



Article Big Data Analysis Framework for Water Quality Indicators with Assimilation of IoT and ML

Sanjeev Kimothi ¹⁽¹⁾, Asha Thapliyal ², Shaik Vaseem Akram ¹⁽¹⁾, Rajesh Singh ¹⁽¹⁾, Anita Gehlot ¹, Heba G. Mohamed ³⁽¹⁾, Divya Anand ^{4,5,*(1)}, Muhammad Ibrahim ⁶⁽¹⁾ and Irene Delgado Noya ^{5,7}

- ¹ Division of Research & Innovation, Uttaranchal University, Dehradun 248007, India; sanjeevkimothi@uttaranchaluniversity.ac.in (S.K.); shaikhvaseemakram@uttaranchaluniversity.ac.in (S.V.A.); rajeshsingh@uttaranchaluniversity.ac.in (R.S.); anita.ri@uttaranchaluniversity.ac.in (A.G.)
- ² Uttarakhand Space Application Centre (USAC), Dehradun 248001, India; ashath08@gmail.com
- ³ Department of Electrical Engineering, College of Engineering, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia; hegmohamed@pnu.edu.sa
- ⁴ School of Computer Science and Engineering, Lovely Professional University, Punjab 144411, India
- ⁵ Higher Polytechnic School, Universidad Europea Del Atlántico, C/Isabel Torres 21, 39011 Santander, Spain; irene.delgado@uneatlantico.es
- ⁶ Big Data Research Center, Jeju National University, Jeju 63243, Korea; ibrahimmayar@jejunu.ac.kr
- ⁷ Department of Project Management, Universidad Internacional Iberoamericana, Campeche 24560, Mexico
- * Correspondence: divyaanand.y@gmail.com

Abstract: According to the United Nations, the Sustainable Development Goal '6' seeks to ensure the availability and sustainable management of water for all. Digital technologies, such as big data, Internet of Things (IoT), and machine learning (ML) have a significant role and capability to meet the goal. Water quality analysis in any region is critical to identify and understand the standard of water quality and the quality of water is analyzed based on water quality parameters (WQP). Currently, water pollution and the scarcity of water are two major concerns in the region of Uttarakhand, and the analysis of water before it is supplied for human consumption has gained attention. In this study, a big data analytics framework is proposed to analyze the water quality parameters of 13 districts of Uttarakhand and find the correlation among the parameters with the assimilation of IoT and ML. During the analysis, statistical and fractal methods are implemented to understand the anomalies between the water quality parameters in 13 districts of Uttarakhand. The variation in WQP is analyzed using a random forest (RF) model, and the dataset is segmented location wise and the mean, mode, standard deviation, median, kurtosis, and skewness of time series datasets are examined. The mean of the parameters is adjusted with the coefficient of variation based on the standard values of each parameter. The turbidity in almost all the experimental sites has a normal distribution, with the lowest mean value (0.352 mg/L) and highest (11.9 mg/L) in the Pauri Garhwal and Almora districts, respectively. The pH of the water samples is observed to be in the standard range in all the experimental sites, with average and median values being nearly identical, at 7.189 and 7.20, respectively. However, the pH mode is 0.25. The Cl^{-} concentration varies with mean values from the lowest (0.46 mg/L) to the highest (35.2 mg/L) over the experimental sites, i.e., the Bageshwar and Rudraprayag districts, respectively. Based on the analysis, it was concluded that the water samples were found to be safe to drink and in healthy condition in almost all the districts of the state Uttarakhand, except for the Haridwar district, where some increase in contaminants was observed.

Keywords: fractals; water quality indicator; predictability index; remote sensing and GIS; machine learning (ML); Internet of Things (IoT)

1. Introduction

According to the Sustainable Development Goal '6', there is a need to ensure the availability and sustainable management of water and sanitation for all [1]. Concerning



Citation: Kimothi, S.; Thapliyal, A.; Akram, S.V.; Singh, R.; Gehlot, A.; Mohamed, H.G.; Anand, D.; Ibrahim, M.; Noya, I.D. Big Data Analysis Framework for Water Quality Indicators with Assimilation of IoT and ML. *Electronics* **2022**, *11*, 1927. https://doi.org/10.3390/ electronics11131927

Academic Editor: Prasan Kumar Sahoo

Received: 30 May 2022 Accepted: 15 June 2022 Published: 21 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). this, the Indian government is striving towards achieving the goal of the United Nations. The Himalayan rivers are the major source of surface water in India and they are one of the world's most abundant sources of drinking water and become easily polluted due to anthropogenic emissions [2,3] and manmade activity, such as urbanization, industrial growth, wastewater discharge, and agricultural expansion [4]. Water as a natural resource system requires a procedure with a large number of components that may perform computations to study the water quality indices (WQIs) in dynamic phenomena for lakes, rivers, watersheds, etc. [5]. Inconsistency among the input parameters is usually observed due to geographic, geological, and demographic challenges [6,7]. In comparison to surface water, the subsurface is the largest store of drinkable water and it is less polluted, but in the current scenario, the increased inorganic trashes, on the other hand, can contaminate both surface and groundwater.

A recent study involving the hydrogeochemical analysis of the glacier meltwater was conducted over the higher mountainous region of the Chaturangi glacier, Garhwal Himalaya [8] and concluded that the water quality has degraded in the ablation period. It could be related to the rate at which chemical reactions occur, causing rock breakdown and mineral dispersion into the melting water. The water quality index (WQI) is employed to analyze the water quality of the river Yamuna during 2000–2009 and it was observed that the critical parameters are affecting the water quality [9]. The state of Uttarakhand is the part of the central and western Himalayas and the rivers emerging from the peaks of the Himalayas provide freshwater [10] and the low-lying regions in the state of Uttarakhand have been experiencing the cumulative impact, including soil pollution and surface water contamination due to the rise in industrialization, and large construction activities over the last several decades [11]. In developing countries, the majority of the health issues are due to the contamination in the drinking water, which is critically required to be taken into consideration for a sustainable ecosystem and indeed the good quality of the drinking water [12].

The assessment of groundwater vulnerability maps demands the adoption of a variety of approaches and procedures based on hydrogeological knowledge of the research area and the use of predictive models. For estimating ground water quality, there are two popular modelling strategies, (1) deterministic physically models and (2) statistical and ML methods [13]. Several climate and watershed parameters are used in this study to determine the long-term median stream water quality indicators using ML algorithms [14]. Nonlinear interactions linked with ground water can now be handled by recent ML algorithms. Furthermore, rather than relying on a previous assumption, these algorithms discover the link between response variables and predictors, enhancing model prediction accuracy. It is proved that RF-based models are appropriated for the analysis of water quality because they are flexible in handling non-linear relationships, limit model overfitting, have less user defined parameters and are capable of incorporating the quantitative and qualitative variables [15].

With the motivation of these aspects, this study aims to implement a big data analytics framework to analyze the water quality parameters of 13 districts of Uttarakhand. Along with this, an architecture is proposed to analyze the big data that is obtained through the IoT and field. This variation in data is analyzed with an RF model, in which the mean, mode, standard deviation, median, kurtosis, and skewness of time series datasets are examined. The contribution of the study is as follows:

- A big data analytics-based framework is proposed in this study for water quality parameter analysis.
- The variation in WQP is analyzed using a random forest (RF) model, and the mean of the parameters is adjusted with the coefficient of variation based on the standard values of each parameter.
- The predictability analysis of 13 experimental sites is carried out with the fractal approach.
- The results were obtained from the statistical analysis over the different experimental sites using two approaches i.e., skewness, kurtosis curves and correlation matrix tables for each parameter over the studied experimental sites.

This research paper is organized as follows. The WQI applications in the context of the extended fractal analysis for water quality parameters (WQPs) are reviewed in Section 2. Section 3 covers the materials and methods. Section 4 covers the results and discussion. The last and final Section 5 concludes the paper.

2. Background

WQIs provide the environment of a water body in terms of pollution load segmentation and identification of water quality. Dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), pH, NH3-NL, and SS are the parameters for these indices and there is a high need for spatial and temporal monitoring [16]. The comprehensive analysis with the simulation of the water quality datasets is highly required to a large extent, as it is challenging to assess spatial variation among the WQIs. The remote sensing technique and ground-level survey of the parameters will be useful in a spatial and temporal range study of the quality assessment for policymakers [17]. Each water quality parameter may be used to analyze the quality of drinking water and its quality management initiatives for stakeholders. The influence of water parameters and quality index assessment was investigated for 28-years of data in the upper uMngeni watershed for the 11 samples (pH, electrical conductivity, temperature, turbidity, total suspended particles, NH4-N, NO3-N, PO4-P, and total phosphorus [18]. There is an urgent need for a comprehensive analysis of WQIs, especially in the Himalayan region including statistical, fractal, and geospatial techniques [19]. The identification of interlinkages between environment and water pollutants can help to enhance the forecasting of chemical exposure and helps in the evaluation of pollutant export over complex ecosystems [20]. The links between some particular water pollutants and the selected hydrological character variables were elucidated using a modified least-square analysis, due to the substantial co-dependence of the hydrological processes [21].

The WQIs were used to categorize groundwater from the permanent wells in the Kanchipuram district of Tamil Nadu, where 13 groundwater quality parameters were used to conclude water quality [22,23]. There are two traditional approaches to analyzing the WQI data and they are statistical and fractal approaches. The statistical and fractal approaches may provide the variation in the parameters spatially, as the Himalayan region has undulating geography, as well as huge network sources of surface water. However, information from only ground observations of a single indicator fails to resolve the complicated geographical terrain, which has an influence mostly on the hydrological cycle either directly or indirectly [24]. Fractal analysis was performed on the quality of rural domestic wastewater under the condition of dissolved oxygen stability [25]. Fractal analysis i.e., the Hurst exponent (H), fractal dimension (FD), and predictability (PI) of the water parameters will lead to better knowledge, as well as the ability to explain the better-approximated values [5] and statistically, it can lead to a reduction in the uncertainty in the analysis of large datasets. The water quality has been evaluated at multiple places for each water parameter using statistical methods. To determine the trend and predictability of water quality and regression, correlation coefficient, autoregressive integrated moving average (ARIMA), Box-Jenkins, residual autocorrelation function (ACF), residual partial autocorrelation function (PACF), lag, fractal, Hurst exponent, and predictability index were calculated [26].

Regularly inspecting the quality of the water might be good practice for ensuring that the water is safe to drink. Therefore, it is suggested [27] that a real-time monitoring system is highly required to assess and forecast future water quality indicators using ML and the IoT. The time series pattern in the data was extracted using a long short-term memory neural network (LSTM NN) and the wWQPs were obtained using sensors such as a pH sensor, turbidity sensor, and total dissolved solids (TDS) sensor, and the data was utilized to forecast future parameter values [28]. The sensors, Arduino, and NodeMCU in the IoT module can be embedded in the water supply to continuously monitor the parameters and this assists users to receive early alerts in case of increased containment WQPs [29]. The water quality prediction model necessitates the usage of high-quality data. Big data is produced at a high rate in the process of building and operating smart water quality monitoring systems based on the IoT, making water quality data difficult. A network model was examined [30] and found that it is capable of merging distributed observation data with geologically separated local models using smart sensor-based federated learning. Furthermore, an optimal scheduler is proposed to improve the overall network system's efficiency by making use of real-time massive data arrivals. A deep learning approach for forecasting water quality in IoT systems with the long-short term memory (LSTM) algorithm can be used for forecasting the water quality indicators [31].

In the Karoon River of Iran, the ANN approaches, such as multilayer perceptron (MLP), radial basis network, and adaptive neuro-fuzzy inference system (ANFIS) models are used [32] to compute DO, BOD, and chemical oxygen demand COD levels. The models included nine input water quality variables, including EC, pH, Ca, Mg, Na, turbidity, PO₄, NO₃, and NO₂, all of which were measured in river water. In addition, when RMSE, and MAE indices, and coefficients of correlations (r2) for predicting DO, BOD, and COD were compared, the MLP/BP model outperformed the ANN-RBF and ANFIS models [33]. It is suggested that the ANN can be utilized as an effective tool to compute and forecast river water quality metrics.

These approaches provide complete and efficient results, which are useful in the comprehension of surface water recharge modeling, especially in delicate ecology [33]. The understanding of water-sensitive parameters and major contaminated source identification is advantageous for drinking water model construction and water quality analysis modeling.

3. Materials and Methods

In this section, we present the water quality parameter that is used for the analysis. Along with this, we have presented the statistical and fractal approach for the ground water quality analysis over the 13 districts of Uttarakhand.

3.1. Big Data-Based Framework for Water Quality Analysis

In this study, we proposed a big data-based framework for carrying out the water quality analysis of Uttarakhand. Figure 1 illustrates the framework, which comprises three different components named as ground water quality monitoring, big data analytics platform, and water data. The first component describes the problems that are identified as relevant to the ground water quality, which consist of water quality degradation, land subsidence, interconnected ground water depletion, and ground water storage reduction. At present, this study focused on the problems based on water quality degradation, and land subsidence. In water quality degradation, the quality of water is analyzed based on the contamination level identified in the water. This study carried out the research by analyzing the WQP, such as turbidity, Cl^- , Fe^{2+} , As^{++} , NO_3^- , pH, Ca, Mg, FluoF-), total dissolved salts (TDS), alkalinity, hardness, and sulphate (SO_4^{2-}) of 13 districts to conclude the water quality.

Land subsidence is a broad term that indicates the lower vertical movement of the Earth's surface, produced by both natural and human forces. Regarding the data, the field data, remote sensing data, and simulated data are considered for the water quality analysis. In the big data analytics platform, the data of the 13 district WQPs are collected based on WQP, and the usable data are concluded from the data processing. Downscaling refers to the strategies used to regionalize information from global climate models and generate fine-scale climate change projections. Fine resolution information is required to better localize WQP. Through the process of downscaling, big data analytics solutions can address the mismatch between regional size data and local-scale information. Statistical modeling, data mining, and pattern discovery and categorization are the analytics methods used to analyze the collected data for making the decisions.



Figure 1. Big data-based framework for water quality analysis.

3.2. Assimilation of IoT and ML (References)

Figure 2 illustrates the architecture that presents the implementation of IoT and ML for the identification of the WQIs of every site based on the WQP data. The sensor node is obtained from the field data related to the water parameters of every district of the Uttarakhand. The field data are communicated to the cloud server LoRa and gateway. The field data available in the cloud server are pre-processed for the transforming of the raw data into an understandable format. Currently, there are different data-driven models, statistical multivariate methods, such as weights of evidence, logistic regression, and a set of AI or ML methods, such as genetic algorithms, adaptive neuro-fuzzy inference system, support vector machines (SVM), artificial neural networks (ANN), and, more recently, random forest (RF) that exist [34]. In hydrogeology investigations, multivariate statistical approaches and ANN are the most widely utilized [35]. Unfortunately, these techniques have a multitude of limitations, including their sensitivity to outlier values in logistic regression and the opacity of neural networks. Currently, ML has advanced significantly in recent years, and new methods have been developed to tackle some of the concerns stated for the frequently used methods [36].

The developing form of the ML technique that employs ensembles of regressions is gaining attention and this ensemble learning technique employs the same basic process to generate repeated numerous predictions, which are then averaged to form a unique model [37]. RF is one of ensemble learning, which is rapidly used for land-cover categorization from sensed data, as well as other domains connected to the environment and water resources [38]. An ensemble of regression (or classification) tree models used in the RF algorithm technique and a succession of separate trees is constructed based on random sub samples from the original data [39]. Each subsample has a decision tree, which is used to forecast the response variable (or a class). The integration of many trees increases the likelihood of developing an effective prediction model. The Algorithm 1's accuracy is mostly determined by the strength of the individual tree classifiers and their interdependence [40].

Here, an RF model is applied on the pre-processed data to identify the variability in WQIs of every site. The WQPs of 13 districts (Uttarakhand) are retrieved in the excel format. The dataset is classified as per the latitude and longitude of the respective district. Mean, mode, standard deviation, median, kurtosis, and skewness are used to examine the time series datasets. On the basis of the standard values of each parameter in every district (Site), the variation in WQP is evaluated using the random forest (RF) model [41] and the means

Sensor Sensor Jode '2 Node '1 LoRa Sensor data Sensor Sensor Gateway Node '3' Node 'n' with LoRa + Data Collection Layer Wi-Fi **Analytics Layer** Cloud Server Random Forest model Variability in www.hallalombh WQIs (H, D, PI) are the outcome Comb Rehalism of the every site

of the parameters are tuned with the coefficient of variation (CV). Now, the generated time series dataset is employed in the fractional Brownian motion (FBM).

Figure 2. Architecture of the assimilation of IoT and ML for WQIs.

Algorithm 1 Proposed random forest algorithm

- 2: For each element $x \in X$, there is a response y: $y \in Y$, Where $Y = y_1, y_2 \dots y_n$
- 3: Each set is bagging B times, where B is independent parameter
- 4: This training set selects a random sample with replacement of the training set and also fits
- decision trees to these sample.
- For $b = 1 \dots B$
- 4.1: Sample, with replacement, n training examples from X_1 , $Y_1 = X_b$, Y_b
- 4.2: Train the regression tree f_b on X_b , Y_b

3.3. Water Quality Parameters Used

The current study is carried out in the various stress conditions, which include the different elevations. The stress conditions refer to the geographical and geological variation in 13 districts of the state Uttarakhand. The datasets of the surface water quality parameters used in the study are obtained from the website of the Ministry of Water and Sanitation, the Govt. of India, for the year 2016. Daily monitoring of the WQPs at each site show anomalies, which is an indicator of the irregular pattern of water quality, and it is essential to analyze the linkages among them.

The fractal indices FD, H, and PI of the mentioned thirteen water quality indices (WQIs) are generated. The statistical analyses, such as mean, mode, standard deviation, median, kurtosis, skewness, and coefficient of variation (CV) of each parameter, along with



^{1:} Given a training set $X = x_1, x_2 \dots x_n$

each experimental site are computed and compared with the geospatial distribution. The box plots of all 13 water quality parameters (1-turbidity, 2-chloride (Cl⁻), 3-iron (Fe²⁺), 4-arsenic (As⁺⁺), 5-nitrate (NO₃⁻), 6-pH, 7-calcium (Ca), 8-magnesium (Mg), 9-fluoride (F⁻), 10-total dissolved salts (TDS), 11-alkalinity, 12-hardness, 13-sulphate (SO₄⁻)obtained from different experimental sites (13 districts) are shown in Figure 3. In India, water quality assessment and regulation will be taken care of by the Ministry of Drinking Water and Sanitization, Govt. of India, which has given a standard limit of WQPs. Table 1 illustrates the WQPs as per the Indian standard acceptance range (source: Methodology Manual for Groundwater quality mapping, Rajiv Gandhi national drinking water mission [42]).



Figure 3. Cont.



Figure 3. Cont.



Figure 3. Box plots of the ground water quality parameters in various sites (1–13).

| Table 1. | Observed | water quali | ty parameters | s with thei | r symbols | and site v | wise dist | ribution of | of the |
|----------|-------------|--------------|---------------|-------------|-----------|------------|-----------|-------------|--------|
| observe | d water qua | ality parame | ers [42]. | | | | | | |

| WQP | BIS Guideline Values | Observation Sites | District | Observed WQP in Each Site |
|----------------------------------|-----------------------------|--------------------------|------------------------|-------------------------------------------|
| Turbidity | 2.5 NTU | Site1 | Almora | 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 |
| Cl ⁻ | 200–1000 mg/L | Site2 | Bageshwar | 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 |
| Fe ²⁺ | 0.1 mg/L | Site3 | Chamoli | 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 |
| As ⁺⁺ | 0.05 mg/L | Site4 | Champavat | 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 |
| NO ₃ - | 45 mg/L | Site5 | Dehradun | 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 |
| pН | 6.5 to 8.5 | Site6 | Nainital | 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 |
| Са | 75 mg/L | Site7 | Pauri Garhwal | 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13 |
| Mg | 30 mg/L | Site8 | Pithoragarh | 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13 |
| F | 1.0–1.5 mg/L | Site9 | Tehri Garhwal | 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13 |
| TDS | 200 mg/L | Site10 | Uttarkashi | 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13 |
| Alkalinity | 200 mg/L | Site11 | Udam Singh Nagar (USN) | 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13 |
| Hardness | 200–600 mg/L | Site12 | Haridwar | 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13 |
| (SO ₄ ²⁻) | 200–400 mg/L | Site13 | Rudrapraygh | 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13 |

3.4. Statistics Analysis

The statistics, including indices and multivariate of WQP time series, are analyzed using the mode, mean and median to calculate the standard deviation from the series datasets, and spatially averaged frequency values. The sample dataset variability is assessed using standard deviation (Std), and kurtosis for data peakedness estimation. Skewness techniques are used to estimate the symmetry between the data points in a particular location. The range of variation in the data series within a sample time series is determined by the coefficient of variation. Regression analysis is carried out to understand the interrelationship in the water quality parameters (WQPs) through regression between the dependent variable (Y) and independent variable (X), which is represented by the following regression equation:

1

$$\mathcal{X} = mX + C \tag{1}$$

where *C* is the integration constant.

Coefficient of regression =
$$(m_{yx}) = r * (\sigma_y / \sigma_x)$$
 (2)

Coefficient of correlation =
$$\frac{\left[\Psi(XY) - \Psi(X)E(Y)\right]}{\left[\left(\Psi(X^2) - \Psi(X)^2\right)\left(\Psi(Y^2) - \Psi(Y)^2\right)\right]^{1/2}} = \left[\frac{\text{COV}(X,Y)}{\sigma_Y\sigma_X}\right] (3)$$

The predicted values for the variables are $\psi(X)$, $\psi(Y)$, $\psi(XY)$ and σ is the standard deviations of the variables.

3.5. Mathematical Analysis

WQPs are used in fractal prediction patterns, such as the chaotic, random, or deterministic structural form, to analyze the irregular pattern in the data time series. Identification, categorization, and mapping of intensive water characteristics necessitate a continual assessment of the natural resource data.

3.5.1. Fractal Dimension (FD)

Scientists and hydrologists are concerned about distinguishing between both clean and polluted water and developing relationships between their characteristics. Furthermore, local officials, especially those in developing countries, are finding it difficult to make drinkable water available with the increased levels of contaminants. The oscillations of environmental and also soil parameters, along with the dynamic interactions between subsurface hydrology, influence the statistical association of water systems.

3.5.2. Hurst Exponent (H)

The Hurst exponent (H) is calculated using typical wavelet approaches and evaluated by the regression equation. The coefficient of each water parameter indicates whether or not it has Brownian time series (or true random walk) behavior with other variables. Table 2 summarizes the statistical and fractal analyses and each WQI analysis.

The fractal and fractal dimension concept [43] was formulated by Benot Mandelbort in the year 1975 and defined the "fraction or fractured", which are governed by their selfsimilarity characteristic. This means that they exhibit similar features across a wide range of scales and that a single system has comparable qualities to the entire fractal. Owing to its efficiency and automated computational complexity, fractal dimension estimates from a fractal set have a variety of approaches. In several domains, box-counting is a popular technique for analyzing image properties such as texture segmentation, classification, and graphic analysis. Other prominent advantages of FD approaches include being able to distinguish deterministic and randomness in time series datasets in the form of variance and spectra distribution. The fractal is a mathematical approach used in fractal geometry to investigate naturally complex phenomena, such as the structures of clouds and geographic boundaries, as well as to differentiate between glacial and fluvial morphology. Fractal based on the degree of multifractality analysis for landscapes of glaciers and rivers has been examined and has found that more glaciers have a more complicated structure than rivers. The Hausdorff dimension is the most fundamental definition of FD; however, other common interpretations that are simple to calculate include box-counting and box dimension.

The Hurst exponent (*H*) is a factual variable analysis with exponential-scaling; it is an indicator of a time series long-term memory. The *H* and FD are also inextricably linked, indicating the roughness of a surface. The *H* can be persistent ($0.5 < H \le 1$) or antipersistent $0 \le H < 0.5$) in a time series, and when the data are not inter-correlated, H = 0.5 indicates that the series is unpredictable. Because it provides statistical self-similarity relationships, this method is applied in a variety of complicated engineering domains. The Hurst exponent of time series is defined as follows in terms of the exponential growth scaling relation.

| | (a): Skewness Variation in the Water Quality Indices in 13 Experimental Sites | | | | | | | | | | | | | |
|--------------------------------------------------------------------------------|-------------------------------------------------------------------------------|------------------|------------------|--------|--------|--------|---------|---------|-------------|--------|------------|----------|-----------|--|
| Parameters/ Sites | Cl- | Fe ²⁺ | As ⁺⁺ | Ni | pН | Ca | Mg | Fl | SO_4^{2-} | TDS | Alkalinity | Hardness | Turbidity | |
| 1 | 6.905 | 6.252 | 1.193 | 2.312 | -1.170 | 6.159 | 4.308 | 1.009 | 5.686 | 1.916 | 4.257 | 2.313 | 5.851 | |
| 2 | 5.808 | 10.923 | NA | 7.853 | 0.148 | 0.122 | 0.222 | 2.133 | 8.012 | 3.089 | 1.277 | 0.416 | 2.572 | |
| 3 | 1.874 | 2.359 | 2.418 | 1.980 | -6.358 | 0.823 | 3.632 | 1.712 | 1.018 | 1.073 | 0.369 | -0.414 | 1.305 | |
| 4 | 4.613 | 6.275 | 3.346 | 3.212 | 0.583 | -0.439 | 1.345 | -0.181 | NA | 5.386 | 1.966 | 1.711 | 12.730 | |
| 5 | 3.356 | 11.994 | 5.453 | 3.882 | -1.264 | 0.569 | 1.314 | 1.281 | 2.036 | 0.729 | 0.581 | 0.610 | 2.370 | |
| 6 | 5.257 | 17.132 | 15.162 | 4.524 | 0.031 | 1.545 | 0.775 | -0.514 | 4.337 | -0.005 | 0.917 | 0.606 | 9.251 | |
| 7 | 5.694 | 12.756 | NA | 4.055 | -2.259 | 2.183 | 2.111 | 9.850 | 4.440 | 2.499 | 1.681 | 1.243 | 3.230 | |
| 8 | 14.016 | 8.887 | 20.273 | 4.499 | -9.093 | -0.795 | -0.012 | 2.625 | 3.492 | 0.617 | 0.960 | -0.828 | 0.452 | |
| 9 | 1.528 | 11.713 | NA | 1.011 | -7.377 | 2.841 | 9.567 | 0.768 | 2.400 | 1.781 | 0.331 | 1.494 | 15.363 | |
| 10 | 9.792 | 2.487 | NA | 1.807 | -3.082 | 0.433 | 0.875 | 1.296 | 4.635 | 2.336 | 1.667 | 3.992 | -0.072 | |
| 11 | 7.197 | 1.944 | NA | 5.579 | -3.385 | 1.283 | -0.016 | 7.930 | 2.669 | 1.763 | 0.958 | 1.241 | 0.332 | |
| 12 | 3.698 | 5.475 | NA | 2.978 | -0.038 | -0.365 | 0.534 | 0.653 | 0.855 | 2.731 | -0.882 | -0.697 | 2.326 | |
| 13 | 4.738 | 6.389 | 2.067 | 3.730 | -1.031 | 2.128 | 1.641 | 2.592 | 3.378 | 3.672 | 1.554 | 0.657 | 9.024 | |
| (b): Kurtosis Variation in the Water Quality Indices in 13-Experimental Sites. | | | | | | | | | | | | | | |
| Parameters/ Sites | Cl- | Fe ²⁺ | As ⁺⁺ | Ni | pН | Ca | Mg | Fl | SO_4^{2-} | TDS | Alkalinity | Hardness | Turbidity | |
| 1 | 56.972 | 41.528 | -0.580 | 6.503 | -0.542 | 42.616 | 23.577 | 1.556 | 36.720 | 11.238 | 17.204 | 5.354 | 37.192 | |
| 2 | 36.427 | 125.136 | NA | 76.130 | -0.306 | -1.294 | -0.882 | 3.472 | 71.894 | 11.006 | 0.940 | -1.073 | 13.487 | |
| 3 | 3.494 | 9.040 | 3.890 | 5.284 | 44.215 | 0.170 | 14.986 | 3.116 | 0.822 | 0.394 | 0.544 | 2.258 | 1.530 | |
| 4 | 23.047 | 53.385 | 9.306 | 8.761 | 1.306 | -0.516 | -0.050 | -1.037 | NA | 35.433 | 3.620 | 2.367 | 163.653 | |
| 5 | 10.825 | 164.378 | 47.773 | 19.055 | -0.348 | -0.321 | 1.734 | 1.503 | 7.979 | -0.568 | -0.586 | -0.361 | 7.101 | |
| 6 | 28.123 | 293.666 | 240.317 | 20.643 | 1.704 | 3.533 | 0.320 | -0.249 | 20.001 | -0.058 | 1.348 | 3.150 | 115.733 | |
| 7 | 38.313 | 223.392 | NA | 16.440 | 3.253 | 4.487 | 4.264 | 150.762 | 18.306 | 6.399 | 2.867 | 0.946 | 11.343 | |
| 8 | 245.498 | 127.897 | 411.000 | 39.478 | 96.398 | -0.746 | -0.585 | 7.950 | 19.345 | -1.409 | -0.319 | 3.340 | 3.711 | |
| 9 | 1.387 | 150.451 | NA | 1.165 | 74.926 | 11.549 | 127.310 | -0.157 | 7.224 | 3.376 | -0.683 | 8.679 | 257.864 | |
| 10 | 100.525 | 15.268 | NA | 6.166 | 8.261 | 0.036 | 0.624 | 0.893 | 37.757 | 9.225 | 2.497 | 25.920 | -0.919 | |
| 11 | 54.539 | 6.625 | NA | 33.459 | 10.514 | 2.522 | -0.547 | 62.923 | 9.809 | 1.799 | 0.589 | 3.236 | 0.445 | |
| 12 | 12.769 | 30.247 | NA | 7.812 | -2.106 | -1.362 | -1.229 | -0.725 | 0.575 | 7.058 | 0.103 | -0.641 | 6.968 | |
| 13 | 23.210 | 47.562 | 2.293 | 19.727 | -0.904 | 7.648 | 2.906 | 6.845 | 15.905 | 16.924 | 3.342 | 1.227 | 98.599 | |

Table 2. (a): Skewness variation in the water quality indices in 13 experimental sites. (b): Kurtosis variation in the water quality indices in 13 experimental sites.

The *H* of a real-valued time series is defined as follows in terms of the exponential growth scaling relation:

$$\left\langle \frac{R(n)}{S(n)} \right\rangle = Cn^{H}$$
, as 'n' approaches to infinity (4)

C indicates constant, angular brackets $\langle \cdots \rangle$ imply the anticipated value, *S*(*n*) is the standard deviation of the initial '*n*' data of the series {*X*₁, *X*₂, ..., *X*_{*n*}}; *R*(*n*) is their range, which is as follows:

$$R(n) = max\{X_1, X_2, \cdots, X_n\} - min\{X_1, X_2, \cdots, X_n\}$$
(5)

H is calculated from the R/S technique and calculated from the wavelets approach for the time series.

3.5.3. Evaluation of Wavelets Approach for H

If f(t) is a self-affine random process and position parameter 't' (i.e., time or distance), a > 0 is a dilation factor and w(t) is an initial wavelet.

$$w_{t,a}(t') = \frac{1}{\sqrt{a}} w\left(\frac{t'-t}{a}\right) \tag{6}$$

If the continuous wavelet transform of f(t) is its shifted $w_{t,a}(t')$, dilated, and scaled version, then f(t) is specified as the following:

$$W(t,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} w_{t,a}(t') f(t') dt'$$
(7)

Trend analysis f(t) is self-affine and the dispersion of W(t, a) will range monotonically with the following expansion factor:

$$V(a) = \langle W^2 \rangle - \langle W^2 \rangle \propto a^{\delta} \tag{8}$$

when the exponent ' δ ' is in the middle of -1 and 3 (i.e., $-1 \le \delta \le 3$), the *H* is specified as the following:

$$H_w = \begin{cases} \frac{\delta+1}{2} & if \quad -1 \le \delta < 1 \quad (\text{FGN})\\ \frac{\delta-1}{2} & if \quad 1 \le \delta \le 3 \quad (\text{FBM}) \end{cases}$$
(9)

FGN implies fractal Gaussian noise, FBM implies fractional Brownian motion; H is defined in terms of fractal dimension (D) [43] as the following:

$$H = 2 - D \tag{10}$$

where *H* is the Hurst exponent; *D* is the fractal dimension lying between 1.0 and 2.0. Now, the predictability index (*PI*) is given as

$$PI = 2|D - 1.5| = 2|0.5 - H|$$
(11)

The data series is uncertain for values of *PI* near to zero, whereas, the data series is predictable for values close to 1.

4. Results and Discussion

4.1. Distribution of the WQPs under Different Stress Conditions: A Quantitative and Qualitative Analysis

Figure 4 shows the quantitative correlation between the major water quality metrics, indicating that fluctuation in these characteristics is caused by variability throughout the originating ecosystem and is influenced by the topographical conditions through which it flows down.

The mean, median, mode, and standard deviation of the daily data series for the year 2016 have been plotted for all 13 experimental sites. In site 3 (Chamoli district), the highest mean value is observed particularly for the Cl^- , Fe^{2+} , As^{++} , NO_3^- and hardness, and apart from this, the TDS is the second highest among all the experimental sites. The mean value and standard deviation (127.6) of Cl^- is observed to be the highest in experiment site 4 (i.e., Champavat). Ca mean concentration is found to vary among the experiment sites.

The mode values of water quality parameters for turbidity, Cl^- , Fe^{2+} , As^{++} , NO_3^- , pH, and SO_4^{2-} are within almost same range (0–1) in all experimental sites, except for Ca, Mg, alkalinity, and hardness, which show the highest variation in the district Nainital, Bageshwer, Champavat, USN and Haridwar, respectively.

The median values of the parameters are also almost within the same range for all experimental sites, except Cl⁻, which is found to have the highest concentration in Chamoli and Cahasahigh, in Pithoragarh, followed by Bageshwer, USN, Chamoli district. The concentration of TDS is observed to be high in the experimental sites, i.e., Nainital, Pithoragarh, Almora, Uttarkashi, USN, Champavat, and Bageswar. The highest alkalinity is observed in Nainital followed by Uttarakashi, Pithoragarh, Bageswer, Pauri Garhwal, and Almora districts. Hardness concentration is found to be the highest in Haridwar and followed by Pithoragarh, Bageshwer, Dehradun, USN, Uttarakashi, and least in Almora.

The standard deviations of the parameters such as turbidity, Fe^{2+} , As^{++} , NO_3^- , pH, and F^- are within almost the same ranges among all experimental sites, except some

fluctuations were observed for Cl⁻, Ca, Mg, TDS, alkalinity, hardness and SO_4^{2-} . The statistical parameters, i.e., mean, mode, median, and standard deviation of hardness sampling, are found more in experimental site 3 (Chamoli district). The variation in the turbidity, Cl⁻, Fe²⁺, NO₃⁻, pH and F⁻ parameters have almost reached the satisfaction level in all experimental sites.



Figure 4. Statistical analysis of the GWQ parameters variation under different stress conditions.

4.2. Spatial Variation in WQPs in Different Stress Conditions

Geo-spatial maps are prepared using remote sensing and GIS techniques, which represent the distribution of WQPs over the region. The geo-spatial maps of the GWQ parameters under different geographical and geological stress conditions over the 13 districts of state Uttarakhand, which reveals through the extrapolation maps the distribution in different stress conditions, are shown in Figure 5.



Figure 5. Cont.



Figure 5. Cont.



Figure 5. Cont.



Figure 5. Geospatial maps of the GWQ parameters under different geographical and geological stress conditions.

Minimum and maximum distribution of the WQPs is observed in the studied region of the state Uttarakhand. The minimum and maximum observed point values of As⁺⁺ are 0.0–0.007 mg/L. The maximum range is observed at the Rudrapryag district. The Cl distribution minimum of 4.43 mg/L and maximum of 36.1 mg/L content level in water is observed in the study region and higher quantities are found in the Rudrapryag and Champavat districts. Ca concentration is observed to be higher in the Pithoragarh, Bageshwar and Haridwar regions with minimum 7.77 mg/L and maximum 123.4 mg/L distribution. Higher 306.6 mg/L alkalinity in water quality is observed in the Haridwar district. Total dissolve salt concentration (TDS) varies from the minimum to maximum range, which is 3.08 to 258.5 mg/L, in the study region and a high level is observed in Nainital district.

 $\rm F^-$ varies from 0.10 to 2.66 mg/L and a higher concentration is observed in the USN district. The hardness level in the water quality is observed to vary from 45.1 to 371.8 mg/L in the whole study region and the highest concentration is found in the Haridwar district. Mg content level in the water samples is found to be between 5.03 and 55.9 mg/L and higher values are observed in Pithoragarh, followed by the Bageshwar district. Higher levels of SO⁴⁻ is observed in the USN district water level, which varies between 0.012 and 40.5 mg/L. NO₃⁻ level is estimated and a higher level is observed in the Haridwar district, which is between 1.36 and 14.6 mg/L. The pH values lie almost in the range, except in Haridwar, Rudrapyag and some regions of the Dehradun and Almora districts. Fe²⁺ concentration's minimum is 0.03 mg/L and maximum 1.05 mg/L and a higher concentration is found in the Haridwar region. Turbidity level in the water samples is found in the range of 0.35 to 11.9 NTU and a higher concentration is observed in the Almora district.

4.3. WQI Distribution and Predictability Analysis

Figure 6 demonstrated the variation in the WQIs in different experimental sites (1-13) and all WQPs in each site. The Hurst exponent (*H*), fractal dimension (FD), and predictability (PI) obtained are graphically represented to show the anomalies among the parameters. The interpretation of these indices has been carried out for all 13 experimental sites and site-wise descriptions of the indices are demonstrated and analyzed.



Figure 6. Cont.



Figure 6. Cont.







Figure 6. Variation in the Hurts exponent (H), fractal dimension (D) and predictability (PI) of the water quality parameters in the 13 different locations of the State Uttarakhand.

For experimental site 1, the turbidity has a persistent relationship with each WQI, i.e., NO_3^- , Fe^{2+} , As^{++} , pH, Ca, Mg, Fl, SO_4^{2-} , TDS, alkalinity and hardness, but has an anti-persistent relationship with Cl⁻. Similarly, Cl⁻, Fe^{2+} , NO_3^- and F^- have shown persistent behavior with all WQIs, whereas the pH, Ca, Mg, TDS, alkalinity, and SO_4^{2-} have shown anti-persistent behavior with all WQIs. For experimental site 2, the turbidity has a persistent relationship with each WQI, i.e., Fe^{2+} , Ni, pH, Ca, Mg, Fl, SO_4^{2-} , TDS, alkalinity and hardness, but has an anti-persistent relationship with Cl⁻, and As^{++} . Similarly, Cl⁻, Fe^{2+} , NO_3^- and F^- have shown persistent behavior with all WQIs, whereas the pH, Ca, Mg, TDS, alkalinity, and SO_4^{2-} have shown anti-persistent behavior with all WQIs.

For experimental site 3, the parameters that have shown the most persistent relationship are turbidity, Cl^- , Fe^{2+} , As^{++} Ni, Mg, SO_4^{2-} , and partially F^- , but the parameters pH, Ca, TDS, alkalinity and hardness have an anti-persistent relationship. Similarly, for site 4, the persistent behavior is observed between the turbidity, Cl^- , Fe^{2+} , As^{++} , NO_3^- , Ca, partially Mg, TDS and SO_4^{2-} . The parameters pH, Ca, F^- , alkalinity, and hardness show anti-persistent behavior among other WQIs. In site 5, persistent behavior is observed between almost all WQIs, except Ca, TDS, and alkalinity. In experimental site 6, the persistence relationship is observed among almost all the WQIs, except F⁻. In site 7, persistence behavior is observed among Fe²⁺, pH, Ca, Mg, and hardness, but the anti-persistence relationship is observed partially in turbidity, Cl⁻, NO₃⁻, F⁻, TDS, SO₄²⁻ and alkalinity. In experimental site 8, all the WQIs, such as turbidity, Cl⁻, Fe²⁺, NO₃⁻ and TDS, have shown persistence behavior except pH, Ca, Mg, and partially hardness and SO₄²⁻, which have shown the anti-persistence relationship. In experimental site 9, turbidity, Cl⁻, Fe²⁺, pH, Ca, Mg, NO₃⁻ and TDS have shown persistence behavior and F⁻, hardness and SO₄²⁻ have shown the anti-persistence relationship. In experimental site -0, turbidity, Cl⁻, Fe²⁺, NO₃⁻, TDS and alkalinity have shown persistence behavior, and pH, Ca, Mg, F⁻, hardness and SO₄²⁻ have shown the anti-persistence relationship. In the experimental site 11, all the studied WQIs have shown persistence behavior except the pH parameter and in site 12, all WQIs have shown unpredictability among each parameter.

In experimental site 13, turbidity, Fe^{2+} , NO_3^- , F^- , TDS and alkalinity have shown less predictability, whereas Cl⁻, pH, Ca, Mg, hardness, and SO_4^{2-} have shown the antipersistence relationship.

4.4. WQPs Analysis

In this section, the WQPs are analyzed by the variation in each parameter under the different stress conditions or experimental sites. The variation in the parameters is analyzed by their distribution statistics, such as mean, median, mode and standard deviation (Figure 4) among the datasets and spatial variation in terms of skewness and kurtosis are represented in Figure 7. The correlation matrix is established among the different WQPs and is summarized in Table 3 for each experimental site (1–13). Table 3 describes the correlation among the WQP of the 13 sites, where it presents the correlation of one WQP with another WQP. In site '1', the alkalinity is strongly correlated with Ca and Mg. Hardness is also strongly correlated with Mg. It has been observed from 13 sites that the alkalinity and hardness have a strong correlation ratio with Mg, Ca and TDS, except in site 2, site 5, site 6, and site 7. This table demonstrates that the WQPs are within the standard values except a few sites, as the alkalinity is high in these few sites. Considering skewness, kurtosis, SD and correlation matrix results among the different data series over the different sites 1–13, a comprehensive analysis has been carried out and described in the following sub sections.

| | | | | | | Site | e '1' | | | | | | |
|---------------------------|-----------|-------|-------|-------|-------------------|-------|-------|-------|------------------|-------|------------|----------|-------------|
| Parameters/r ² | Turbidity | Cl- | Fe | Ar | NO ₃ - | pН | Ca | Mg | \mathbf{F}^{-} | TDS | Alkalinity | Hardness | SO_4^{2-} |
| Turbidity | 1 | | | | | | | | | | | | |
| Cl- | 0.021 | 1 | | | | | | | | | | | |
| Fe | 0.267 | 0.013 | 1 | | | | | | | | | | |
| Ar | 0.160 | 0.031 | 0.113 | 1 | | | | | | | | | |
| NO ₃ - | 0.002 | 0.182 | 0.004 | 0.019 | 1 | | | | | | | | |
| pH | 0.015 | 0.011 | 0.013 | 0.114 | 0.089 | 1 | | | | | | | |
| Ca | 0.002 | 0.059 | 0.001 | 0.010 | 0.020 | 0.009 | 1 | | | | | | |
| Mg | 0.006 | 0.135 | 0.005 | 0.034 | 0.107 | 0.039 | 0.488 | 1 | | | | | |
| Fl | 0.000 | 0.011 | 0.001 | 0.008 | 0.067 | 0.002 | 0.012 | 0.019 | 1 | | | | |
| TDS | 0.032 | 0.115 | 0.025 | 0.071 | 0.055 | 0.277 | 0.005 | 0.098 | 0.020 | 1 | | | |
| Alkalinity | 0.003 | 0.083 | 0.001 | 0.013 | 0.356 | 0.013 | 0.724 | 0.611 | 0.031 | 0.041 | 1 | | |
| Hardness | 0.044 | 0.208 | 0.030 | 0.210 | 0.015 | 0.084 | 0.330 | 0.590 | 0.011 | 0.150 | 0.346 | 1 | |
| SO_4^{2-} | 0.000 | 0.356 | 0.004 | 0.000 | 0.011 | 0.007 | 0.152 | 0.229 | 0.001 | 0.088 | 0.267 | 0.146 | 1 |

Table 3. Correlation matrix for water quality parameters over different sites 1–13.

| | | | | | | Site | e '2' | | | | | | |
|---------------------------|-----------|----------|-------|-------|------------------------------|-------|-------|-------|------------------|-------|------------|----------|-------------|
| Parameters/r ² | Turbidity | Cl^{-} | Fe | Ar | NO_3^- | pН | Ca | Mg | F^{-} | TDS | Alkalinity | Hardness | SO_4^{2-} |
| Turbidity | 1 | | | | | | | | | | | | |
| Cl- | 0.018 | 1 | | | | | | | | | | | |
| Fe | 0.015 | 0.001 | 1 | | | | | | | | | | |
| Ar | 0.050 | 0.006 | 0.026 | 1 | | | | | | | | | |
| NO ₃ - | 0.000 | 0.017 | 0.016 | 0.291 | 1 | | | | | | | | |
| pH | 0.035 | 0.014 | 0.002 | 0.025 | 0.001 | 1 | | | | | | | |
| Ca | 0.002 | 0.000 | 0.097 | 0.004 | 0.024 | 0.000 | 1 | | | | | | |
| Mg | 0.020 | 0.001 | 0.109 | 0.003 | 0.262 | 0.009 | 0.126 | 1 | | | | | |
| Fl | 0.011 | 0.000 | 0.008 | 0.013 | 0.002 | 0.004 | 0.229 | 0.000 | 1 | | | | |
| TDS | 0.026 | 0.000 | 0.007 | 0.197 | 0.002 | 0.003 | 0.051 | 0.032 | 0.000 | 1 | | | |
| Alkalinity | 0.066 | 0.025 | 0.119 | 0.000 | 0.096 | 0.010 | 0.398 | 0.041 | 0.185 | 0.119 | 1 | | |
| Hardness | 0.103 | 0.064 | 0.044 | 0.016 | 0.035 | 0.013 | 0.231 | 0.060 | 0.089 | 0.099 | 0.473 | 1 | |
| SO_4^{2-} | 0.014 | 0.007 | 0.024 | 0.004 | 0.014 | 0.001 | 0.011 | 0.005 | 0.022 | 0.167 | 0.001 | 0.059 | 1 |
| | | | | | | Site | e '3' | | | | | | |
| Parameters/r ² | Turbidity | Cl- | Fe | Ar | NO ₃ - | pН | Ca | Mg | F ⁻ | TDS | Alkalinity | Hardness | SO_4^{2-} |
| Turbidity | 1 | | | | | | | | | | | | |
| Cl- | 0.207 | 1 | | | | | | | | | | | |
| Fe | 0.036 | 0.005 | 1 | | | | | | | | | | |
| Ar | 0.100 | 0.028 | 0.045 | 1 | | | | | | | | | |
| NO ₃ - | 0.079 | 0.044 | 0.016 | 0.440 | 1 | | | | | | | | |
| pН | 0.221 | 0.053 | 0.027 | 0.038 | 0.111 | 1 | | | | | | | |
| Ca | 0.173 | 0.060 | 0.005 | 0.009 | 0.127 | 0.417 | 1 | | | | | | |
| Mg | 0.277 | 0.07 | 0.03 | 0.07 | 0.21 | 0.28 | 0.27 | 1 | | | | | |
| Fl | 0.210 | 0.089 | 0.064 | 0.039 | 0.100 | 0.246 | 0.120 | 0.259 | 1 | | | | |
| TDS | 0.320 | 0.060 | 0.015 | 0.110 | 0.124 | 0.350 | 0.346 | 0.401 | 0.252 | 1 | | | |
| Alkalinity | 0.224 | 0.008 | 0.023 | 0.081 | 0.133 | 0.407 | 0.491 | 0.439 | 0.200 | 0.665 | 1 | | |
| Hardness | 0.366 | 0.171 | 0.024 | 0.091 | 0.221 | 0.442 | 0.520 | 0.553 | 0.346 | 0.725 | 0.701 | 1 | |
| SO_4^{2-} | 0.208 | 0.192 | 0.025 | 0.088 | 0.053 | 0.228 | 0.224 | 0.383 | 0.342 | 0.414 | 0.363 | 0.570 | 1 |
| | | | | | | Site | e '4' | | | | | | |
| Parameters/r ² | Turbidity | Cl^{-} | Fe | Ar | NO ₃ ⁻ | pН | Ca | Mg | \mathbf{F}^{-} | TDS | Alkalinity | Hardness | SO_4^{2-} |
| Turbidity | 1 | | | | | | | | | | | | |
| Cl- | 0.001 | 1 | | | | | | | | | | | |
| Fe | 0.049 | 0.002 | 1 | | | | | | | | | | |
| Ar | 0.052 | 0.002 | 0.046 | 1 | | | | | | | | | |
| NO ₃ - | 0.025 | 0.036 | 0.020 | 0.276 | 1 | | | | | | | | |
| pН | 0.007 | 0.025 | 0.035 | 0.064 | 0.050 | 1 | | | | | | | |
| Ca | 0.045 | 0.017 | 0.062 | 0.009 | 0.027 | 0.280 | 1 | | | | | | |
| Mg | 0.015 | 0.001 | 0.029 | 0.026 | 0.005 | 0.262 | 0.500 | 1 | | | | | |
| Fl | 0.004 | 0.153 | 0.061 | 0.031 | 0.144 | 0.138 | 0.087 | 0.059 | 1 | | | | |
| TDS | 0.006 | 0.001 | 0.002 | 0.001 | 0.150 | 0.056 | 0.040 | 0.021 | 0.013 | 1 | | | |
| Alkalinity | 0.007 | 0.077 | 0.030 | 0.002 | 0.067 | 0.332 | 0.496 | 0.201 | 0.068 | 0.025 | 1 | | |
| Hardness | 0.023 | 0.031 | 0.054 | 0.029 | 0.007 | 0.594 | 0.709 | 0.424 | 0.106 | 0.057 | 0.655 | 1 | |
| SO_4^2 | 0.029 | 0.018 | 0.007 | 0.018 | 0.025 | 0.154 | 0.172 | 0.046 | 0.074 | 0.006 | 0.215 | 0.284 | 1 |

Table 3. Cont.

| | | | | | | Site | e '5' | | | | | | |
|---------------------------|-----------|----------|-------|-------------------|-------------------|-------|-------|-------|-------|-------|------------|----------|-------------|
| Parameters/r ² | Turbidity | Cl^{-} | Fe | Ar | NO_3^- | pН | Ca | Mg | Fl | TDS | Alkalinity | Hardness | SO_4^{2-} |
| Turbidity | 1 | | | | | | | | | | | | |
| Cl- | 0.168 | 1 | | | | | | | | | | | |
| Fe | 0.019 | 0.003 | 1 | | | | | | | | | | |
| Ar | 0.004 | 0.003 | 0.001 | 1 | | | | | | | | | |
| NO ₃ - | 0.000 | 0.000 | 0.000 | 0.002 | 1 | | | | | | | | |
| pH | 0.001 | 0.001 | 0.004 | 0.000 | 0.000 | 1 | | | | | | | |
| Ca | 0.020 | 0.027 | 0.067 | 0.003 | 0.045 | 0.070 | 1 | | | | | | |
| Mg | 0.024 | 0.014 | 0.000 | 0.000 | 0.000 | 0.027 | 0.348 | 1 | | | | | |
| Fl | 0.001 | 0.016 | 0.030 | 0.005 | 0.020 | 0.000 | 0.040 | 0.001 | 1 | | | | |
| TDS | 0.047 | 0.002 | 0.196 | 0.002 | 0.015 | 0.014 | 0.272 | 0.002 | 0.098 | 1 | | | |
| Alkalinity | 0.086 | 0.117 | 0.022 | 0.006 | 0.130 | 0.011 | 0.022 | 0.012 | 0.169 | 0.093 | 1 | | |
| Hardness | 0.035 | 0.008 | 0.011 | 0.001 | 0.001 | 0.008 | 0.113 | 0.152 | 0.036 | 0.001 | 0.130 | 1 | |
| SO_4^{2-} | 0.002 | 0.010 | 0.016 | 0.004 | 0.001 | 0.007 | 0.002 | 0.007 | 0.166 | 0.023 | 0.037 | 0.033 | 1 |
| | | | | | | Site | e '6' | | | | | | |
| Parameters/r ² | Turbidity | Cl- | Fe | Ar | NO ₃ - | pН | Ca | Mg | Fl | TDS | Alkalinity | Hardness | SO_4^{2-} |
| Turbidity | 1 | | | | | | | | | | | | |
| Cl- | 0.004 | 1 | | | | | | | | | | | |
| Fe | 0.000 | 0.000 | 1 | | | | | | | | | | |
| Ar | 0.000 | 0.000 | 0.000 | 1 | | | | | | | | | |
| NO ₃ - | 0.003 | 0.415 | 0.000 | 0.106 | 1 | | | | | | | | |
| pH | 0.019 | 0.025 | 0.003 | 0.001 | 0.008 | 1 | | | | | | | |
| Ca | 0.000 | 0.130 | 0.001 | 0.001 | 0.024 | 0.002 | 1 | | | | | | |
| Mg | 0.002 | 0.036 | 0.007 | 0.001 | 0.000 | 0.000 | 0.499 | 1 | | | | | |
| Fl | 0.001 | 0.040 | 0.001 | 0.008 | 0.038 | 0.054 | 0.001 | 0.009 | 1 | | | | |
| TDS | 0.006 | 0.010 | 0.006 | 0.001 | 0.006 | 0.043 | 0.001 | 0.004 | 0.086 | 1 | | | |
| Alkalinity | 0.015 | 0.046 | 0.002 | 0.001 | 0.438 | 0.011 | 0.215 | 0.061 | 0.063 | 0.068 | 1 | | |
| Hardness | 0.007 | 0.008 | 0.000 | 0.000 | 0.019 | 0.041 | 0.151 | 0.068 | 0.012 | 0.175 | 0.485 | 1 | |
| SO_4^{2-} | 0.005 | 0.351 | 0.000 | 0.000 | 0.025 | 0.003 | 0.088 | 0.003 | 0.077 | 0.019 | 0.118 | 0.125 | 1 |
| | | | | | | Site | e '7' | | | | | | |
| Parameters/r ² | Turbidity | Cl- | Fe | NO ₃ - | p] | Н | Ca | Mg | Fl | TDS | Alkalinity | Hardness | SO_4^{2-} |
| Turbidity | 1 | | | | | | | | | | | | |
| C1- | 0.014 | 1 | | | | | | | | | | | |
| Fe | 0.033 | 0.001 | 1 | | | | | | | | | | |
| NO ₃ - | 0.025 | 0.050 | 0.011 | 1 | | | | | | | | | |
| pH | 0.001 | 0.015 | 0.001 | 0.022 | 1 | l | | | | | | | |
| Ca | 0.019 | 0.001 | 0.015 | 0.090 | 0.0 | 07 | 1 | | | | | | |
| Mg | 0.007 | 0.019 | 0.001 | 0.090 | 0.0 | 04 | 0.305 | 1 | | | | | |
| Fl | 0.097 | 0.004 | 0.002 | 0.069 | 0.0 | 06 | 0.020 | 0.050 | 1 | | | | |
| TDS | 0.040 | 0.060 | 0.000 | 0.049 | 0.0 | 00 | 0.000 | 0.005 | 0.031 | 1 | | | |
| Alkalinity | 0.101 | 0.075 | 0.006 | 0.000 | 0.0 | 01 | 0.151 | 0.017 | 0.071 | 0.136 | 1 | | |
| Hardness | 0.035 | 0.026 | 0.003 | 0.005 | 0.0 | 00 | 0.069 | 0.213 | 0.013 | 0.091 | 0.110 | 1 | |
| SO4 ²⁻ | 0.006 | 0.210 | 0.019 | 0.003 | 0.0 | 07 | 0.044 | 0.001 | 0.104 | 0.066 | 0.230 | 0.072 | 1 |

Table 3. Cont.

| | | | | | Si | ite '8' | | | | | | |
|---------------------------|-----------|-----------------|-------|-------------------|-------|---------|-------|-------|-------|------------|----------|-------------------|
| Parameters/r ² | Turbidity | Cl ⁻ | Fe | NO_3^- | pН | Ca | Mg | Fl | TDS | Alkalinity | Hardness | SO_4^{2-} |
| Turbidity | 1 | | | | | | | | | | | |
| Cl- | 0.022 | 1 | | | | | | | | | | |
| Fe | 0.001 | 0.000 | 1 | | | | | | | | | |
| NO ₃ - | 0.064 | 0.002 | 0.001 | 1 | | | | | | | | |
| pH | 0.104 | 0.001 | 0.000 | 0.055 | 1 | | | | | | | |
| Ca | 0.299 | 0.063 | 0.000 | 0.036 | 0.100 | 1 | | | | | | |
| Mg | 0.357 | 0.080 | 0.000 | 0.064 | 0.131 | 0.787 | 1 | | | | | |
| Fl | 0.050 | 0.000 | 0.002 | 0.176 | 0.050 | 0.003 | 0.035 | 1 | | | | |
| TDS | 0.159 | 0.004 | 0.001 | 0.041 | 0.048 | 0.034 | 0.070 | 0.243 | 1 | | | |
| Alkalinity | 0.228 | 0.059 | 0.003 | 0.092 | 0.146 | 0.139 | 0.241 | 0.415 | 0.521 | 1 | | |
| Hardness | 0.356 | 0.046 | 0.001 | 0.137 | 0.152 | 0.540 | 0.623 | 0.344 | 0.308 | 0.616 | 1 | |
| SO_4^{2-} | 0.012 | 0.000 | 0.000 | 0.005 | 0.009 | 0.014 | 0.022 | 0.064 | 0.000 | 0.030 | 0.039 | 1 |
| | | | | | Si | ite '9' | | | | | | |
| Parameters/r ² | Turbidity | Cl ⁻ | Fe | NO_3^- | pН | Ca | Mg | Fl | TDS | Alkalinity | Hardness | SO_4^{2-} |
| Turbidity | 1 | | | | | | | | | | | |
| Cl- | 0.001 | 1 | | | | | | | | | | |
| Fe | 0.098 | 0.004 | 1 | | | | | | | | | |
| NO ₃ - | 0.000 | 0.004 | 0.112 | 1 | | | | | | | | |
| pH | 0.000 | 0.800 | 0.003 | 0.000 | 1 | | | | | | | |
| Ca | 0.048 | 0.004 | 0.037 | 0.150 | 0.006 | 1 | | | | | | |
| Mg | 0.000 | 0.057 | 0.003 | 0.031 | 0.085 | 0.000 | 1 | | | | | |
| Fl | 0.000 | 0.000 | 0.056 | 0.001 | 0.001 | 0.002 | 0.398 | 1 | | | | |
| TDS | 0.011 | 0.116 | 0.014 | 0.013 | 0.150 | 0.012 | 0.006 | 0.024 | 1 | | | |
| Alkalinity | 0.431 | 0.012 | 0.120 | 0.188 | 0.004 | 0.188 | 0.014 | 0.009 | 0.000 | 1 | | |
| Hardness | 0.002 | 0.470 | 0.001 | 0.014 | 0.625 | 0.003 | 0.017 | 0.216 | 0.080 | 0.023 | 1 | |
| SO4 ²⁻ | 0.004 | 0.525 | 0.003 | 0.006 | 0.605 | 0.000 | 0.020 | 0.300 | 0.050 | 0.018 | 0.848 | 1 |
| | | | | | Si | te '10' | | | | | | |
| Parameters/r ² | Turbidity | Cl ⁻ | Fe | NO ₃ - | pН | Ca | Mg | Fl | TDS | Alkalinity | Hardness | SO4 ²⁻ |
| Turbidity | 1 | | | | | | | | | | | |
| Cl- | 0.000 | 1 | | | | | | | | | | |
| Fe | 0.001 | 0.004 | 1 | | | | | | | | | |
| NO ₃ - | 0.013 | 0.001 | 0.006 | 1 | | | | | | | | |
| pH | 0.003 | 0.081 | 0.031 | 0.019 | 1 | | | | | | | |
| Ca | 0.010 | 0.099 | 0.007 | 0.045 | 0.394 | 1 | | | | | | |
| Mg | 0.226 | 0.031 | 0.037 | 0.046 | 0.017 | 0.153 | 1 | | | | | |
| Fl | 0.018 | 0.032 | 0.176 | 0.038 | 0.225 | 0.602 | 0.191 | 1 | | | | |
| TDS | 0.002 | 0.795 | 0.010 | 0.007 | 0.042 | 0.011 | 0.010 | 0.002 | 1 | | | |
| Alkalinity | 0.025 | 0.276 | 0.002 | 0.016 | 0.074 | 0.096 | 0.182 | 0.093 | 0.167 | 1 | | |
| Hardness | 0.030 | 0.042 | 0.004 | 0.016 | 0.071 | 0.201 | 0.410 | 0.270 | 0.000 | 0.417 | 1 | |
| SO42- | 0.007 | 0.003 | 0.025 | 0.085 | 0.261 | 0.372 | 0.238 | 0.518 | 0.008 | 0.037 | 0.211 | 1 |
| | | | | | | | | | | | | |

Table 3. Cont.

| | | | | | Sit | te '11' | | | | | | |
|------------------------------|-----------|-----------------|-------|-------------------|-------|---------|-------|-------|-------|------------|----------|-------------------|
| Parameters/r ² | Turbidity | Cl ⁻ | Fe | NO_3^- | pН | Ca | Mg | Fl | TDS | Alkalinity | Hardness | SO_4^{2-} |
| Turbidity | 1 | | | | | | | | | | | |
| Cl- | 0.003 | 1 | | | | | | | | | | |
| Fe | 0.005 | 0.030 | 1 | | | | | | | | | |
| NO ₃ - | 0.000 | 0.789 | 0.049 | 1 | | | | | | | | |
| pH | 0.028 | 0.006 | 0.043 | 0.020 | 1 | | | | | | | |
| Ca | 0.003 | 0.067 | 0.003 | 0.058 | 0.032 | 1 | | | | | | |
| Mg | 0.001 | 0.361 | 0.011 | 0.420 | 0.038 | 0.622 | 1 | | | | | |
| Fl | 0.030 | 0.205 | 0.010 | 0.242 | 0.028 | 0.045 | 0.228 | 1 | | | | |
| TDS | 0.000 | 0.537 | 0.030 | 0.530 | 0.059 | 0.090 | 0.371 | 0.198 | 1 | | | |
| Alkalinity | 0.013 | 0.419 | 0.013 | 0.386 | 0.054 | 0.379 | 0.568 | 0.166 | 0.606 | 1 | | |
| Hardness | 0.009 | 0.303 | 0.004 | 0.308 | 0.082 | 0.445 | 0.661 | 0.209 | 0.572 | 0.747 | 1 | |
| SO_4^{2-} | 0.003 | 0.840 | 0.029 | 0.826 | 0.006 | 0.076 | 0.448 | 0.236 | 0.563 | 0.462 | 0.330 | 1 |
| | | | | | Si | te '12' | | | | | | |
| Parameters/r ² | Turbidity | Cl ⁻ | Fe | NO ₃ - | pН | Ca | Mg | Fl | TDS | Alkalinity | Hardness | SO4 ²⁻ |
| Turbidity | 1 | | | | | | | | | | | |
| Cl- | 0.051 | 1 | | | | | | | | | | |
| Fe | 0.001 | 0.377 | 1 | | | | | | | | | |
| NO ₃ - | 0.011 | 0.626 | 0.296 | 1 | | | | | | | | |
| pH | 0.190 | 0.000 | 0.107 | 0.015 | 1 | | | | | | | |
| Ca | 0.105 | 0.003 | 0.000 | 0.045 | 0.137 | 1 | | | | | | |
| Mg | 0.239 | 0.004 | 0.005 | 0.017 | 0.169 | 0.408 | 1 | | | | | |
| Fl | 0.191 | 0.001 | 0.018 | 0.000 | 0.000 | 0.018 | 0.061 | 1 | | | | |
| TDS | 0.007 | 0.094 | 0.000 | 0.082 | 0.008 | 0.096 | 0.253 | 0.005 | 1 | | | |
| Alkalinity | 0.023 | 0.079 | 0.005 | 0.144 | 0.100 | 0.368 | 0.518 | 0.023 | 0.636 | 1 | | |
| Hardness | 0.191 | 0.000 | 0.001 | 0.022 | 0.160 | 0.304 | 0.341 | 0.293 | 0.083 | 0.196 | 1 | |
| SO_4^{2-} | 0.037 | 0.265 | 0.113 | 0.325 | 0.053 | 0.279 | 0.212 | 0.005 | 0.222 | 0.342 | 0.083 | 1 |
| | | | | | Si | te '13' | | | | | | |
| Parameters/r ² | Turbidity | Cl^{-} | Fe | NO_3^- | pН | Ca | Mg | Fl | TDS | Alkalinity | Hardness | SO_4^{2-} |
| Turbidity | 1 | | | | | | | | | | | |
| Cl- | 0.007 | 1 | | | | | | | | | | |
| Fe | 0.005 | 0.186 | 1 | | | | | | | | | |
| NO ₃ ⁻ | 0.173 | 0.007 | 0.016 | 1 | | | | | | | | |
| pH | 0.067 | 0.000 | 0.000 | 0.022 | 1 | | | | | | | |
| Ca | 0.007 | 0.004 | 0.056 | 0.016 | 0.031 | 1 | | | | | | |
| Mg | 0.029 | 0.032 | 0.190 | 0.001 | 0.083 | 0.008 | 1 | | | | | |
| Fl | 0.005 | 0.056 | 0.224 | 0.015 | 0.002 | 0.242 | 0.143 | 1 | | | | |
| TDS | 0.002 | 0.072 | 0.262 | 0.000 | 0.053 | 0.194 | 0.037 | 0.255 | 1 | | | |
| Alkalinity | 0.001 | 0.072 | 0.237 | 0.001 | 0.012 | 0.488 | 0.117 | 0.397 | 0.536 | 1 | | |
| Hardness | 0.001 | 0.046 | 0.006 | 0.000 | 0.010 | 0.015 | 0.225 | 0.028 | 0.005 | 0.031 | 1 | |
| SO4 ²⁻ | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.214 | 0.045 | 0.016 | 0.005 | 0.064 | 0.011 | 1 |
| | | | | | | | | | | | | |

Table 3. Cont.



Figure 7. Cont.





Figure 7. Kurtosis and skewness curve for each (13) parameters over 13 experimental site.

4.4.1. Variation in Turbidity, Alkalinity, Hardness, pH and TDS

The turbidity in almost all the experimental sites has a normal distribution, with the lowest mean value (0.352 mg/L) and highest (11.9 mg/L) in the Pauri Garhwal and Almora districts, respectively. The median (0 to 2.3), and mode (0–1)) values show roughly equal variations among all the sites. The sample data points are found to have a standard deviation of the lowest value of 1.04 in USN and highest of 45.1 in Almora, and 31.1 in the Rudrapyarg district. The SD values indicate that the WQPs are close together for other experimental sites, except for the Almora and Rudraprayag districts. The curve between skewness and kurtosis is not symmetrical in most of the cases. Figure 7 reveals that the positive skewness and kurtosis value indicates that the data series are platykurtic for all sites except the Uttarkashi district, where it shows negative variation.

Alkalinity has shown the higher average value, which indicates the alkaline nature in the data series. In most of the experimental sites, the alkalinity has shown the highest median and mode values that are not substantially identical, indicating that the sample data are largely distributed.

In a few experimental sites, the dataset shows a high standard deviation and a skewness value that is close to zero. The curve between the skewness and kurtosis indicates that the series is symmetrical and platykurtic. In most of the experimental sites, alkalinity has shown persistence behavior with Cl⁻, F-, TDS, and SO₄²⁻ parameters, and it shows Brownian motion, while Alkalinity has also shown an anti-persistence nature with turbidity, Fe²⁺, pH, Ca, Mg, and hardness parameters in few cases.

The hardness of the WQPs has shown large mean, median, and standard deviation values for almost all the studied experimental sites, which shows the distributed datasets that are diversely significant from the mode value and the data series does not behave normally. While the M=mode has zero values in most of the sites, a few sites have large values, especially for the locations of the Kumaun region. In most of the experimental sites, the datasets have a skewness value that is close to one or zero. The curve between skewness and kurtosis indicates that the series is symmetrical and platykurtic. The data series of the few sites have negative skewness values with the Brownian time series. Turbidity, Cl^- , NO_3^- , pH, Ca, F-, TDS, alkalinity, and SO_4^{2-} all exhibit persistent activity, while Fe parameters exhibit anti persistent behavior.

The pH of the water samples is observed to be in the standard range in all the experimental sites, with average and median values being nearly identical, at 7.189 and 7.20, respectively; however, the pH mode is 0.25. These numbers are nearly comparable and indicate that the behavior is normal and the pH is symmetrical because the standard deviation (SD) is 0.691. The negative skewness and corresponding negative kurtosis are observed in almost all the experimental sites, except in the Tehri and Pithogagarh site which shows high kurtosis; hence, except for these two sites, the curve is platykurtic. It is observed that the element pH has the Brownian time-series behavior with F^- , persistent behavior with Ca, Mg, Fe²⁺ and NO₃⁻, and anti-persistent behavior with various parameters such as turbidity, Cl⁻, TDS, alkalinity, hardness and SO₄²⁻.

The TDS is observed for all the experimental site WQPs and it can be observed that the fitness curve does not follow the usual behavior because the mean, median, and mode values have a highly varied nature. The TDS values are not closely related within the time series dataset for all the experimental sites; hence, the standard deviation is very high. TDS has shown persistence behavior with the distribution of other elements, such as Cl^- , pH, Fl, and SO_4^{2-} , TDS, and turbidity, which exhibit a Brownian time series. It has a negative skew and a platykurtic curve for a few experimental sites. NO_3^- , Cl^- and alkalinity have persistent behavior in almost all the experimental sites, while Fe²⁺, Ca, Mg, and hardness parameters have anti-persistent activity.

4.4.2. Variation in Chloride, Flouride, Nitriate and Sulphate

The Cl⁻ concentration varies with mean values from the lowest (0.46 mg/L) to the highest (35.2 mg/L) across the experimental sites, i.e., Bageshwar and Rudraprayag districts, respectively. Even the median varies from 0 to 1, except 3.1 and 9 for the Pithoragrah and Almora districts, respectively, although the mode values are in the order of 0, indicating that the data are laid normally between the sample points and there is also a normal distribution of the SD, except for the large standard deviation, i.e., 135.8 and 127.6 for the Champavat and Rudrapyagh districts. The kurtosis and skewness variation has shown that the curve is symmetrical, except for some experimental sites with larger kurtosis. With the parameters F^- , SO_4^{2-} , and turbidity, Cl⁻ has Brownian time series (true random walk) behavior. As a result, the curve has a platykurtic shape. Turbidity, NO_3^- , TDS and alkalinity show persistent behavior, while Fe^{2+} , pH, Ca, Mg, and hardness show anti-persistent behavior.

The F⁻ element datasets for all the experimental sites have shown normal behavior and also the mean and median values are closely related. The skewness and kurtosis curves indicate the platykurtic nature of all the sites, which also have a lower standard deviation value, demonstrating the closeness between the data points. F⁻ has a Brownian time series (true random walk) with all the studied parameters, such as NO₃⁻ and hardness parameters. F⁻ displays persistent behavior with turbidity, Cl⁻, SO₄²⁻, TDS, and alkalinity, as well as anti-persistent behavior with Fe²⁺, pH, Ca, and Mg in a few experimental sites.

It is observed that for all the experimental sites, the NO_3^- element varies with normal distribution and it is evident from the series mean and median values. Apart from these, the standard deviation of the dataset overall experimental sites appears to be within the standard values, and this denotes that the sample data are near one another. The skewness and kurtosis variation is symmetrical; hence, the platykurtic curve, as with kurtosis, is less than 3. In most experimental sites, NO_3^- has persistent behavior with other elements, such as Cl^- , F^- , SO_4^{2-} , TDS, alkalinity, hardness, and turbidity, while Ca, Mg, Fe²⁺ and PH parameters have anti-persistent behavior.

In almost all the experimental sites, SO_4^{2-} has less indication with the 0-mode value and the mean and median values show many differences and are not the same. The WQPs datasets are dispersive, as evidenced by the higher standard deviation. The skewness and kurtosis values of the different data series indicated that the series are symmetrical and platykurtic. SO_4^{2-} in most of the cases provides real random-walk flow and exhibits both persistent and anti-persistent behavior with turbidity, Cl⁻, TDS, and alkalinity parameters in comparing the different experimental sites.

4.4.3. Variation in Iron, Arsenic, Calcium and Magnesium

The average, median, and mode values of Fe^{2+} are nearly equal; hence, its distribution is observed under the standard or normal conditions. Dataset variation is observed to be Close to each other and this is evident by the standard deviation among the WQPs, which vary from 0.03 to 7 in all the experimental sites. The skewness and kurtosis curves are platykurtic for almost all the experimental sites because of the low variations, while the high skewness indicates that it is not symmetrical and also the kurtosis value is too large. Brownian time series (True random walk) behavior is observed with NO_3^- , F^- , and hardness parameters in the sample dataset having heavier outliers and Fe^{2+} . It displays persistent behavior with pH, Ca, and Mg parameters, but anti-persistent behavior with Cl^- , TDS, alkalinity, and SO_4^{2-} factors.

The datasets for the As^{++} element are found in a few experimental sites and observed with fewer amounts. The mean, median, and mode values of As^{++} are nearly equal; hence, its distribution is observed under the standard or normal conditions. Dataset variation is observed to be close to each other and this is evident by the standard deviation, which varies from 0.0 to 0.07 in the respective experimental sites. The skewness and kurtosis curves are platykurtic for almost all the experimental sites because of the low range variations.

Ca element variation in all the experimental sites is observed, meaning that the mean, median, and mode values are largely dispersed, and abnormalities are found in a few datasets. The high standard deviation suggests that the Ca levels are extremely discrete and generally low values are observed. The curve between skewness and kurtosis is not platykurtic and is positively skewed for almost all the sites. Ca normally shows persistent behavior with the distribution of other elements, such as Cl⁻, TDS, alkalinity, hardness, Mg, F⁻, SO₄²⁻, TDS, and turbidity, while the Fe²⁺ and pH parameters show anti-persistent behavior.

The Mg element is observed to show normal variation except in the Kumaun region (plane regions of the Uttarakhand). The dataset series for all the experimental sites are found to be dispersive; hence, the skewness and kurtosis curves do not reflect normal behavior. The average, median, and mode of Mg are under the standard range. The standard deviation considerably indicates that the Mg values are widely dispersed. The curve is platykurtic and it is positively skewed for all the experimental sites, except for the Tehri Garhwal. In almost all the experimental sites, the Mg element shows a Brownian time series (true random walk) with pH and alkalinity parameters. Mg exhibits persistent behavior with turbidity, Cl^- , NO_3^- , Ca, TDS, hardness, F^- , and SO_4^{2-} , while it exhibits anti-persistent activity with Fe²⁺ and As⁺⁺.

4.5. The Findings from the Results

In this study, the analysis of the variations and distribution of the WQPs in 13 experimental sites are discussed with spatial variation in the WQP remote sensing and GIS methods under different stress conditions.

In addition to this, the predictability analysis from the fractal approach in 13 experimental sites is also evaluated. Skewness and kurtosis curves are implemented on each parameter of 13 experimental sites. In all 13 experimental sites (districts), the non-platykurtic curve was observed in the water quality parameters to identify the comprehensive characteristics of the water quality indices. The curve between skewness and kurtosis is not symmetrical, as evidenced by the positive skewness among the series of data points. Kurtosis has shown large values, which indicate that the curve for datasets such as turbidity, Cl^- , Fe^{2+} , Ca, and SO_4^{2-} are platykurtic. For almost all the experimental sites, the NO_3^- , Ca, Mg, TDS, alkalinity, and hardness variation is in normal form as the high coefficient of variation among each one is observed. In most of the WQIs in totality, Brownian time-series behavior, while Ca, Mg, Fe^{2+} , and NO_3^- exhibit persistent behavior, and turbidity, Cl^- , TDS, alkalinity, hardness, and SO_4^{2-} exhibit anti-persistent behavior.

Turbidity, Cl⁻, NO₃⁻, F⁻, TDS, alkalinity and SO₄²⁻ have a persistent nature with Fe, Ca, Mg, and NO₃⁻ for almost all the experimental sites, and all have a stable relationship with the hardness of the water. When comparing Fe²⁺ with Cl⁻, TDS, alkalinity, and SO₄²⁻ parameters, an anti-persistent tendency can be observed. NO₃⁻ has an anti-persistent relationship with Ca, Mg, Fe²⁺, and pH, while with other factors, including turbidity, Cl⁻,

TDS, alkalinity, hardness, $SO_4^{2^-}$ has an anti-persistent performance. Mg has anti-persistent behavior with only Fe^{2+} , whereas Ca has anti-persistent activity with both Fe and pH parameters in almost all the sites. With Fe^{2+} , pH, Ca, and Mg, the parameters F^- , alkalinity, TDS, and SO_4^{2-} possess an anti-persistence nature. The various indices show a consistent pattern, indicating that the fluctuations in WQPs are within an adequate level of one another. The fractal and statistical analysis were combinedly found to be a better approach for calculating linkage among the water quality indices. The water samples were found to be safe to drink and in healthy condition in almost all the districts of the state Uttarakhand, except for the Haridwar district, where some increase in contaminants was observed.

The current proposed architecture enables us to identify the WQIs of every district based on the ground data from the 13 districts of the Uttarakhand state. The proposed architecture empowers us to confirm the water quality of every district. In addition to this, IoT and ML-based framework was implemented to carry out the statistical multivariate approaches using the RF model to find the correlation among the parameters. The broad execution of this architecture in the future scope allows for the administration to recognize which districts' ground water quality is worsening, as well as to envision water quality data in real time on a cloud server, enabling the administration to comply with the requirements to purify the water before supplying it to the people. The calibrating of the many sensors, as well as the availability of a few sensors with the technology for acquiring real-time data, are the suggested architecture's limitations.

5. Conclusions

Sustainable management and availability of water quality ensure the health of the human beings associated with that region. This can be achieved with the integration of big data, IoT and ML. In this study, a big data framework is proposed to find the correlation between WQPs of 13 districts of the Uttarakhand state. In addition to this, IoT and ML-based framework was implemented to carry out the statistical multivariate approaches using an RF model. The variation in WQP is analyzed using an RF model, and the dataset is segmented location wise and the mean, mode, standard deviation, median, kurtosis, and skewness of time series datasets are examined. The means of the parameters are adjusted with the coefficient of variation based on the standard values of each parameter. The water samples were found to be safe to drink and in healthy condition in almost all the districts of the state Uttarakhand, except for the Haridwar district, where some increase in contaminants was observed.

Author Contributions: Conceptualization, S.K., A.T. and S.V.A.; methodology, R.S., A.G. and H.G.M.; validation, D.A.; formal analysis, M.I. and I.D.N.; investigation, R.S. and A.G.; writing—original draft preparation, S.K., A.T. and S.V.A.; writing—review and editing, D.A., M.I. and I.D.N.; supervision, H.G.M. and D.A.; project administration, H.G.M. All authors have read and agreed to the published version of the manuscript.

Funding: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022TR140), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Data Availability Statement: Not applicable.

Acknowledgments: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022TR140), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Clean Water and Sanitation Projects | SDG 6—UN India. Available online: https://in.one.un.org/page/sustainable-development-goals/clean-water-sanitation-sdg-6/ (accessed on 28 May 2022).
- Simonetti, V.C.; Frascareli, D.; Gontijo, E.S.J.; Melo, D.S.; Friese, K.; Silva, D.C.C.; Rosa, A.H. Water quality indices as a tool for evaluating water quality and effects of land use in a tropical catchment. *Int. J. River Basin Manag.* 2021, 19, 157–168. [CrossRef]

- Singh, H.; Singh, D.; Singh, S.K.; Shukla, D.N. Assessment of river water quality and ecological diversity through multivariate statistical techniques, and earth observation dataset of rivers Ghaghara and Gandak, India. *Int. J. River Basin Manag.* 2017, 15, 347–360. [CrossRef]
- 4. MacKenzie, R. Clearing the Waters over Hygiene Debate. Nurs. Times 2000, 96, 24.
- 5. Parmar, K.S.; Bhardwaj, R. Statistical, time series, and fractal analysis of full stretch of river Yamuna (India) for water quality management. *Environ. Sci. Pollut. Res.* 2015, 22, 397–414. [CrossRef] [PubMed]
- Wahl, D.; Ness, B.; Wamsler, C. Implementing the urban food–water–energy nexus through urban laboratories: A systematic literature review. *Sustain. Sci.* 2021, 16, 663–676. [CrossRef]
- El-Gafy, I. Water-food-energy nexus index: Analysis of water-energy-food nexus of crop's production system applying the indicators approach. *Appl. Water Sci.* 2017, 7, 2857–2868. [CrossRef]
- Konapala, G.; Mishra, A. Quantifying climate and catchment control on hydrological drought in the continental United States. Water Resour. Res. 2020, 56, e2018WR024620. [CrossRef]
- 9. Sharma, D.; Kansal, A. Water quality analysis of River Yamuna using water quality index in the national capital territory, India (2000–2009). *Appl. Water Sci.* 2011, *1*, 147–157. [CrossRef]
- 10. Ai, L.; Shi, Z.H.; Yin, W.; Huang, X. Spatial and seasonal patterns in stream water contamination across mountainous watersheds: Linkage with landscape characteristics. *J. Hydrol.* **2015**, *523*, 398–408. [CrossRef]
- Sinha, D.K.; Rastogi, G.K.; Kumar, R.; Kumar, N. Correlation study among water quality parameters an approach to water quality management. J. Environ. Sci. Eng. 2009, 51, 111–114. [PubMed]
- Jabbar, F.K. Assessment and Prediction of Surface Water Vulnerability from Non-Point Source Pollution in Midwestern Watersheds. Ph.D. Thesis, Missouri University of Science and Technology, Rolla, MO, USA, 2019.
- 13. Alnahit, A.O.; Mishra, A.K.; Khan, A.A. Stream water quality prediction using boosted regression tree and random forest models. *Stoch. Environ. Res. Risk Assess.* 2022, 4. [CrossRef]
- 14. Shen, L.Q.; Amatulli, G.; Sethi, T.; Raymond, P.; Domisch, S. Estimating nitrogen and phosphorus concentrations in streams and rivers, within a machine learning framework. *Sci. Data* **2020**, *7*, 161. [CrossRef] [PubMed]
- Bisht, H.; Kotlia, B.S.; Kumar, K.; Taloor, A.K.; Arya, P.C.; Sah, S.K.; Agnihotri, V.; Tewari, M.; Upadhyay, R. Hydrogeochemical analysis and identification of solute sources in the meltwater of Chaturangi glacier, Garhwal Himalaya, India. *Appl. Water Sci.* 2022, 12, 29. [CrossRef]
- Gajbhiye, S.; Sharma, S.K.; Awasthi, M.K. Application of Principal Components Analysis for Interpretation and Grouping of Water Quality Parameters. *Int. J. Hybrid Inf. Technol.* 2015, *8*, 89–96. [CrossRef]
- Gad, A.; Ali, R.R.; El-zeiny, A. Negative impacts of man-made activities on water quality, Egypt. Natl. Authroity Remote Sens. Space Sci. 2011, 1–20.
- 18. Lei, C.; Wagner, P.; Fohrer, N. Influences of land use changes on the dynamics of water quantity and quality in the German lowland catchment of the Stör. *Hydrol. Earth Syst. Sci. Discuss.* **2022**, *26*, 2561–2582. [CrossRef]
- 19. Antolini, F.; Tate, E.; Dalzell, B.; Young, N.; Johnson, K.; Hawthorne, P.L. Flood Risk Reduction from Agricultural Best Management Practices. *J. Am. Water Resour. Assoc.* 2020, *56*, 161–179. [CrossRef]
- 20. Grazzini, F.; Craig, G.C.; Keil, C.; Antolini, G.; Pavan, V. Extreme precipitation events over northern Italy. Part I: A systematic Cl-assification with machine-learning techniques. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 69–85. [CrossRef]
- Rawat, K.S.; Singh, S.K.; Jacintha, T.G.A.; Nemčić-Jurec, J.; Tripathi, V.K. Appraisal of long term groundwater quality of peninsular India using water quality index and fractal dimension. J. Earth Syst. Sci. 2017, 126, 4. [CrossRef]
- Rawat, K.S.; Singh, S.K. Water Quality Indices and GIS-based evaluation of a decadal groundwater quality. *Geol. Ecol. Landsc.* 2018, 2, 240–255. [CrossRef]
- 23. Williams, Z.C.; Pelletier, J.D.; Meixner, T. Self-Affine Fractal Spatial and Temporal Variability of the San Pedro River, Southern Arizona. J. Geophys. Res. Earth Surf. 2019, 124, 1540–1558. [CrossRef]
- Prada, D.A.; Reátiga, A.; Correa, S.N.; Acevedo, A.; Mantilla, J.; Florville, T.; Nuñez, J. Fractal analysis in rural domestic wastewater quality under dissolved oxygen stability in wetlands planted with Chrysopogon zizanioides. *J. Phys. Conf. Ser.* 2019, 1386, 012136. [CrossRef]
- Wu, M.Y.; Xue, L.; Jin, W.B.; Xiong, Q.X.; Ai, T.C.; Li, B.L. Modelling the Linkage Between Landscape Metrics and Water Quality Indices of Hydrological Units in Sihu Basin, Hubei Province, China: An Allometric Model. *Procedia Environ. Sci.* 2012, 13, 2131–2145. [CrossRef]
- Lopez, A.L.; Haripriya, N.A.; Raveendran, K.; Baby, S.; Priya, C.V. Water quality prediction system using LSTM NN and IoT. In Proceedings of the 2021 IEEE International Power and Renewable Energy Conference (IPRECON), Kollam, India, 24–26 September 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6.
- Ahmed, U.; Mumtaz, R.; Anwar, H.; Shah, A.A.; Irfan, R. Efficient Water Quality Prediction Using Supervised. *Water* 2019, 11, 2210. [CrossRef]
- 28. Wu, D.; Wang, H.; Seidu, R. Smart data driven quality prediction for urban water source management. *Future Gener. Comput. Syst.* **2020**, *107*, 418–432. [CrossRef]
- 29. Park, S.; Jung, S.; Lee, H.; Kim, J.; Kim, J.H. Large-scale water quality prediction using federated sensing and learning: A case study with real-world sensing big-data. *Sensors* **2021**, *21*, 1462. [CrossRef]

- 30. Thai-Nghe, N.; Thanh-Hai, N.; Ngon, N.C. Deep learning approach for forecasting water quality in IoT systems. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 686–693. [CrossRef]
- 31. Emamgholizadeh, S.; Kashi, H.; Marofpoor, I.; Zalaghi, E. Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. *Int. J. Environ. Sci. Technol.* **2014**, *11*, 645–656. [CrossRef]
- Adamowski, J.; Chan, H.F. A wavelet neural network conjunction model for groundwater level forecasting. J. Hydrol. 2011, 407, 28–40. [CrossRef]
- 33. Najafzadeh, M.; Ghaemi, A. Prediction of the five-day biochemical oxygen demand and chemical oxygen demand in natural streams using machine learning methods. *Environ. Monit. Assess.* **2019**, *191*, 380. [CrossRef]
- 34. Zhou, Y. Real-time probabilistic forecasting of river water quality under data missing situation: Deep learning plus post-processing techniques. J. Hydrol. 2020, 589, 125164. [CrossRef]
- Chen, T.; Zhang, H.; Sun, C.; Li, H.; Gao, Y. Multivariate statistical approaches to identify the major factors governing groundwater quality. *Appl. Water Sci.* 2018, 8, 215. [CrossRef]
- 36. Rashidi, H.H.; Tran, N.K.; Betts, E.V.; Howell, L.P.; Green, R. Artificial intelligence and machine learning in pathology: The present landscape of supervised methods. *Acad. Pathol.* **2019**, *6*, 2374289519873088. [CrossRef]
- Rahimi, N.; Eassa, F.; Elrefaei, L. An ensemble machine learning technique for functional requirement classification. *Symmetry* 2020, 12, 1601. [CrossRef]
- Talukdar, S.; Singha, P.; Mahato, S.; Pal, S.; Liou, Y.-A.; Rahman, A. Land-use land-cover classification by machine learning classifiers for satellite observations—A review. *Remote Sens.* 2020, *12*, 1135. [CrossRef]
- Rani, D.; Gill, N.S.; Gulia, P.; Chatterjee, J.M. An Ensemble-Based Multiclass Classifier for Intrusion Detection Using Internet of Things. *Comput. Intell. Neurosci.* 2022, 2022, 1668676. [CrossRef]
- 40. Chatzis, S.P.; Siakoulis, V.; Petropoulos, A.; Stavroulakis, E.; Vlachogiannakis, N. Forecasting stock market crisis events using deep and statistical machine learning techniques. *Expert Syst. Appl.* **2018**, *112*, 353–371. [CrossRef]
- 41. Mandelbrot, B.B. Stochastic models for the Earth's relief, the shape and the fractal dimension of the coastlines, and the number-area rule for islands. *Proc. Natl. Acad. Sci. USA* **1975**, *72*, 3825–3828. [CrossRef]
- 42. *Manual on Aquifer Mapping;* Government of India, Ministry of Water Resources, Central Ground Water Board: Faridabad, India, 2012; 72p.
- 43. Schepers, H.E.; Van Beek, J.H.G.M.; Bassingthwaighte, J.B. Four methods to estimate the fractal dimension from self-affine signals (medical application). *IEEE Eng. Med. Biol. Mag.* **1992**, *11*, 57–64. [CrossRef]