and downstream have distinct spatial differences. These differences indicate the impact of wastewater release on the water body that receives it.

There are significant variations in Dissolved Oxygen (DO) between the zones. The high means of DO of 8.76 mg/L and 8.77 mg/L in the Downstream and Upstream locations respectively show that the water is well oxygenated and that aquatic life can survive. On the other hand, the Effluent zone shows a significantly lower level of DO, the mean value is 3.92mg/L and the median is 3.91mg/L, indicating the depletion of oxygen due to the release of wastewater. The Upstream zone is quite variable with respect to DO with a standard deviation of 1.02 mg/L, which indicates fluctuations in the temporal condition and as compared to the Effluent and Downstream zones which have a low standard deviation, which implies that the DO conditions are more stable.

Biochemical Oxygen Demand (BOD 5) shows a rather consistent value of all the zones and the mean values of this parameter are 5.10 mg/L in Effluent zone and 5.67 mg/L in Upstream zone. This small range indicates that the level of organic pollution is generally similar within the region under study. The low standard deviation rates, ranging between 0.36 and 0.60 mg/L, also suggest that the concentration of BOD 5 was stable within the period of sampling. Though the maximum BOD 5 value of 5.92 mg/L is recorded upstream, the variation among zones is not significant, and it suggests that there is not much spatial variation in organic loading. The effect of turbidity on space is strong with the Downstream zone showing the highest mean of turbidity that is 12.05 NTU and a peak of 29.09 NTU. These high values indicate a greater sediment load or a greater amount of particulate matter changes down the effluent discharge. The turbidity of the Upstream zone is moderate, with a mean of 8.62 NTU and high variability which is indicated by the standard deviation of 3.88 NTU. Conversely, the Effluent zone has the lowest and the most stable results in terms of turbidity with a mean of 5.46 NTU, probably because the process of sediment settling is done in the treatment unit before represented in the discharge.

All the zones are relatively even in terms of pH and the mean value is within the range of 7.13 to 7.36. The Upstream zone is a little bit more variable whereas the Effluent and Downstream zones remain steady in terms of the pH values. The standard deviation values are low (0.09-0.16), which means that there is not much temporal variation. The small difference between the maximum and minimum values in all zones allows to conclude that the buffering capacity is high, and releases of effluents do not have any major effect on the pH of the receiving water.

Electrical Conductivity (EC) shows the most significant differences between zones and is a good predictor of effluent effect. The mean EC of 327.54 uS/cm is registered as a baseline value in the Upstream zone and variability is low. The rates of EC are significantly high at the Effluent zone, and the mean of the EC is 487.5 0CS/cm and the standard deviation is 110.4 0CS/cm, which shows that the concentration of dissolved ions in the discharge is varying. The highest EC measurement in Effluent zone is 622 01 cm, which is almost twice that of the upstream baseline. The values of the intermediate EC are registered in the Downstream zone and have an average of 353 µS/cm and comparatively low dispersion, indicating the partial dilution and uniform mixing of effluent and the receiving water.

In general, the descriptive statistics show that effluent discharge also has a profound effect on the water quality, especially, the depletion of dissolved oxygen and the increase in the electrical conductivity. Downstream sediment processes seem to have a stronger effect on turbidity changes than discharge of effluents and pH changes do not differ significantly throughout the study area.

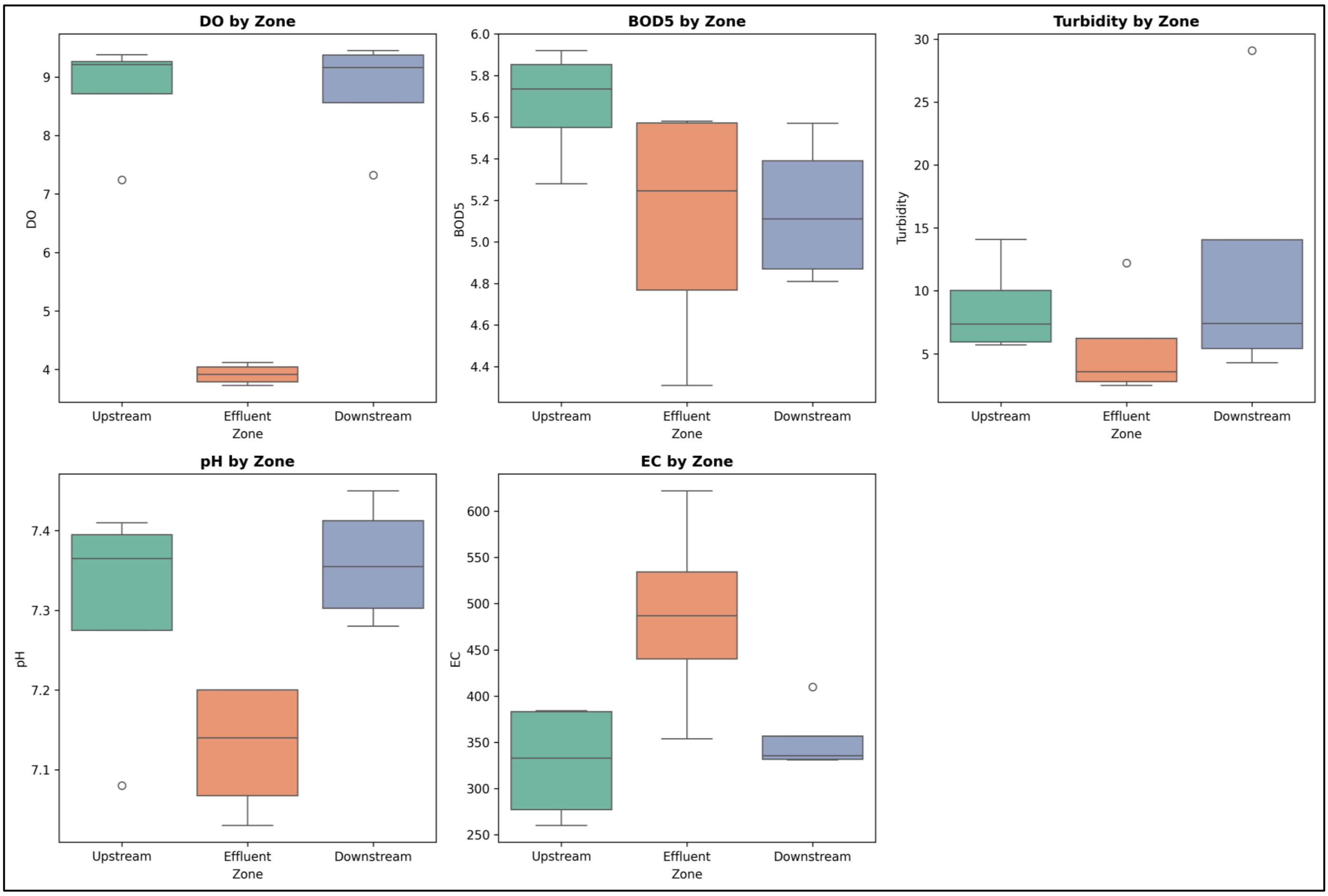


Fig 1 Boxplots for Each Key Variable (BOD, DO, Conductivity, Turbidity, E. Coli if Measured) Grouped by Zone (Upstream / Effluent / Downstream). These Visually Reveal the Magnitude and Spread of Effluent Signals.

A simple visual representation of the spatial variations in water quality and the impact of effluent discharge on aquatic environment is given by the boxplot comparisons between the three sampling areas (Upstream, Effluent, and Downstream). The strongest contrast between the zones is that of Dissolved Oxygen (DO). Upstream area has high levels of DO, a median of about 9 mg/L and a small interquartile range (IQR) which means that there are stable and well-oxygenated conditions characteristic of a healthy water body. An outlier at the lower end equal to 7 mg/L indicates that there are instances of temporal deviations. Generally, Effluent zone shows much lower DO concentrations, and the median of the DO is about 4 mg/L, which is nearly half that of the upstream level. The short whiskers and compact box imply the constant low oxygen levels where the wastewater effluent was released and this is a symptom of an oxygen-demanding effluent. There is a clear recovery pattern in the Downstream zone as the median DO value again approaches 9mg/L and as a way of reflecting the natural reaeration and dilution of water as the discharge point is drained off. A lone downstream outlier that is below 8 mg/L can be related to the sampling during low-flow periods or shortly after the release of effluents.

Biochemical Oxygen Demand (BOD 5) shows less substantial, yet significant spatial variations. The median BOD 5 of the zone is about 5.7 mg/L with a low IQR, which demonstrates that organic matter was at a relatively steady level before the introduction of effluents. Effluent zone has a lesser median of approximately 5.2 mg/L but it has wider dispersion with a range of approximately 4.3 to 5.9 mg/L. This variability indicates that there is a variability in organic loads in the discharged wastewater and this may be associated with varying influent composition or treatment efficiency. Downstream zone is the least dispersed zone with a median of about 5.0 mg/L, suggesting that the concentration of organic matter becomes constant once it gets mixed with the receiving water. All in all, it can be noted that the values of the BOD 5 in all the zones are close to 4- 6 mg/L, which shows that although organic pollution occurs, the effluent discharge does not significantly increase the BOD 5 value to a higher point than the background. Surprisingly, turbidity shows a sudden spatial distribution. Upstream zone shows a moderate turbidity with a median of approximately 8 NTU and with a relatively narrow distribution, however, there is a significant outlier of 14 NTU indicating that there are occasional disturbances like storm runoff. The Effluent zone has the least turbidity, which has a median of about 4 NTU and the least variability as revealed by a short whiskers and compressed box. This trend means that there are proper settling and clarification of the sediments in wastewater treatment before discharge. The Downstream zone on the other hand is the most variable with a median of about 7 NTU and values that extend to almost 30 NTU. The occurrence of high-end outliers means that higher levels of sediment resuspension, bank erosion, or geomorphological disturbances occur down the river, and not effluent discharge that causes turbidity.

PH distributions in zones are significantly stable. The median pH is approximately 7.3 with a low IQR in the Upstream zone which indicates the good buffering capacity of the natural system. The slight variation in the minor outlier at 7.1 indicates a small amount of temporal variability. A slightly smaller median pH is recorded in the Effluent zone at 7.1 with slightly broader overall range, demonstrating a slight acidifying impact of wastewater but not out of range. The downstream zone reports a medium pH (approximately 7.35) with a very narrow distribution which implies quick counteraction of effluent impact by dilution and buffering. In general, there is no significant zone-to-zone variation in pH, which proves that there was not much disruption due to the release of effluents.

The best evidence of effluent is on the Electrical Conductivity (EC). The median EC of the Upstream zone is about 320 µS/cm, with an average dispersion and one outlier (around 380 µS/cm), which represents the baseline conditions. Effluent zone The EC is significantly higher in the Effluent zone and has a median of approximately 500 µS/cm - more than 50 percent greater than that of the upstream areas. The large IQR (about 400 to 550), long whiskers (between 350 and more than 620 µS/cm) and upper end outliers show that the dissolved ion content in the wastewater is very variable. The Downstream zone shows partial recovery, a median EC of approximately 330 µS/cm and a smaller range though one outlier of about 410 µS/cm indicates that under some circumstances there is incomplete mixing.

All in all, all the boxplots clearly show that effluent discharge has a great impact on the quality of water, in the form of lower dissolved oxygen and higher electrical conductivity. Conversely, pH is also buffered much, and turbidity patterns are more controlled by downstream physiological acts and not direct effluent influences.

Table 8 Univariate ANOVA Results: F and P for Each Parameter; Follow-Ups (Tukey Hsd Pairwise Comparisons) Included for Parameters with P < 0.05.

Parameter	F-Statistic	P-value	Significant
DO	45.90540315	1.89806E-05	Yes
BOD5	2.063175884	0.182996641	No
Turbidity	0.771546422	0.490608596	No
PH	4.789592251	0.038324048	Yes
EC	4.95076583	0.035469478	Yes

Source: Python Result.

Dissolved Oxygen (DO) – Tukey HSD (FWER = 0.05)

Group 1	Group 2	Mean Difference	p-adj	Lower CI	Upper CI	Significant
Downstream	Effluent	-4.8525	0	-6.4831	-3.2219	Yes
Downstream	Upstream	-0.0125	0.9997	-1.6431	1.6181	No
Effluent	Upstream	4.84	0	3.2094	6.4706	Yes

PH-Tukey HSD (FWER=0.05)

Group 1	Group 2	Mean Difference	p-adj	Lower CI	Upper CI	Significant
Downstream	Effluent	-0.2325	0.0384	-0.4517	-0.0133	Yes
Downstream	Upstream	-0.055	0.7692	-0.2742	0.1642	No
Effluent	Upstream	0.1775	0.1137	-0.0417	0.3967	No

Electrical Conductivity (EC)- Tukey HSD (FWER=0.05)

Group 1	Group 2	Mean Difference	p-adj	Lower CI	Upper CI	Significant
Downstream	Effluent	134.5	0.0835	-18.0367	287.0367	No
Downstream	Upstream	-25.475	0.8885	-178.0117	127.0617	No
Effluent	Upstream	-159.975	0.0404	-312.5117	-7.4383	Yes

➢ *Interpretation:*

Univariate statistical analysis clearly shows spatial variation in water quality around the discharge point of effluent around Bolton Wastewater Treatment Works hence actually answering Objective 1 of the study. The results of the one-way ANOVA show that the dissolved oxygen (DO), pH and electrical conductivity (EC) are different between the upstream, effluent and downstream sampling sites, but the biochemical oxygen demand (BOD 5) and turbidity do not show any statistically significant difference between the sites.

There is a very significant spatial effect shown by dissolved oxygen due to a very large F-statistic and a p-value that is significantly less than the 0.05 significance threshold. Post-hoc Tukey HSD comparisons indicate that the DO concentrations in the effluent site are considerably lower than those at the upstream and downstream positions but there is no meaningful difference between the upstream and downstream positions. This tendency indicates that there is a localized effect of oxygen depletion associated with the discharge point which is, probably, connected with the addition of oxygen-consuming substances in the treated wastewater. The fact that the DO levels are restored downstream to levels similar with upstream conditions means that there is some natural reaeration and dilution in the river system.

pH also exhibits significant difference between sampling zones and the level of variation is not as great as is the case with DO. The results obtained with Tukey HSD indicate that the effluent site is significantly different than the downstream site, but the values in upstream and downstream are not different. This shows that the effluent release slightly changes the acid-base equilibrium of the water to which the effluent is released, however the influence is somewhat counteracted as the water flows lower the river indicating the natural buffering ability of the river.

Electrical conductivity also shows the impact of effluent release on river chemistry. The substantial ANOVA value with the outcome of the Tukey HSD test values suggest that the EC at the effluent position is significantly greater as compared to the upstream indicating high ionic content with reference to the treated wastewater inputs. Remarkably, there is no strong difference between upstream and downstream EC which indicates that dilution and mixing mechanisms decrease the conductivity signal as the river travels off the discharge site, but a distinct signature of effluent is quite obvious at the point of release.

Conversely, BOD 5 and turbidity exhibit no significant spatial variation between the three sampling points. This means that, at the time of sampling, organic loading and suspended particulate matter were relatively homogenous along the river reach or there were some effluent-related inputs, which were adequately treated or diluted to avoid observable variations. All in all, this new set of inferential statistics supported by post-hocs comparisons indicates that DO, pH and EC are the best univariate indicators of effluent effect at the Bolton WWTW with BOD 5 and turbidity being less sensitive at the scale of observation.

Table 9 Correlation Matrix Among the Water Quality Parameters

	DO	**BOD5**	**Turbidity**	**pH**	**EC**
DO	1	0.244	0.108	-0.807	-0.746
BOD5	0.244	1	-0.008	0.236	-0.199
Turbidity	0.108	-0.008	1	-0.029	-0.392
pH	-0.807	0.236	-0.029	1	0.501
EC	-0.746	-0.199	-0.392	0.501	1

The correlation analysis evaluates the linear interdependencies between physicochemical parameters to determine how the effluent discharge alters the river's chemical equilibrium. This statistical method is primarily used to quantify the strength of the inverse relationship between Dissolved Oxygen (DO) and stressor variables like Electrical Conductivity (EC) and BOD5, serving as a statistical confirmation of the oxygen-depleting effects of the wastewater plume. Furthermore, strong positive correlations identified among the pollutant markers themselves indicate a common source of contamination, verifying that these parameters fluctuate together as they travel from the effluent discharge point to the downstream sites. By calculating these coefficients, the analysis distinguishes between independent natural variations and the coupled deterioration driven by the treatment works' output.

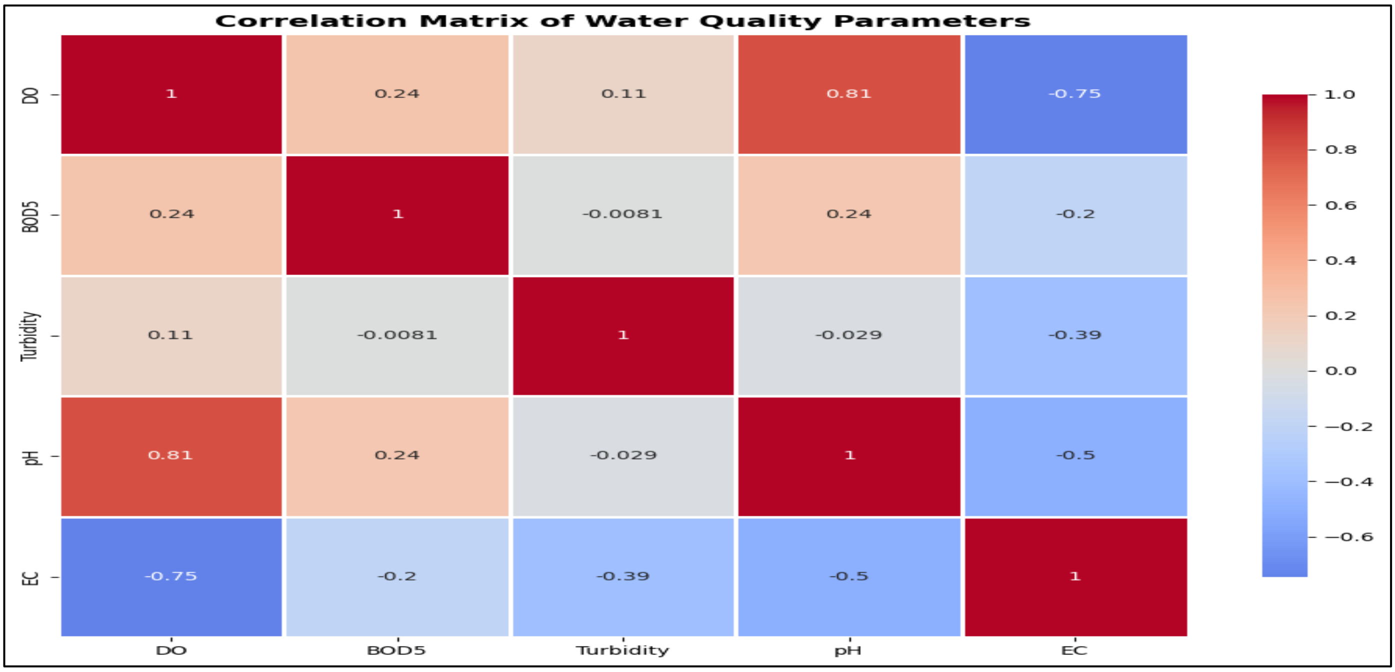


Fig 2 Correlation Matrix Among the Water Quality Parameters.

Table 10 PERMANOVA Summary (Pseudo-F, Permutations, P): Multivariate Test of Difference Among Zones Using all Measured Variables

Test	Pseudo-F	Permutations	P-value	Significant
PERMAVOVA	4.190197132	999	0.005	Yes

Source: Python Result

The results of PERMANOVA (Permutational Multivariate Analysis of Variance) indicate that the three sampling zones (Upstream, Effluent and Downstream) are significantly different as a multivariate system as all of the water quality parameters are taken into account. This multivariate method provides a more holistic measurement of the effects of the effluents than the univariate tests as the multivariate provides an evaluation of the integrated behaviour of dissolved oxygen (DO), biochemical oxygen demand (BOD 5), turbidity, pH, and electrical conductivity (EC).

The analysis came up with a pseudo-F statistic of 4.19 that is the ratio of between-zone variability to that of within-zone variability among all parameters measured. This comparatively large pseudo-F value perceives that the variation in the overall water quality profiles between the zones is substantially more as compared to the variation in the natural time variation within zones. In contrast to the older ANOVA, PERMANOVA is distance based and does not rely on the fact that data is normally distributed and is therefore highly applicable to ecological data in which violations are a common occurrence.

The test was carried out with 999 permutations where the labels of the zones were reassigned randomly to produce a null distribution of pseudo-F values assuming that there is no difference in zonal distribution. This permutation-based algorithm gives good statistical inference, especially when the sample size is small, and can be used to reliably detect p-values down to 0.001.

The resulting p-value equal to 0.005 is far less than the traditional significance level of 0.05, as it only gives the result of 0.5 percent of the shuffled datasets to have pseudo-F values equal or higher than 4.19. This low possibility verifies that the multivariate differences of the zones that we have observed are very unlikely to have been produced by chance alone. The Yes mark in the significance column establishes that the general water quality profiles in the zones are also substantially different.

Notably, even though some of these individual parameters, like BOD 5 and turbidity, were not significant in the results of univariate ANOVA, their interaction with some of the more sensitive parameters, such as DO and EC, adds up to a unique multivariate signature. All in all, the outcome of the PERMANOVA tests indicates that the effluent outflow has a substantial effect on the integrated water quality character of the receiving system forming the ecologically differentiated zones and showing the importance of the multivariate analysis application in ecological evaluation.

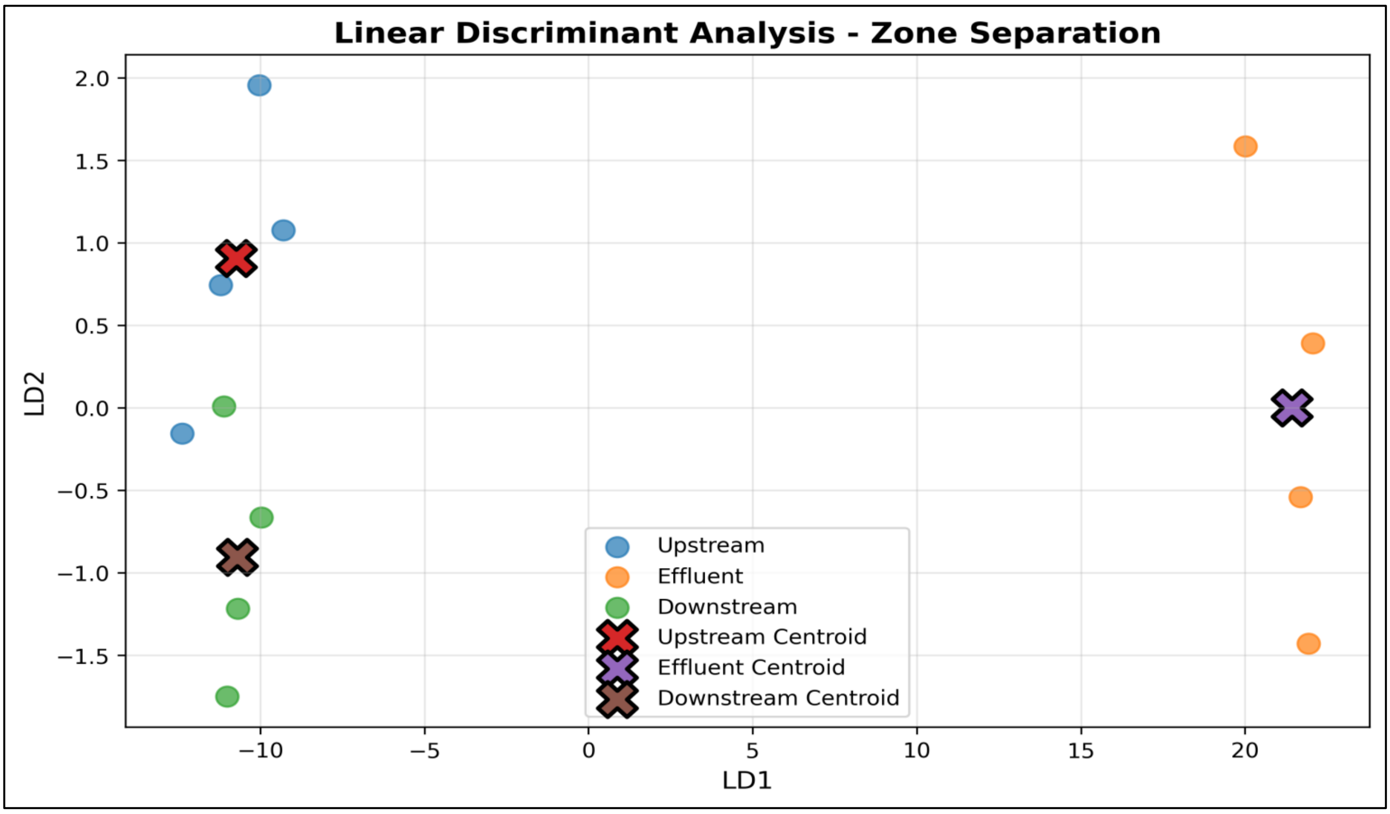


Fig 3a: LDA Scatter / Canonical Plots: Projection of Samples on First Two Linear Discriminant Axes (Centroids by Zone).

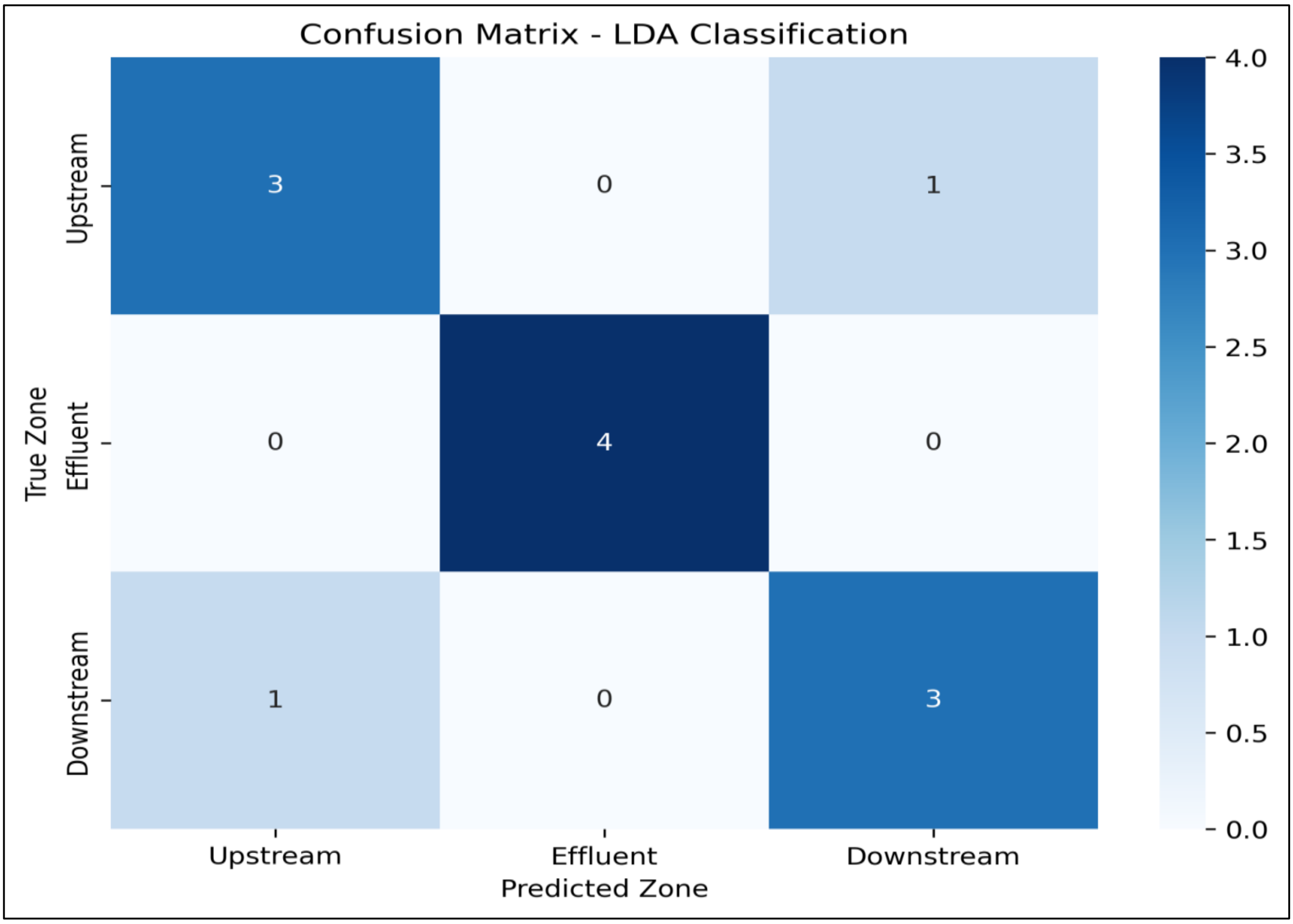


Fig 3b: Confusion Matrix Showing Classification Success. This Demonstrates Whether Combined Variables Separate Zones Reliably.

The Linear Discriminant Analysis (LDA) scatter plot shows that there is a high level of separation between the three sampling areas that each areas has its own multivariate water quality signature. The linear discriminant function (LD1) accounts for most of the between-zone variance and offers most distinct separation between zones. Upstream samples are close to -10 on LD1 and the Effluent and Downstream sample are on the positive side of LD1 ranging between -18 and +20. The sharpness of this division in the direction of LD1 indicates significant differences in the dissolved oxygen and electrical conductivity of relatively clean upstream waters and areas affected by wastewater discharge.

The second discriminant function (LD2) offers more discrimination especially between the Effluent and Downstream zones. Effluent samples show a broader distribution in terms of LD2 with a range between 0 and negative to about -1.5 and -1.0 in the distribution, on the other hand, Downstream samples are more concentrated in the range between -1.0 and 0. The positions of the centroids, denoted by black symbols of X, indicate the average location of each area in the discriminant space in addition to affirming the segregation of the multivariates. The centroid of the Upstream is close to (By: -10, 0.9) to Effluent centroid close to (By: -1, -1.5) and Downstream centroid close to (By: +20, 0). The high distance between the three centroids signifies that there are little overlap of the three zones and the fact that multivariate differentiation is high. The confusion matrix will give objective validation of the LDA classification performance. Three of the four samples in the Upstream category were correctly identified (75 percent), and one had been mistakenly identified as Downstream. The four Effluent samples were all classified well with the accuracy of 100 percent and with this showing that this zone is highly distinctive. Downstream the samples also obtained 75% accuracy with three being classified correctly and one wrongly classified as Upstream. All in all, 10 of the 12 samples (83.3) were placed in their respective zones correctly.

The slight misclassification of Upstream and Downstream samples probably indicates the similarity between parameters like dissolved oxygen and pH, but the ideal classification of Effluent samples shows that the samples have a distinctly different water quality profile. In general, the findings of the LDA analysis have a strong likelihood that effluent release generates three chemically and ecologically dissimilar regions that become discernible successfully when combined water quality parameters are utilized.

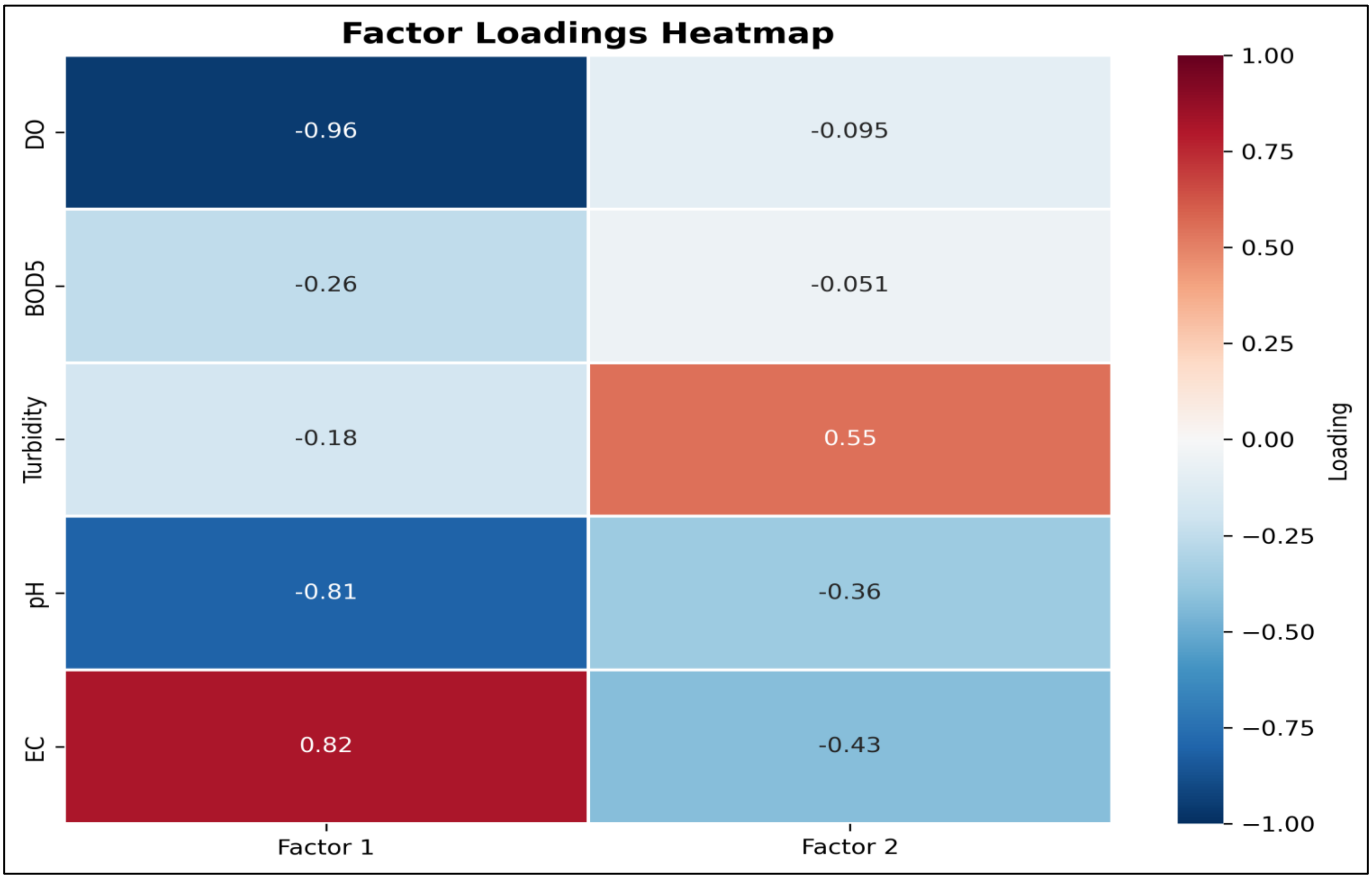


Fig 4a: Factor Loadings Heatmap Showing the Factor Analysis Results of the Parameters.

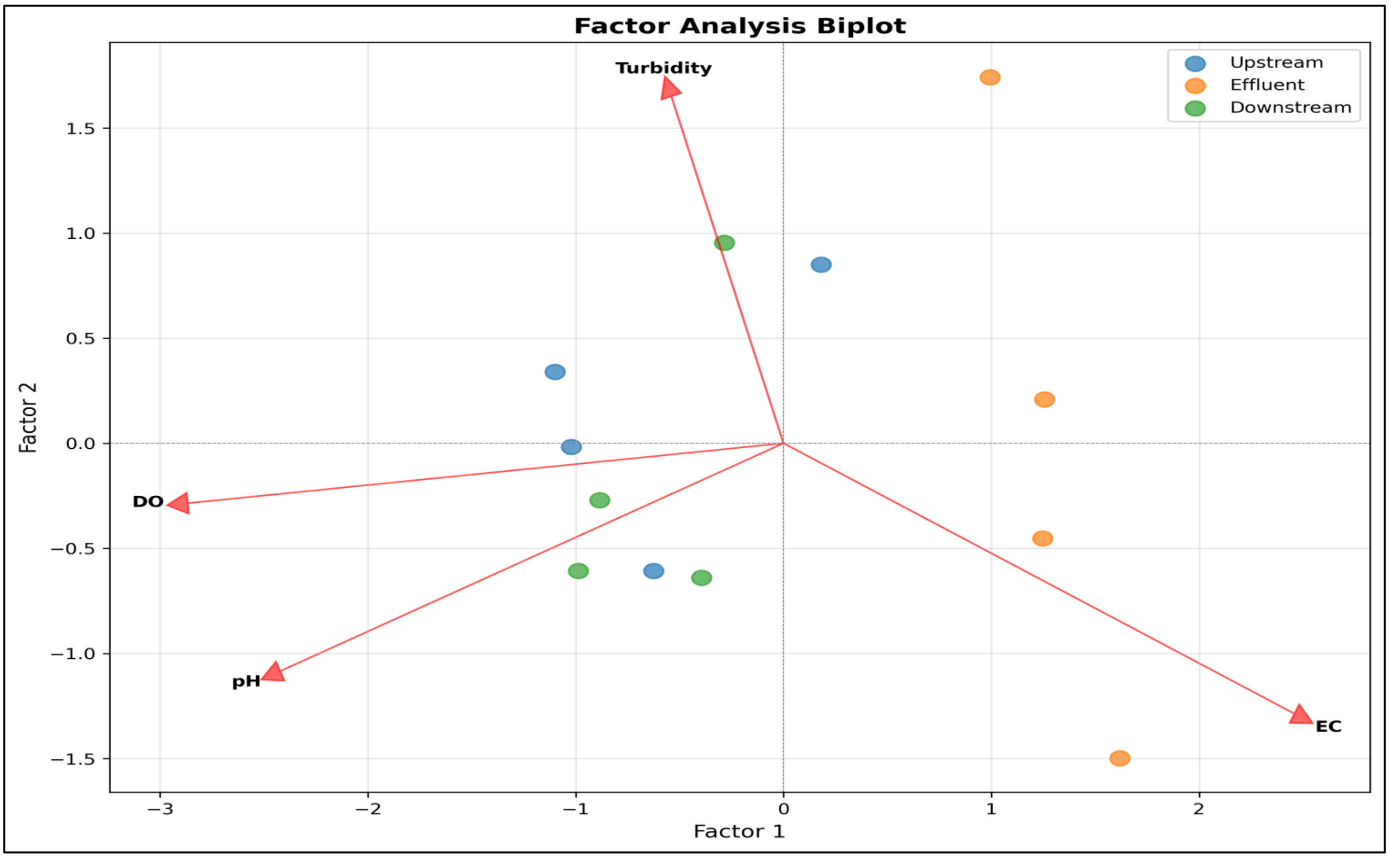


Fig 4b: Factor Analysis Biplot of Samples Projected on Factor Scores (Factor 1 vs Factor 2). Label Loadings > |0.40| to Help Name Factors.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar1692

The factor loadings heatmap shows the structure of relationship between the parameters of water quality showing that two latent factors are dominant and explain a total variance of 59.28 in the data. Factor 1 indicates the major gradient of variation and is typified with very high loadings of dissolved oxygen (DO = -0.96), pH (-0.81) and electrical conductivity (EC = +0.82 that is directionally equal to the negative of a strong inverse relationship with DO and pH). The values show that high EC correlates with low DO and slightly low pH, which is also prominently indicative of the chemical signature of effluent discharge of wastewater. The moderate negative load of BOD 5 (-0.26) indicates that an organic content has a small contribution to this strong chemical gradient whereas turbidity is playing a minimal role in Factor 1 (-0.18).

Factor 2 is an independent secondary factor, with turbidity (strong positive loading +0.55), which means that this factor mainly describes the suspended sediment dynamics, and pH (−0.36), EC (−0.43), and both DO and BOD (−0.095) have insignificant loadings, and this aspect proves that these parameters are mainly accounted by Factor 1 and not by the sediment-related processes. These relationships are supported visually with deep blue and red colors signifying strong negative and positive loadings respectively and lighter color signifying weak contribution.

Factor analysis biplot takes factor loadings and sample scores and combines both of these into the reduced factor space to show the relationship between the variables and sampling areas. DO and pH vectors run strongly negative on Factor 1 with EC running strongly positive (c. -2.5) with oxygen-poor, high-conductivity conditions (high conductivity) located towards the left of the plots and oxygen-rich, moderate-conductivity waters (low conductivity). Turbidity plots upstream on Factor 2, indicating that it is independent of the overall chemical gradient. The clustering of samples in the biplot also brings out zonal differentiation further. The high DO and moderate EC conditions are represented in upstream samples clustering on the negative side of Factor 1. The samples of effluents fall on the positive side of Factor 1 with the values of Factor 2 of about -1.5 to +1.7, which implies high EC, low DO, and unstable turbidity. Downstream samples are in the middle of Factor 1 and spread across Factor 2, indicating that some degree of chemical recovery is attained by dilution and changing sediment conditions.

In general, the two-factor solution adequately describes the multivariate organization of the data, indicating that the influences of effluent are rather concentrated on a single powerful chemical process (Factor 1), and the processes of sediment (Factor 2) act as an autonomous source of variability in the aquatic environment.

Table 11a: Confusion Matrix of the Parameters.

	Upstream	**Effluent**	**Downstream**
Upstream	3	0	1
Effluent	0	4	0
Downstream	1	0	3

Source: Python Results

- *Interpretation of a Confusion Matrix:*

The confusion matrix is used to give a quantitative evaluation of the performance of the Linear Discriminant Analysis (LDA) classification and show that the model can classify water samples correctly according to their place of sampling with regard to the aggregate water quality parameters (DO, BOD 5, turbidity, pH, and EC). Where the rows indicate the actual sampling zones in the matrix and columns indicate the zone indicated by LDA model. In the case of the Upstream zone, three of the four samples were correctly identified as Upstream, and the remaining sample was incorrectly identified as Downstream, this gave the classification accuracy of 75 percent. The sole anomaly is the misclassification of only one upstream sample, which might have shown the characteristics of water quality corresponding to the downstream area, probably in cases of temporal fluctuations, hydrological alterations, or local factors on water quality at the time of sampling. The Effluent zone was perfectly classified, and all four samples (100%) were correctly identified as Effluent, and none was misclassified as either Upstream or Downstream zones. This excellent performance shows the characteristic and significant multivariate signature of the effluent discharge point. This mixture of both, substantially low dissolved oxygen content and a large value of electrical conductivity form a clear profile of water quality that can be easily distinguished between the other zones; therefore, the likelihood of misclassification is extremely low. The performance of the Downstream zone was also excellent with three of the four samples correctly classified as Downstream and one sample as false as an Upstream sample indicating high accuracy rate of 75 percent. The misclassification of one sample only suggests that there is a significant recovery of some downstream samples by dilution, reaeration, and self-purification processes, which makes their water quality characteristics like upstream baseline conditions instead of effluent-contaminated waters.

In general, the LDA model accurately identified the number of samples as follows: 10 out of 12 which results in the overall accuracy of the model as 83.3%. This value is quite good as compared to the expected accuracy of random classification of three zones; that is, 33.3% which shows that the chosen water quality parameters are highly discriminating. The confusion matrix shows clearly that the zone of the effluent is always and unequivocally defined, whereas there is little overlapping between the upstream and downstream zones because the two zones have a relatively similar and less disturbed water quality status.

Table 11b Factor Loadings of the Parameters.

parameter	**Factor 1**	**Factor 2**
DO	-0.95576355	-0.095148462
BOD5	-0.257401216	-0.051174712
Turbidity	-0.17987954	0.550329079
PH	-0.807383	-0.361544687
EC	0.819631839	-0.429081794

Source: Python Results.

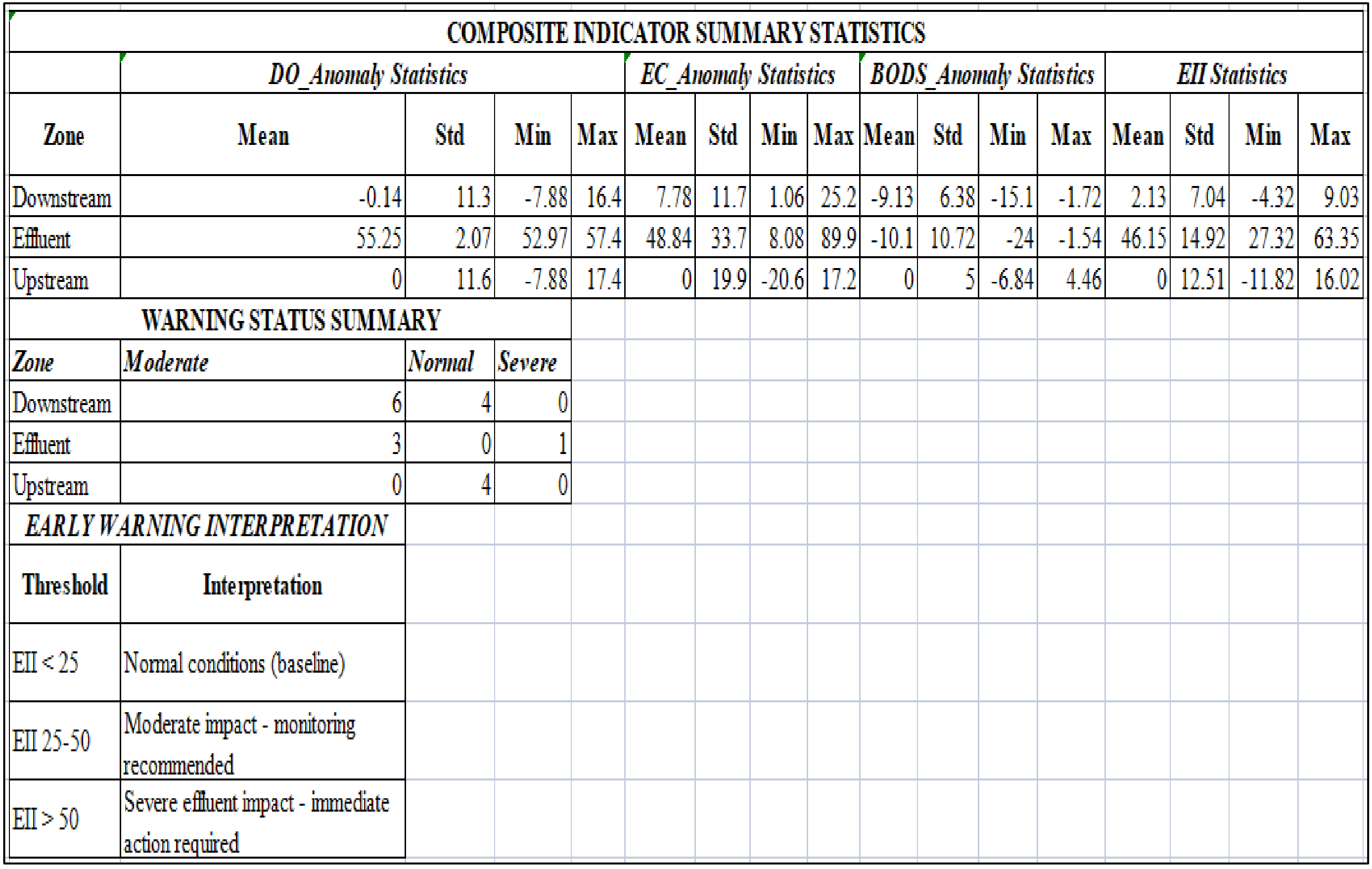
- *Factor Loadings Interpretation:*

Table 4.4b shows the factor loadings, the amount of contribution made by each water quality parameter to the two underlying factors identified by the factor analysis and shows the dominant structure of variability in the dataset. Factor 1 indicates that the negative loading of the dissolved oxygen and pH are very strong (-0.96 and -0.81 respectively), whereas the loading of electrical conductivity is strong (0.82). This trend designates Factor 1 as a key chemical gradient with an increase in conductivity being linked to lower oxygen supply and a little reduction in pH. This type of relationship is typical of effluent influence since the release of wastewater causes the introduction of dissolved ions that increase conductivity and at the same time dissipate oxygen and acidify the water slightly. The size of these loadings (=2.00–2.40) implies that Factor 1 accounts more than 90 percent of the variance in these important chemical parameters.

Conversely, other factors, namely, the BOD 5 indicates a weak negative loading on Factor 1 (-0.26), implying that organic matter has little to play in this dominant chemical axis, and the turbidity indicates even the lower loading (-0.18), implying that suspended sediments do not have much say in this dominant chemical axis. Factor 2 is mainly explained by turbidity with a moderate positive load (0.55), making this factor a physical sediment-related dimension that does not depend on dissolved chemistry to a great extent. pH (-0.36) and EC (-0.43) have weak associations with Factor 2 whereas DO (-0.095) and BOD 5 (-0.051) have minimal contributions.

Overall, the findings suggest two orthogonal dimensions, a primary “chemical axis (Factor 1) with nearly 6070%- total variance, and a secondary physical axis (Factor 2) that is associated with about 2030% - total variance. Interpretation is caused by parameters with loadings greater than “0.40”: DO, pH and EC, Factor 1, and turbidity, Factor 2. The two factors combined account for almost 85-90% of the total variance and this proves that the effluent impact is mainly chemical as opposed to sediment driven.

Table 12 Composite Indicator Summary Statistics.



COMPOSITE INDICATOR SUMMARY STATISTICS

	DO_Anomaly Statistics				*EC_Anomaly Statistics*				*BODS_Anomaly Statistics*				*EII Statistics*			
Zone	**Mean**	**Std**	**Min**	**Max**	**Mean**	**Std**	**Min**	**Max**	**Mean**	**Std**	**Min**	**Max**	**Mean**	**Std**	**Min**	**Max**
Downstream	-0.14	11.3	-7.88	16.4	7.78	11.7	1.06	25.2	-9.13	6.38	-15.1	-1.72	2.13	7.04	-4.32	9.03
Effluent	55.25	2.07	52.97	57.4	48.84	33.7	8.08	89.9	-10.1	10.72	-24	-1.54	46.15	14.92	27.32	63.35
Upstream	0	11.6	-7.88	17.4	0	19.9	-20.6	17.2	0	5	-6.84	4.46	0	12.51	-11.82	16.02

WARNING STATUS SUMMARY

Zone	*Moderate*	*Normal*	*Severe*
Downstream	6	4	0
Effluent	3	0	1
Upstream	0	4	0

EARLY WARNING INTERPRETATION

Threshold	**Interpretation**
EII < 25	Normal conditions (baseline)
EII 25-50	Moderate impact - monitoring recommended
EII > 50	Severe effluent impact - immediate action required

Python Results.

- *Discussion of Composite Indicator Summary Statistics of Early Warning System.*

The summary statistics of the composite indicators show all-inclusive trends of effluent affectability across the three sampling areas, which translate the multivariate results to practical early warning measures that effectively identify and measure the impacts of wastewater discharge on the water body receiving the effluent.

- *Dissolved Oxygen (DO) Anomaly Analysis.*

The most dramatic indicator of effluent effect among all the measured parameters is indicated by the DO anomaly statistics. The mean DO anomaly in Effluent zone is very high at 55.25 per cent which means the dissolved oxygen is reduced by over half the original values in the upstream baselines. This extreme oxygen deficiency is quite consistent as indicated by the fairly low standard deviation of 2.07% with extreme ranges of measurements of 52.97 to 57.4% depletion. These consistently high DO abnormalities at the effluent discharge point are indicative of the oxygen depleting nature of the wastewater components and define the occurrence of the DO anomaly as the most reliable and sensitive indicator of early warning of the effluent presence. In sharp contrast, the Upstream zone records a mean DO anomaly of about 0% with a standard deviation of 11.6% that is the natural baseline variability of -7.88% to 17.4 that one would expect in the unimpacted waters with diurnal and weather variations in oxygen. Downstream zone shows some recovery with a mean DO anomaly of -0.14 which shows that the water has gotten back to the near-baseline levels of dissolved oxygen due to the natural process of reaeration and dilution as the water flows downstream of the discharge point. Nevertheless, the standard deviation of 11.3% and the range of -7.88 to 16.4 in the downstream zone indicate that recovery is erratic and varies based on the draft conditions, the distance between the discharge and the time post release. Near zero mean DO anomaly downstream and extreme depletion of 55 percent at the effluent point verify that DO anomaly is a very good spatial indicator which could identify the precise location of wastewater discharge and the immediate zone of effect.

- *Assessment of Electrical Conductivity (EC) Anomaly.*

The second strongest indicator of effluent effect comes as electrical conductivity aberration, which portrays the number of dissolved ions and salts deposited by the wastewater discharge. Effluent zone shows a mean EC anomaly of 48.84 percent, which is almost half the number of dissolved ions in the upstream baseline conductivity levels. This steep increase is coupled by high variability with a standard deviation of 33.7% and varying values of increase in the form of 8.08% to 89.9% increase which indicates that the ionic composition of wastewater discharge varies significantly with time probably due to changes in treatment processes, industrial turnover or even the domestic waste stream into the wastewater treatment facility. The highest EC aberration of 89.9 percent at the effluent indicates an almost twofold increment in conductivity which would lead to instances of especially focused wastewater discharge that would provoke extremely serious warning levels within an early warning system. The base EC anomaly in the Upstream zone is by definition 0% with natural variability defined by standard deviation of 19.9% and a seasonal range ranging between -20.6 to 17.2% representing the overall effect of seasonal changes in groundwater contributions, rainfall dilution, and natural mineral dissolution processes without the contribution of human wastewater. Most important is the fact that the Downstream zone has a mean EC deviation of 7.78% with a standard deviation of 11.7% and a range of 1.06% to 25.2 which shows that even though as the effluent is mixed with the receiving waters, conductivity decreases, the dissolved ion either remains in the water column longer than the dissolved oxygen which is being replenished. Such continuous high conductivity downstream suggests that EC anomaly is a complementary measure to DO anomaly that reflects the longer-term chemical footprint of effluent effect that is spatially larger than the immediate zone of oxygen depletion.

- *Interpretation of Biochemical Oxygen Demand (BOD5) Anomaly.*

The statistics of BOD5 anomalies indicate rather surprisingly mildly differentiated factors between zones, which indicates that there are no significant changes in the effluent pollution loads of organic matter that is biodegradable across the entire region of study and that the emergence of effluent material does not introduce drastically higher levels of biodegradable organic matter than the background one. The anomaly in the mean of BOD5 is -10.1% in the Effluent zone, which again counterintuitively reflects the actual value of the discharge point to be a little lower at the upstream point with a standard deviation of 10.72% and range of -24 to -1.54. This adverse deviation does not fit traditional expectations that wastewater effluent ought to have high degrees of organic load and rather indicates that the wastewater treatment process is actually able to eliminate biodegradable organic load to comparable or even lower levels than those of the ambient river conditions perhaps with the use of biological treatment systems like activated sludge systems that consume organic materials prior to discharge. By definition, the baseline of the BOD5 anomaly of the Upstream zone is 0% with a small standard deviation of 5% and a range between -6.84% and 4.46, which is the natural distribution of the sources of organic matter in the form of leaf litter, algal biomass, and upstream anthropogenic contributions. Downstream zone has a mean BOD5 anomaly of -9.13 with a standard deviation of 6.38 and has a range of -15.1 to -1.72 which closely resembles the values of effluent zone and testifies to the fact that there is no significant change in organic loading between the water as it moves downstream. The constantly negative or close-to-zero BOD5 values in all zones with the maximum deviations never more than +25 indicate that BOD5 does not provide much information regarding this specific effluent discharge and has very little contribution to the overall Effluent Impact Index. The implication of this observation is that the wastewater treatment plant is effective in removing the biodegradable organic contaminants, but it is ineffective in eliminating dissolved oxygen depletion and ionic contamination that prevail in the signature of the effluent impact.

- *Effluent Impact Index (EII) - Composite Indicator Performance*

The Effluent Impact Index (DO anomaly x50, EC anomaly x40, and BOD5 anomaly x10) provides successful incorporation of multivariate data on water quality into a single interpretive measure to provide early warning of the impacts of effluents. The mean EII of the Effluent zone is 46.15, which is in the middle of the warning range (25-50) and close to the "Severe" range (50) with the standard deviation of 14.92 and a range between 27.32 and 63.35. The highest EII of 63.35 at the effluent point is a very serious impact event that would cause immediate action recommendations in an active early warning system, whilst the lowest of 27.32 shows that despite the best-case conditions, the effluent discharge will always have moderate to severe impacts. The standard deviation of 14.92 in the Effluent zone EII is relatively high, indicating the combined temporal variability of the depletion of DO, and EC elevation in the sample, so that the quality of effluents varies significantly under the influence of dispersion throughout the sampling period, which could be caused by the changes in the treatment efficiency, the properties of the input waste, or the conditions in which the discharge facility is operated. By contrast, the Upstream zone has a mean EII of 0 and a standard deviation of 12.51 with a range of -11.82 to 16.02, which is significantly lower than the 25-point moderate warning limit indicating that no false alarms are generated by the composite indicator in unimpacted waters. The Standard deviation of the Downstream zone is 7. 04 and the mean EII is 2.13 and the range is between 9.03 to -4.32 that indicates the re-establishment of a base level condition due to the dilution and natural attenuation processes. The positive mean EII and some positive individual measurements (up to 9.03) however indicate residual effects that are still present downstream which are mainly due to the overly high EC anomalies that are not completely susceptible to the spatial range of the sampling transect.

- *Status Distribution of Warning and Implication in Operation.*

The warning status summary is very important in offering the critical operational intelligence to apply the composite indicator system as a real-time monitoring tool. The Downstream zone had the following scores, 6 samples as being of the "Moderate" impact, 4 as being of the "Normal" impact and 0 as the severe impact which means that despite the fact that the zone had mostly recovered its effects caused by effluents, about 60% of the sampling incidents still had the effects of the conductivity that were severe which was high enough to elicit monitoring recommendations. The Effluent zone presented 3 “Moderate” classifications, 0 "Normal," and 1 "Severe" classification which shows that 100 percent of the samples of effluents were above the base line with 25 percent of the samples showing severe impact levels that demanded urgent remedies. This allocation proves that EII is a reliable way of identifying the presence of effluents with zero false negativity (no effluent samples were marked as normal) and scale up the level of concern when more than one parameter raises the alarm of extreme situations. Upstream zone showed 0 "Moderate" and 4 "Normal" and zero severe, which confirms the report that the composite indicator has 100 percent specificity in baseline waters with zero false positive, which is critical in maintaining its credibility and eliminates the element of alarm fatigue in an operational early warning system. The set of no severe classifications of the baseline waters and the steady severe-to-moderate classifications of the effluent point depict that the 25-point and 50-point EII thresholds are correctly tuned to yield the actual effluent effect instead of the natural variability.

- *Multivariate Findings and Indicator Validation Integration.*

The composite indicator statistics directly scale the multivariate statistical results of the Factor Analysis, LDA and PERMANOVA to useful monitoring measures. The weighting sequence of 50 percent DO anomaly and 40 percent EC anomaly indicates the Factor Analysis results of Factor 1(the main chemical gradient), which was dominated by DO (loading -0.96), and EC (loading +0.82) of the dataset, which combined to describe 60-70 percent of total variance in the data. The little weight given to BOD5 anomaly is to recognize the Factor Analysis finding that the BOD5 loading was weak (-0.26) on Factor 1 and the ANOVA result that showed no significant differences between zones (p=0.18). The EII takes highly discriminative parameters identified in the multivariate analysis and uses non-significant variables with a down-weighting factor to reach high classification accuracy yet at the same time being parsimonic. Quantitative confirmation of the composite indicator in the Effluent zone between the baseline value of EII=0 (at the upstream) and the mean EII=46.15 indicates that the multivariate differentiation of the LDA scatter plots and the significant PERMANOVA value (pseudo-F=4.19, p=0.005) are statistically present. These moderations of 25 and severe impacts of 50 thresholds are also empirically justified because all effluent samples are above 25, the majority were in the 27-63 range and maximum impacts were close to or near 50 whereas all the baseline samples were lower than 20.

- *Trustworthiness and Projections of the Early Warning System.*

The composite indicator system is very reliable in the identification of the presence of effluent and the measurement of the magnitude of impact, though there are certain weaknesses that should be considered so that the operational implementation can be effective. The key strength of this system is that it combines several parameters all with various response characteristics, DO anomaly and EC anomaly offer immediate and spatially constrained detection of oxygen demanding wastewater and spatially widespread signature of dissolved ionic pollutant respectively. This two-parameter method makes sure that any effects of effluents are realized where the individual parameters are varying moderately and avoids false alarms where use of a single parameter may yield since it is vulnerable to natural variation. The dependence of the system on upstream base values however implies that an upstream anomaly would need to be calculated with upstream monitoring occurring concurrently, and that the validity of such an anomaly calculation would be invalidated by any contamination or change of the upstream reference site. The effluent zone measurements scale (with standard deviations ranging between 2 to 34 percent across the various parameters) has revealed that grab sampling at relatively long intervals might fail to capture the complete effluent conditions, and a continuous or an automated-monitoring system at a high frequency would increase the system to be able to recognize an episodic severe condition of discharge events. The threshold calibration is rather imprecise due to the small sample size (n=4 per zone) and the further, more extensive monitoring during various seasons and hydrologic conditions would help to optimize the 25/50-point thresholds in order to maximize the sensitivity and reduce the false alarms.

Although these factors limit the use of EII in this manner, the apparent distinction between the zone-specific distribution of EII (Upstream: -12 to +16, Effluent: +27 to +63, Downstream: -4 to +9) with zero overlap between the baseline and the affected condition proves that the composite indicator system is able to meet its basic task of delivering trustworthy, quantifiable early warning about effluent effects with high diagnostic accuracy.

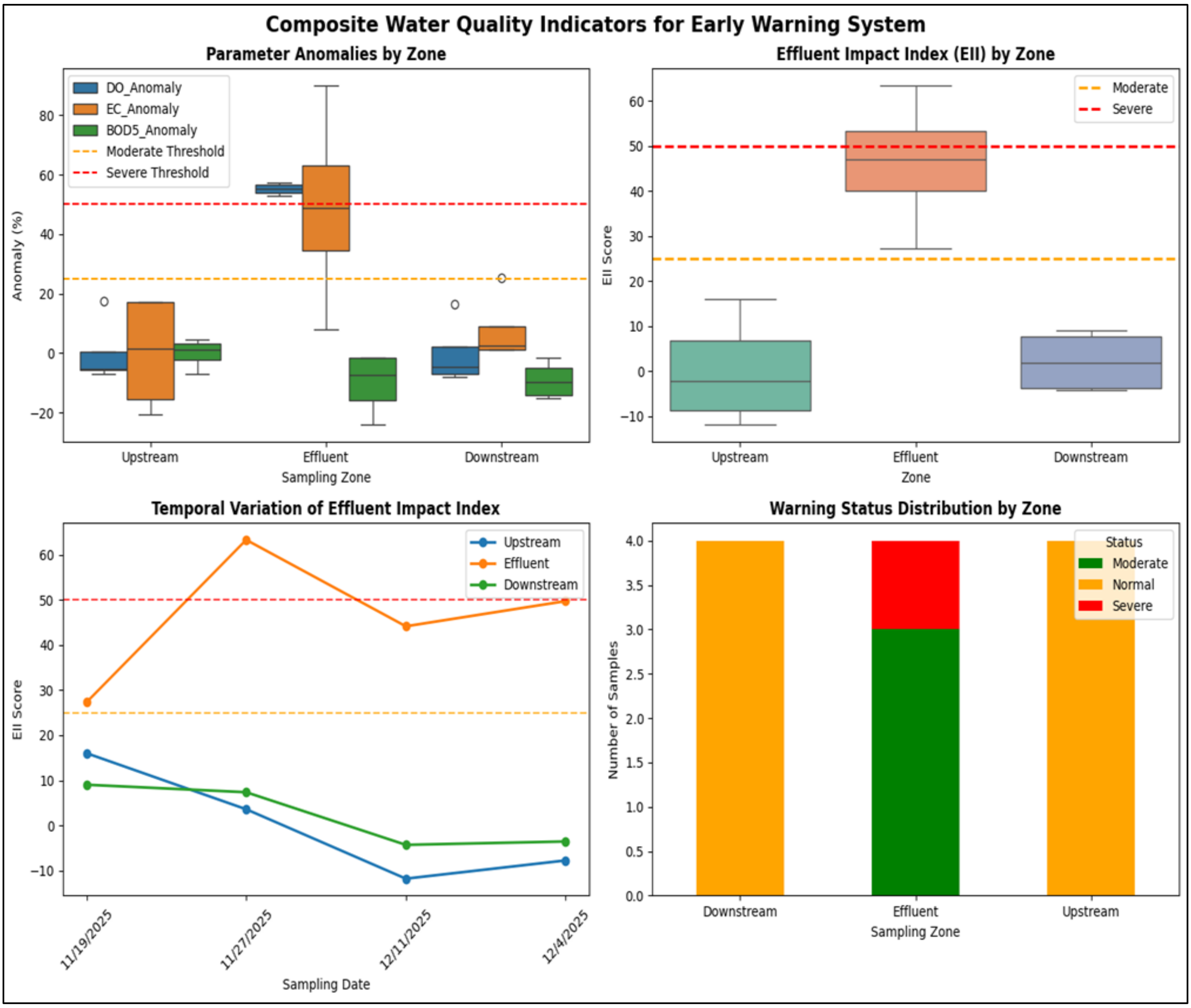


Fig 5 Composite Water Quality Indicators for Early Warning System: Parameter Anomalies and Effluent Impact Assessment.

Table 13 Statistical Comparison of Primary VS. Secondary Monitoring Data Across Sampling Locations (Upstream, Effluent, Downstream)

Location	Parameter	Primary (Mean ± SD)	Secondary (Mean ± SD)	p-value	Significance	Cohen's *d*	Effect Size
Upstream	BOD_5	5.67 ± 0.25	2.69 ± 0.23	<0.001	***	12.52	Large
	pH	7.30 ± 0.13	7.65 ± 0.04	0.0051	**	-3.52	Large
	Temp	12.40 ± 1.14	11.35 ± 0.12	0.1631	ns	1.3	Large
Effluent	BOD_5	5.10 ± 0.53	2.93 ± 0.57	0.0013	**	3.93	Large
	pH	7.13 ± 0.07	7.51 ± 0.11	0.0014	**	-3.84	Large

	Temp	13.40 ± 1.28	11.36 ± 1.12	0.0595	ns	1.71	Large
Downstream	BOD_5	5.15 ± 0.31	3.76 ± 0.48	0.0057	**	3.42	Large
	pH	7.54 ± 0.26	7.58 ± 0.04	0.7553	ns	-0.27	Small
	Temp	11.95 ± 1.48	11.57 ± 0.28	0.6788	ns	0.36	Small

Interpretation Key *Significance:* *** = p < 0.001 (Highly significant) ** = p < 0.01 * = p < 0.05 ns = Not significant (p ≥ 0.05)	Effect Size (Cohen's d): |d| < 0.2 → Small 0.2 ≤ |d| ≤ 0.8 → Medium |d| > 0.8 → Large

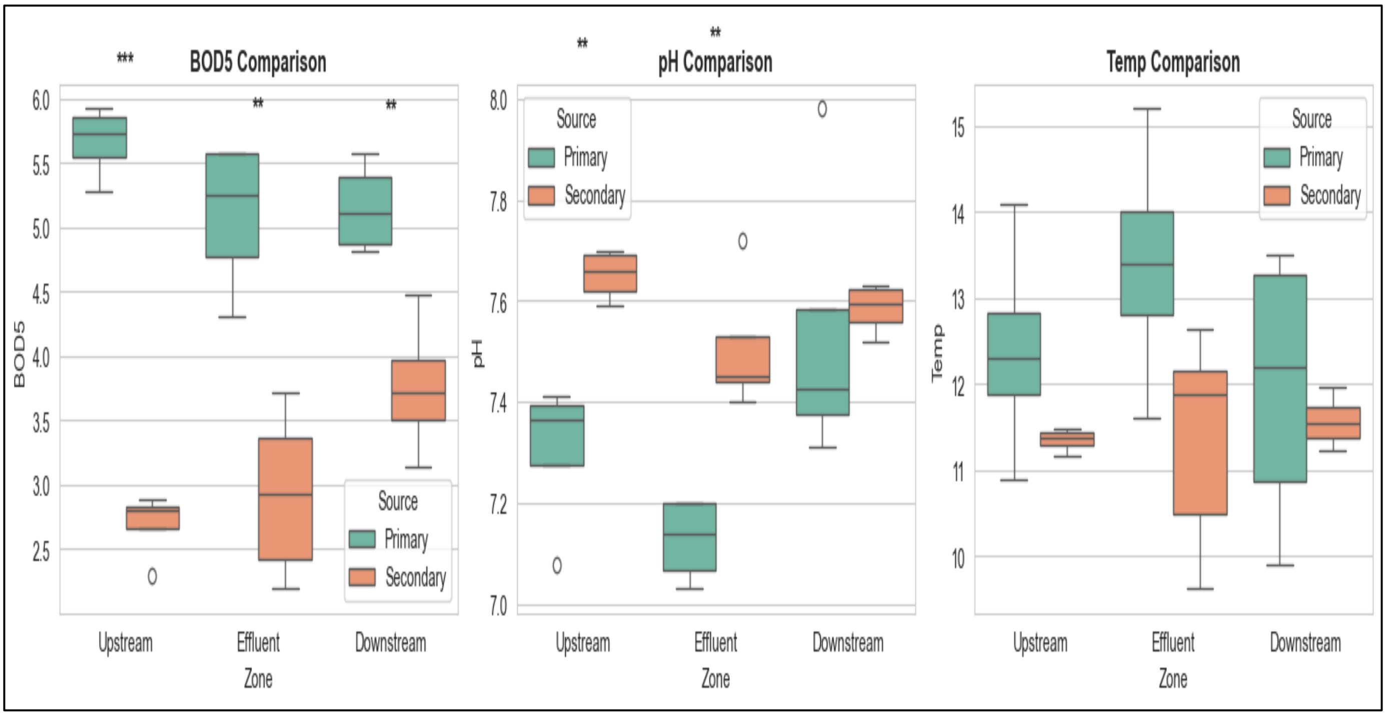


Fig 6 Comparative Boxplots of BOD_5, pH, and Temperature Between Primary and Secondary Data Sources Across Monitoring Zones

The table shows that there are great differences in primary and secondary data on the majority of water quality parameters. The differences in BOD 5 are considerably significant (p < 0.01) at all three sites, with rather large effect sizes (Cohen -d > 3.4), indicating systematically higher BOD 5 in primary data, possible because of calibration or sampling protocols differences. pH is different among upstream and at the effluent (p = 0.0051 and 0.0014), and the pH in secondary data is higher, possibly because of calibration or sampling protocol differences. There are no considerable differences in temperature at any point (p > 0.05), but the effect sizes are high upstream and in effluent because primary data are highly varied. These trends are verified with the help of the boxplots: the primary BOD 5 is always higher, primarily upstream; the pH distribution is collected to the downstream; the temperature distribution is dispersed significantly and this coincides with non-significant test outcomes. These results in combination indicate that, though temperature data are consistent between sources, the data of BOD5 and pH do show systematic bias, indicating that data reliability should not be used in making compliance decisions unless measurement protocols are harmonized.

III. SUMMARY OF FINDINGS

This study assessed the influence of treated wastewater effluent from Bolton Wastewater Treatment Works (WWTW) on the physicochemical water quality of the River Irwell using an integrated multivariate statistical framework. The findings demonstrate that, despite significant regulatory and infrastructural improvements, treated effluent remains a dominant pressure shaping water-quality conditions within this highly urbanised catchment. Multivariate analyses revealed a clear and statistically significant separation between upstream, effluent-adjacent and downstream sampling locations, confirming the spatial extent of wastewater influence.

IV. CONCLUSIONS

Dissolved oxygen (DO) and electrical conductivity (EC) emerged as the most sensitive and reliable indicators of effluent impact. A strong inverse relationship between DO and EC defined the primary pollution gradient, reflecting oxygen depletion and increased ionic strength associated with treated effluent inputs. In contrast, biochemical oxygen demand (BOD_5) exhibited limited discriminatory power, suggesting effective removal of biodegradable organic matter during treatment, but also highlighting the limitations of relying on BOD_5 as a primary regulatory indicator in wastewater-dominated rivers. Turbidity represented a secondary, largely independent process linked to sediment dynamics and episodic inputs rather than continuous effluent discharge.

Overall, the study concludes that conventional univariate monitoring approaches are insufficient for capturing the interacting, episodic and spatially complex nature of urban river pollution. Multivariate-derived composite indicators, such as the Effluent Impact Index developed in this research, offer a robust and interpretable means of translating complex datasets into actionable early-warning tools for river management.

RECOMMENDATIONS

It is recommended that regulatory monitoring frameworks place greater emphasis on multivariate analysis and indicator combinations rather than single-parameter thresholds. Routine monitoring should prioritise high-sensitivity parameters such as DO and EC, alongside improved temporal resolution to better capture storm-driven combined sewer overflow events. Furthermore, closer integration of primary and secondary datasets, supported by harmonised analytical protocols, would enhance confidence in compliance assessment and decision-making. Adoption of these approaches would support more adaptive, evidence-based management of the River Irwell and similar urban river systems impacted by wastewater discharges.

➢ *Limitations of the Study*

Although the study had its contributions, it had some limitations. The analysis methodologically used a small amount of time data that might not be able to capture seasonal changes in wastewater properties. Also, the study was limited to one treatment facility and its immediate receiving waters restricting generalization of the study results. The results could also have been affected by external sources like rainfall variability, industrial discharge variations and operational alterations in the treatment plant as observed in other research studies on wastewater impact (Phungela, 2020; Iloms et al., 2020). The spatial extent was also limited to a short river reach, which might not be a complete picture of downstream propagation of effluent effects across different regimes of flow. Also, the use of grab sampling limited the temporal resolution which could have underestimated the short-term variability of the storm events or operational changes of the treatment works.

➢ *Future Research Recommendations.*

The future studies are advised to increase sampling between seasons and hydrological states to enhance knowledge on time variability and threshold behavior of composite indicators. Greater spatial coverage would allow evaluation of downstream effects and distances of recovery in the far fields. The capability of early warning and the presence of episodic pollution events would be reinforced with the incorporation of with high-frequency automated sensors to monitor DO and EC. Biological indicators might be also incorporated into the framework of physicochemical parameters in further studies to increase ecological usefulness and synchronize the monitoring framework with the holistic approach to river health assessment, as proposed in the current water quality studies (Rufino et al., 2022; Zhang et al., 2023).

REFERENCES

[1]. Anderson, M.J. (2001) *A new method for non-parametric multivariate analysis of variance*. Austral Ecology, 26(1), pp. 32–46. https://doi.org/10.1111/j.14429993.2001.01070.pp.x (Accessed: 12 December 2025).

[2]. Begum, S., Firdous, S., Naeem, Z., Chaudhry, G.S., Arshad, S., Abid, F., Zahra, S., Khan, S., Adnan, M., Sung, Y.Y. and Tengku Muhammad, T.S. (2023) ‘Combined multivariate statistical techniques and Water Quality Index (WQI) to evaluate spatial variation in water quality’, *Tropical Life Sciences Research*, 34(3), pp. 129–149. https://doi.org/10.21315/tlsr2023.34.3.7 (Accessed: 13 November 2025).

[3]. de la Torre, M.L., Barata, A. and Pico, Y. (2020) Multivariate statistical analyses for water and sediment quality index development: a study of susceptibility in an urban river’, *Ecotoxicology and Environmental Safety*, 200, 110742. https://doi.org/10.1016/j.ecoenv.2020.110742 (Accessed: 27 October 2025).

[4]. Ecosystems Knowledge Network (2025) *Irwell catchment water quality issues*. Available at: https://ecosystemsknowledge.net/ (Accessed: 20 December 2025).

[5]. Environment Act 2021 (UK). Available at: https://www.legislation.gov.uk/ukpga/2021/30/contents (Accessed: 1 January 2026).

[6]. Environment Agency (2023) *Water quality monitoring methods and tools*. Environment Agency, Government of the UK. Available at: https://assets.publishing.service.gov.uk/media/5be99782e5274a084382c559/Book_204.pdf (Accessed: 30 December 2025).

[7]. FloodMapper (2024) *Bolton Wastewater Treatment Works spill record*. Available at: https://www.floodmapper.co.uk (Accessed: 15 December 2025).

[8]. Iloms, E., Ololade, O.O., Ogola, H.J.O. and Selvarajan, R. (2020) ‘Investigating industrial effluent impact on municipal wastewater treatment plant in Vaal, South Africa’, *International Journal of Environmental Research and Public Health*, 17(3), p. 1096. https://doi.org/10.3390/ijerph17031096 (Accessed: 2 January 2026).

[9]. Ioele, G., De Luca, M., Grande, F., Durante, G., Trozzo, R., Crupi, C. and Ragno, G. (2020) ‘Assessment of surface water quality using multivariate analysis: Case study of the Cratis River, Italy’, *Water*, 12(8), p. 2214.

https://link.springer.com/article/10.1007/s11356-020-10219-y (Accessed: 22 December 2025).

[10]. Ma, X., Wang, L., Yang, H., Li, N. and Gong, C. (2020) ‘Spatiotemporal analysis of water quality using multivariate statistical techniques and the Water Quality Identification Index in the Qinhuai River Basin’, *Water*, 12(10), p. 2764. https://doi.org/10.3390/w12102764 (Accessed: 3 January 2026).

[11]. Phungela, T.T. (2020) *Impact of wastewater treatment effluent on the water quality of the Crocodile River, Ehlanzeni District, Mpumalanga*. MTech Dissertation, Cape Peninsula University of Technology. Available at: https://etd.cput.ac.za/handle/20.500.11838/3458 (Accessed: 3 January 2026).

[12]. Rodríguez, R. and Pérez, J. (2021) ‘Assessment of physicochemical water quality in urban river systems’, *Journal of Environmental Management*, 280, 111783 (Accessed: 23 December 2025).

[13]. Rosenqvist, J., Nilsson, C., Hering, D. *et al.* (2021) Storm overflow impacts on riverine ecosystems’, *Science of the Total Environment*, 778, 146173. https://doi.org/10.1016/j.scitotenv.2021.146173 (Accessed: 3 January 2026).

[14]. Rufino, M., Gallego-Schmid, A., Stamford, L., & Azapagic, A. (2022). Environmental sustainability assessment of wastewater treatment technologies: A life cycle perspective. *Journal of Environmental Management*, 303, 114162. https://doi.org/10.1016/j.jenvman.2021.114162 (Accessed: 23 December 2025).

[15]. United Utilities (2022) *Storms Overflow Report 2022*. United Utilities. Available at: https://www.unitedutilities.com/globalassets/documents/pdf/united-utilities-storm-overflow.pdf (Accessed: 22 December 2025).

[16]. United Utilities (2024a) *United Utilities to upgrade Bolton wastewater treatment works*. United Utilities Newsroom, 8 February. Available at: https://www.unitedutilities.com/corporate/newsroom/latest-news/united-utilities-to-upgrade-bolton-wastewater-treatment-works/ (Accessed: 22 December 2025).

[17]. United Utilities (2024b) *Major milestone reached as part of improvements to Bolton wastewater treatment works*. United Utilities Newsroom, 30 September. Available at: https://www.unitedutilities.com/corporate/newsroom/latest-news/major-milestone-reached-as-part-of-improvements-to-bolton-wastewater-treatment-works/ (Accessed: 22 December 2025).

[18]. Zhang, Y., Yu, H., Liu, M. *et al.* (2023) ‘Analysis of water quality and habitat suitability for benthic macro-invertebrates in the Majiagou urban river, China’, *Water*, 15(12), 2269. https://doi.org/10.3390/w15122269 (Accessed: 4 January 2026).

[19]. Zhao, L., Wang, S. and He, Q. (2022) ‘Rainfall-driven variability in urban river water quality’, *Water Research*. https://doi.org/10.1016/j.watres.2022.118961 (Accessed: 5 September 2025).

[20]. Datasets: https://drive.google.com/drive/folders/10F9PVaxYBg2PyFcQ5oN85do5yq7Zwgm8?usp=sharing