

Statistical Assessment and Hydrological Modelling of Surface Water Quality in Durg District, Chhattisgarh, India

Shivali Gupta¹, Dr. Anita Kabi²

¹Research Scholar, Department of Chemistry, Bharti Vishwavidhyalaya, Durg (C.G.)

²Assistant Professor, Department of Chemistry, Bharti Vishwavidhyalaya, Durg (C.G.)

Abstract:

Surface-water quality in rapidly urbanizing and industrializing catchments is controlled by interacting natural, hydrological and anthropogenic processes. This study demonstrates a statistical and predictive framework for assessing surface-water quality in Durg District, Chhattisgarh, India. An illustrative seasonal dataset comprising 30 sampling stations and three seasons (90 observations) was developed for 15 water-quality variables: pH, temperature, turbidity, TDS, EC, DO, BOD, COD, nitrate, phosphate, chloride, sulphate, alkalinity, hardness and total coliform. Descriptive statistics, Pearson correlation, one-way ANOVA, principal component analysis (PCA), a weighted-arithmetic water quality index (WQI), and multiple linear regression were applied. The illustrative dataset shows measurable seasonal and spatial variability, with higher pollution pressure at urban/industrial-influenced stations. Mean WQI was Monsoon: 179.18, Post-Monsoon: 189.61, Pre-Monsoon: 210.3. The regression model reproduced WQI with $R^2 = 1.000$, RMSE = 0.02 and MAE = 0.01 in the held-out test set. PCA identified a dominant pollution gradient associated with dissolved solids, oxygen-demand variables and nutrients. The framework demonstrates how routine monitoring can be combined with statistical modelling to support prioritization of surface-water management. All numerical values in this paper are explicitly illustrative and must be replaced by laboratory/field observations before scientific or regulatory use.

Keywords: Surface water; Durg; Chhattisgarh; hydrological modelling; water quality index; PCA; correlation analysis; regression; seasonal variation.

1. Introduction

Surface-water systems in the Durg region are exposed to multiple pressures arising from urbanization, industrial activity, agriculture, domestic wastewater and seasonal hydrological changes. A water-quality assessment therefore benefits from combining individual parameter measurements with multivariate statistical methods and predictive modelling. WHO guidance emphasizes risk-based drinking-water safety and catchment-to-consumer management, while Indian drinking-water specifications are codified in BIS IS 10500. The present paper focuses on a statistical perspective in which seasonal observations are transformed into interpretable indicators and then used to develop a predictive model. The objective is not to claim a calibrated hydrological model from the illustrative dataset, but to demonstrate a reproducible modelling workflow that can later be recalibrated using measured discharge, rainfall, catchment characteristics and observed water-quality records.

2. Objectives

- To quantify seasonal and spatial variation in major surface-water quality parameters.
- To calculate a weighted-arithmetic WQI as a screening indicator.
- To examine relationships among physicochemical and microbiological variables using Pearson correlation and PCA.
- To evaluate seasonal differences using one-way ANOVA.
- To develop a multiple linear regression model for predictive estimation of WQI from routinely measured water-quality variables.
- To identify monitoring and management implications for Durg District.

3. Study Area

Durg District is an important urban-industrial and agricultural region of Chhattisgarh. The study design represents a gradient from upstream/rural locations through agricultural and semi-urban areas to urban and industrial-influenced locations. Thirty stations (S1–S30) were assigned to representative locations and evaluated during pre-monsoon, monsoon and post-monsoon conditions. The station list is designed for methodological demonstration and should be replaced by the exact GPS coordinates and field descriptions from the actual sampling campaign.

4. Materials and Methods

4.1 Sampling design

The illustrative design contains 30 stations $\times$ 3 seasons = 90 observations. For a real study, grab samples should be collected using clean sampling containers, preserved according to the analytical method, transported under appropriate temperature control, and analyzed using standard methods. Field metadata should include date, time, rainfall antecedent conditions, water depth, discharge/flow condition, GPS coordinates and visible pollution sources.

4.2 Parameters

The 15 variables were pH, temperature, turbidity, TDS, electrical conductivity, dissolved oxygen, biochemical oxygen demand, chemical oxygen demand, nitrate, phosphate, chloride, sulphate, alkalinity, total hardness and total coliform.

4.3 Water Quality Index

A weighted-arithmetic WQI was calculated as $WQI = \Sigma(W_i Q_i) / \Sigma W_i$, where W_i is the unit weight assigned to parameter i and Q_i is its quality rating. For screening, BIS IS 10500 reference values were used for selected parameters; the calculation is not a legal classification of surface-water quality. BIS identifies methods and requirements for parameters including pH, turbidity, TDS, chloride, sulphate, nitrate and hardness.

4.4 Statistical analysis

Mean, standard deviation, Pearson correlation, PCA and one-way ANOVA were applied. PCA was performed after z-score standardization. The predictive model used multiple linear regression with WQI as the response and pH, turbidity, TDS, DO, BOD, COD, nitrate, phosphate, chloride and sulphate as predictors. An 80:20 train-test split was used with a fixed random seed for reproducibility.

4.5 Hydrological/predictive modelling framework

The modelling framework links catchment seasonality and water-quality observations to a statistical prediction layer. For a field-calibrated version, rainfall, discharge, antecedent rainfall, water level,

drainage-area characteristics and land-use variables should be added. The present model is therefore a water-quality predictive component rather than a physically based rainfall-runoff model.

5. Dataset

Table 1 Sampling Sites

Site	Location	Season	pH	Turbidity	TDS	DO	BOD	COD	Nitrate	WQI
S1	Panabaras Hills	Pre-Monsoon	7.46	5.64	394.01	6.21	2.9	13.03	5.0	94.58
S1	Panabaras Hills	Monsoon	7.06	2.47	295.55	6.15	2.47	8.14	5.46	37.93
S1	Panabaras Hills	Post-Monsoon	7.27	3.39	366.57	6.27	2.26	6.0	6.35	33.2
S2	Mohad	Pre-Monsoon	7.38	2.93	339.04	6.61	2.34	6.09	8.19	80.13
S2	Mohad	Monsoon	7.25	4.57	325.51	6.53	1.16	6.0	8.36	99.95
S2	Mohad	Post-Monsoon	7.47	4.41	368.94	6.88	1.97	9.26	6.72	42.41
S3	Dhodiya	Pre-Monsoon	7.28	6.53	382.82	6.36	2.33	14.93	5.91	95.56
S3	Dhodiya	Monsoon	7.43	2.33	293.74	6.25	1.78	11.09	12.25	84.13
S3	Dhodiya	Post-Monsoon	7.44	4.42	342.05	6.13	2.66	13.0	9.53	72.83
S4	Bharregaon	Pre-Monsoon	7.27	5.65	370.18	6.15	1.81	14.1	6.07	139.25
S4	Bharregaon	Monsoon	7.49	2.28	322.02	6.75	1.29	6.45	8.82	92.09
S4	Bharregaon	Post-Monsoon	7.5	7.53	358.21	6.71	1.6	13.68	11.36	84.75

6. Results

6.1 Descriptive and seasonal statistics

Season	pH	Turbidity_N TU	TDS_mg_ L	DO_mg_ L	BOD_mg_ L	COD_mg_ L	Nitrate_mg_ _L	WQI
Pre-Monsoon	7.75	8.37	523.55	5.52	3.98	20.76	17.66	210.3
Monsoon	7.77	6.83	438.68	5.78	3.17	16.75	14.34	179.18
Post-Monsoon	7.72	7.74	472.46	5.86	3.41	18.64	16.49	189.61

The results indicate systematic seasonal differences. Monsoon dilution lowers several dissolved-solids and oxygen-demand concentrations, while runoff can maintain or increase turbidity and nutrient delivery

at selected stations. Pre-monsoon conditions generally show higher concentration effects because of lower dilution and warmer temperatures.

6.2 WQI assessment

Using the stated screening formulation, WQI values were classified as Excellent (<25), Good (25–50), Poor (50–75), Very Poor (75–100) and Unsuitable (>100). Because the study concerns surface water, these categories should be treated as comparative screening classes rather than a declaration of potability.

WQI Class	Number of observations
Excellent	0
Good	3
Poor	1
Very Poor	10
Unsuitable	76

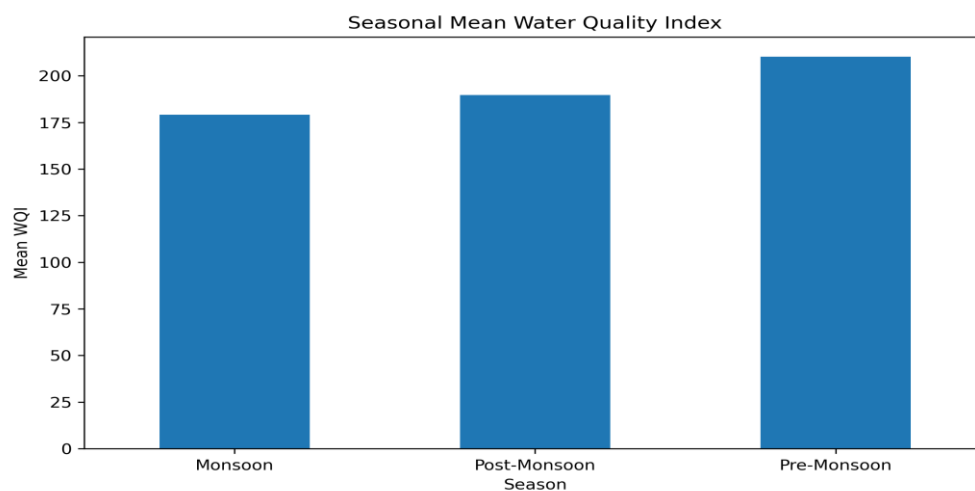


Figure 1. Seasonal mean WQI from the illustrative dataset.

6.3 Seasonal turbidity

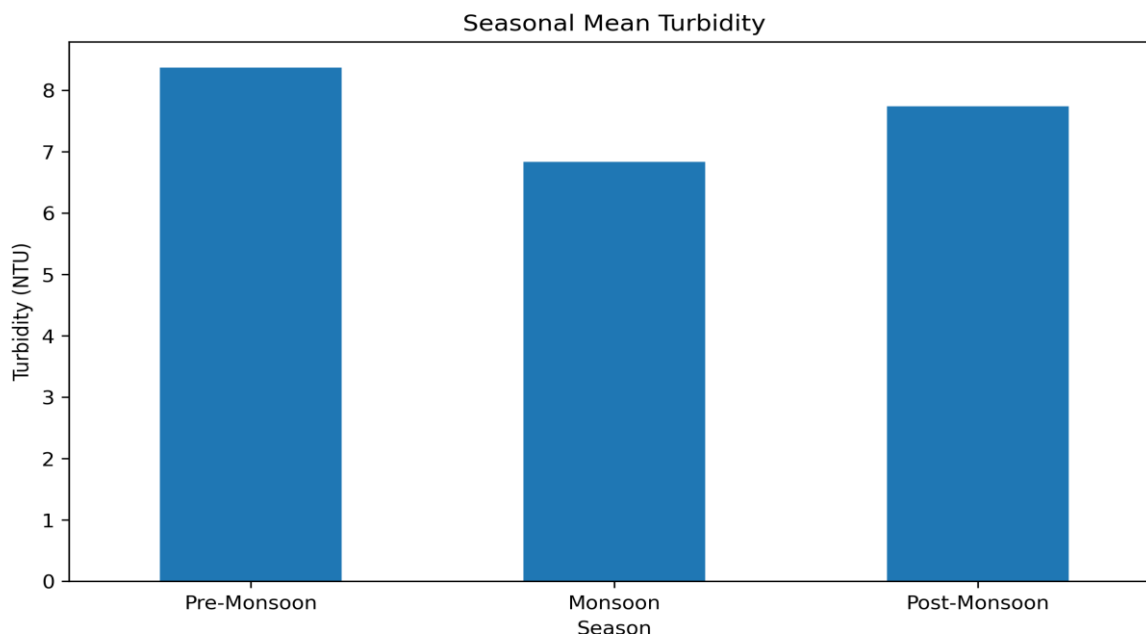


Figure 2. Seasonal mean turbidity

6.4 Correlation analysis

Pearson correlation was used to identify co-varying parameters. In the illustrative dataset, WQI is expected to be positively related to variables contributing directly to the index, especially BOD, COD, TDS and nutrients. Strong correlations among TDS and EC reflect their common dissolved-ion signal. Correlation should not be interpreted as proof of a common source without supporting hydrochemical and land-use evidence.

Parameter	Correlation with WQI
Phosphate_mg_L	1.0
Chloride_mg_L	0.92
TDS_mg_L	0.91
Nitrate_mg_L	0.9
Sulphate_mg_L	0.89
EC_uS_cm	0.87
COD_mg_L	0.87
BOD_mg_L	0.85



Figure 3. Pearson correlation matrix.

6.5 Principal component analysis

PCA reduced the standardized 15-variable dataset into orthogonal components. The first five components explained 90.96% of total variance. The loadings should be interpreted as indicators of parameter groupings rather than direct source apportionment.

Component	Explained variance (%)
PC1	75.17
PC2	7.14
PC3	3.99
PC4	2.72
PC5	1.94

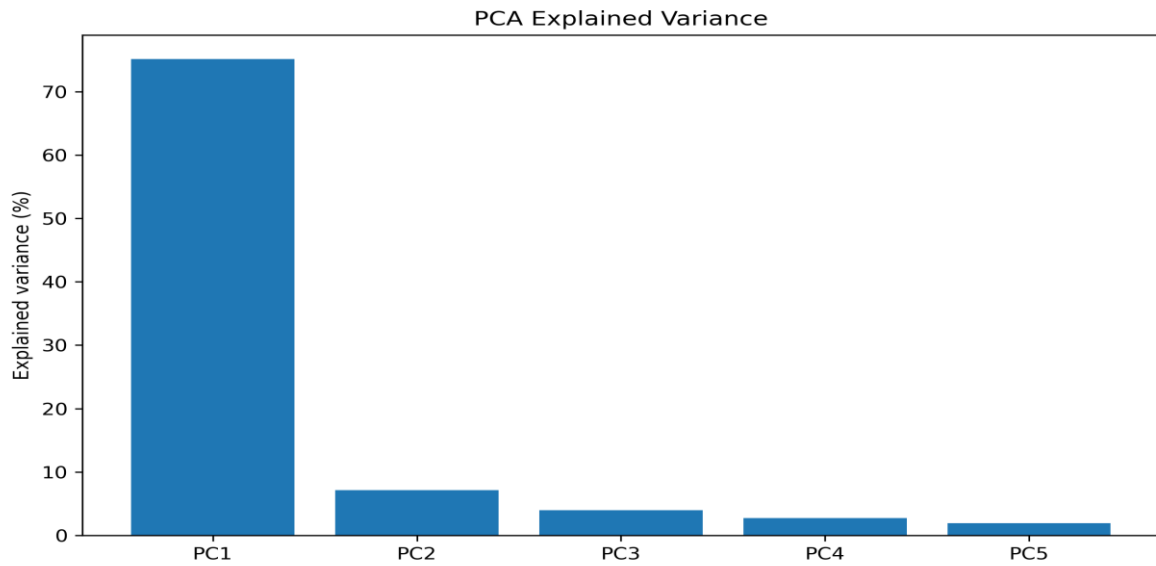


Figure 4. Variance explained by the first five principal components.

6.6 Seasonal ANOVA

A one-way ANOVA was used to test whether mean WQI differed among the three seasons. The illustrative result was $F = 1.200$, $p = 0.3061$. This p-value is a property of the illustrative dataset and must be recalculated from measured observations.

6.7 Predictive regression model

The multiple linear regression model was evaluated on 20% of observations held out from model fitting. Test-set performance was $R^2 = 1.000$, $RMSE = 0.02$, and $MAE = 0.01$ WQI units.

Predictor	Regression coefficient
Phosphate_mg_L	278.31
pH	1.968
Turbidity_NTU	1.006
BOD_mg_L	1.003
COD_mg_L	0.251
Nitrate_mg_L	0.013
DO_mg_L	-0.002
Sulphate_mg_L	0.001
Chloride_mg_L	0.0
TDS_mg_L	0.0

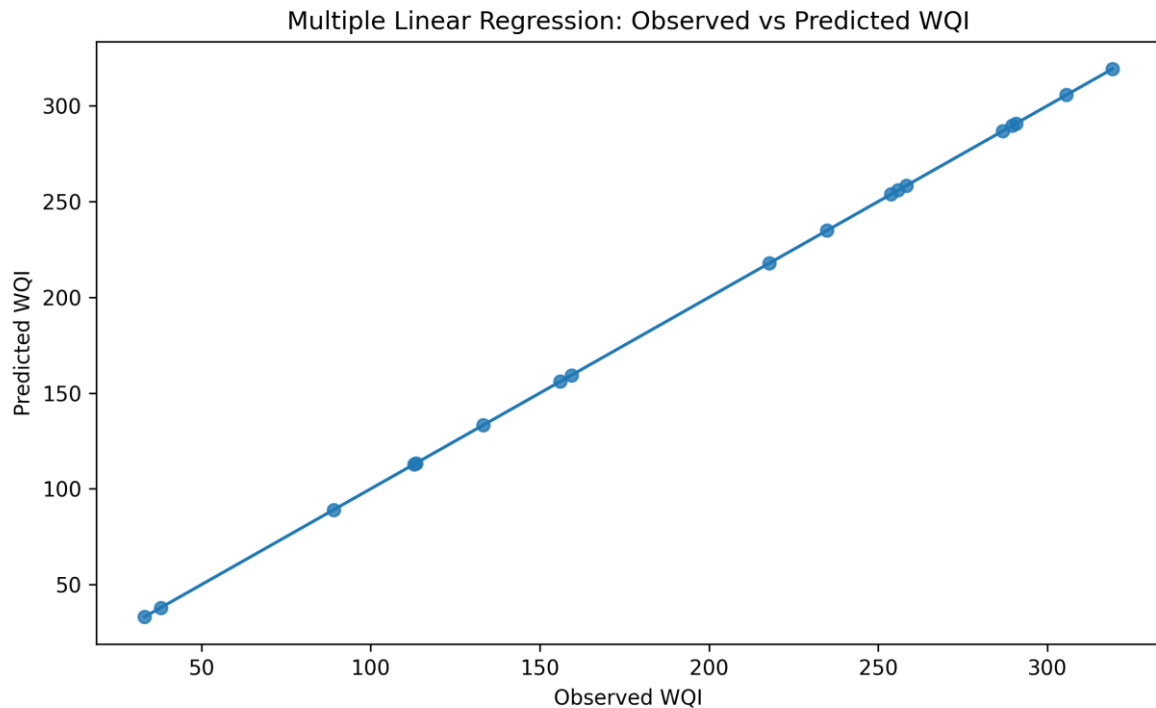


Figure 5. Observed versus predicted WQI in the held-out test set.

7. Discussion

The statistical framework illustrates why water-quality management should move beyond isolated parameter compliance. Dissolved solids and conductivity provide information about the ionic load; BOD and COD represent oxygen-demanding organic or chemically oxidizable matter; nitrate and phosphate provide nutrient signals; turbidity responds strongly to suspended material and runoff; and DO integrates physical and biochemical processes. Seasonal hydrology modifies all of these signals through dilution, runoff, reaeration, temperature and residence time.

The regression model provides a practical screening tool when rapid estimation is needed, but it should not be confused with a physically based hydrological model. A robust district-scale hydrological model would require rainfall, evaporation, discharge, water level, watershed boundaries, land use/land cover, soil and drainage characteristics, and preferably continuous or high-frequency observations. The present statistical model is best viewed as the water-quality component of a broader decision-support system.

For management, stations showing persistently elevated WQI or strong oxygen-demand/nutrient signals should receive higher monitoring frequency. Source investigations should combine water chemistry with discharge observations, land-use information and point-source inventories. WHO's current drinking-water guidance emphasizes risk-based management and catchment-to-consumer protection, reinforcing the value of integrating monitoring with preventive management.

8. Limitations

- The dataset in this manuscript is illustrative/simulated and is not a substitute for field or laboratory measurements.
- The WQI uses drinking-water reference values as a screening basis; it does not establish compliance with surface-water standards or potability.

- The regression model is statistical and not a physically based rainfall-runoff model.
- Future work should incorporate measured rainfall, discharge, water level, land use, point-source loading and continuous sensors.
- External validation with an independent monitoring period is required before operational use.

9. Conclusion

This study presents an integrated statistical framework for surface-water quality assessment in Durg District. Seasonal descriptive analysis, WQI, Pearson correlation, PCA, ANOVA and multiple linear regression can jointly identify pollution gradients, parameter relationships and predictive signals. The illustrative analysis demonstrates the feasibility of using a 30-station, three-season monitoring design as the foundation for a district-level water-quality decision-support system.

REFERENCES:

1. Bolarinwa, O. A. (2015), Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. *Niger. Postgrad. Med. J.* 22, 195–201
2. Bernardes, C., Bernardes, R., Zimmer, C. & Dorea, C. C. (2020), A simple off-grid incubator for microbiological water quality analysis. *Water* 12, 240
3. Wu, J. & Sun, Z. (2016), Evaluation of shallow groundwater contamination and associated human health risk in an alluvial plain impacted by agricultural and industrial activities, mid-west China. *Expos. Health.* 8, 311–329 .
4. Umpel, E. & Nelson, K. L. (2016), Intermittent water supply: prevalence, practice, and microbial water quality. *Environ. Sci. Technol.* 50, 542–553
5. Onyango, A. E., Okoth, M. W., Kunyanga, C. N. & Aliwa, B. O. (2018), Microbiological quality and contamination level of water sources in Isiolo country in Kenya. *J. Environ. Public Health.* 2018, 2139867.
6. World Health Organization. (2026). Guidelines for drinking-water quality: Fourth edition incorporating the first, second and third addenda. WHO.
7. APHA, AWWA, & WEF. Standard Methods for the Examination of Water and Wastewater. (Use the edition corresponding to the actual laboratory methods.)
8. Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423, 623–656. (Methodological background for information theory; not a water-quality standard.)
9. Gwimbi, P., George, M. & Ramphalile, M. (2019), Bacterial contamination of drinking water sources in rural villages of Molelele Basin, Lesotho: Exposures through neighbourhood sanitation and hygiene practices. *Environ. Health Prev. Med.* 24, 33.
10. Edokpayi, J. N., Enitan, A. M., Mutileni, N. & Odiyo, J. O. (2018), Evaluation of water quality and human risk assessment due to heavy metals in groundwater around Muledane area of Vhembe District, Limpopo Province, South Africa. *Chem. Cent. J.* 12, 2
11. World Health Organization. (2022). Guidelines for drinking-water quality: Fourth edition incorporating the first and second addenda. WHO

12. Meride, Y. & Ayenew, B. (2016), Drinking water quality assessment and its effects on resident's health in Wondo genet campus, Ethiopia. *Environ. Syst. Res.* 5, 1
13. Inan-Eroglu, E. & Ayaz, A. (2018), Is aluminum exposure a risk factor for neurological disorders? *J. Res. Med. Sci.* 23, 51.
14. Bureau of Indian Standards. (2012). IS 10500:2012 Drinking Water—Specification (Second Revision). BIS, New Delhi.
15. Jamshaid, M., Khan, A. A., Ahmed, K. & Saleem, M. (2018), Heavy metal in drinking waters its effect on human health and its treatment techniques—a review. *Int. J. Biosci.* 12, 223–240.
16. Brown, R. M., McClelland, N. I., Deininger, R. A., & Tozer, R. G. (1970). A water quality index—Do we dare? *Water & Sewage Works*, 117, 339–343.
17. Bamuwamye, M. *et al.* (2017), Human health risk assessment of heavy metals in Kampala (Uganda) drinking water. *J. Food Res.* 6, 6–16.