Assessing The Impact of Land Cover On Groundwater Quality In a Smart City Using GIS And Machine Learning Algorithms

Rukhsar Anjum (✉ rukhsar.amu3@gmail.com)

  Aligarh Muslim University Faculty of Sciences   https://orcid.org/0000-0001-7488-5591

Sk Ajim Ali

  Aligarh Muslim University Faculty of Sciences

Mansoor Alam Siddiqui

  Aligarh Muslim University Faculty of Sciences

Abstract

The present study aimed to assess the impacts of land cover on groundwater quality by integrating physico-chemical data and satellite imageries. Initially, fourteen groundwater parameters of both pre-monsoon and post-monsoon were collected from nineteen sampled stations and water quality index (WQI) was calculated. Consequently, Google earth, Landsat-8, and Sentinel-2A imageries were considered for land cover mapping including the concentration of settlement and urban built-up, greenery coverage, and micro-level land use. Two machine learning models such as artificial neural network (ANN) and random forest (RF) were used for pixel-based classification and establish spatial relation between water quality and land cover. This study trained and tested the models for the whole study area as well as 500m buffers from each sampled station. The result of model's validation including mean absolute error (MAE), root mean squared error (RMSE), Kappa statistic (K), Overall accuracy of model (OAC), and receiver operating characteristic (ROC), indicated that random forest classifier has better performance than the artificial neural network. The results show that the testing dataset of pre-monsoon season has higher accuracy with MAE 0.343, RMSE 0.397, K-value 0.55, and ROC value 0.838 to envisage the impact of land cover on groundwater quality in comparison to post-monsoon season. The results also reveal that the classification accuracy is greater within 500m buffer areas in comparison to the whole study area with a close to 0 value of MAE and RMSE, and absolutely 1 value of K and ROC. Based on the above findings, the present study suggested to consider a large scale for determining the controlling factors of groundwater degradation.

1. Introduction

Water is the epitome of life. The decay of water both from various non-point and point sources has led to a difficult challenge for the civilians and planners (Bui et al., 2020). The major groundwater systems on the globe don't stay in dynamic equilibrium rather indicate not capable declining water table. As per the 2010 World Bank report, India secured its position to be a largest patron of groundwater with an average estimated use of 230 km³. Water is the rudimentary feature for agriculture, industry, or urban concretization regardless of domestic use. Developing nations frequently shifts from a fast pace of economic boom and each development strategy has a likelihood of producing a cynical impact on the environment. Pervasive ecological changes took place as an outcome of human pursuit (Shaker et al., 2010). Low-quality drinking water is a crucial factor for the developing world because it risks the entire ecosystem along with human health. Most of the Indian cities especially those situated in the central part of the nation surrounded by land from all sides depend mainly on groundwater as a reliable source of drinking water. Overpopulation leads to the huge demand of food supply rising stress on soil besides industrial bloom which becomes the reason of contamination of groundwater with pesticide and toxic fertilizers and industrial heavy metals making the water filthy. Monitoring of water quality is henceforth turned into a necessity for environmental and human health and sustainable water management (Zhang and Li, 2019). There are two main controlling factors of water pollution; natural process and anthropogenic process (Machiwal et al., 2018). Like pollution by using fertilizers in agriculture resulted in

nitrate contamination (Paradis et al., 2016), boreholes causing sea water penetration into the ground in coastal areas (Ferguson and Gleeson, 2012), pollution by toxin penetration into groundwater (Molina et al., 2009; Wu et al., 2017; Shaker et al., 2010; Machiwal et al., 2018; Wang et al., 2017), Contamination of geo-thermal fluids on shallow water (Iskandar et al., 2012), contamination of groundwater from the fractured bed rock (Bondu et al., 2017), and pollution by oxidation etc. Hence it is much needed to evaluate and monitor the groundwater.

Evaluation of groundwater quality is a challenging task. Over a decade, many techniques have been developed and applied for the assessment of water quality out of them water quality index (WQI) is an extensively used method (Chaurasia et al., 2018; Kanga et al., 2020; Kawo and Karuppannan, 2018; Tyagi et al., 2013; Yisa et al., 2012). WQI is a mathematical tool that signifies the values in number and minimizes the complexity of the data set by providing results in single classifying values (Akhtar et al., 2021). Lumb et al. (2006) had used WQI for the purpose of evaluating the water quality of freshwater against Canadian council of ministers of the environment (CCME) guidelines. Water quality parameters were used in many studies (Yan et al., 2016; Venkatraman et al., 2016; Rabeiy, 2018; Kim et al., 2005; Wagh et al., 2019). The water quality index calculated from the selected water quality parameters is a non-dimensional test (Rezaie-Balf et al., 2020; Rufino et al., 2019).

In this study, indices allow a class appraisal of two different seasons from the same sampling stations since the results vary between pre-monsoon and post-monsoon. The variables used to obtain WQI include pH, Alkalinity, Dissolved oxygen (DO), Total Dissolved Solid (TDS), Total suspended solids (TSS), Chloride, Total Hardness, Electrical Conductivity, Calcium, Magnesium, Turbidity, Sulphate, Bicarbonate, and Sodium. There are various methods to calculate WQI, among them rank weight method was considered in the present study. WQI shows the result in generalized form, and various land covers were fitted with WQI to know the relation between them which makes the process complicated. To conquer these difficulties, now researchers have started using WQI combining with Artificial Intelligence (AI) to obtain much more precise data of water quality and their relation with various factors (Yaseenet al., 2018; Bedi et al., 2020; Wang et al., 2017; El Bilali et al., 2021). AI-based modeling excludes sub-index calculations and gives way more reliable output. The use of machine learning algorithms has been increased ten-fold over a decade due to their capacity for quick results, big data handling of various categories at a single time, and precise results. The output result entirely depends on the input data and methodology.

Land cover change and its impact on biosphere elements of the water cycle is the core objective of this study. The breakthrough point for regional sustainable development has one most anticipated aspect i.e. land cover change. Since the accelerating procedure of urbanization and its unplanned sprawl of urban territory, the land cover has changed drastically, resulting in the public attention to impact environmental resources more specifically the quality and quantity of hydrology (Dutta et al., 2018; Szewrański et al., 2018). Land cover change always tends to speed up the process of erosion which affects all links of the hydrological cycle which accelerates the process of the non-point source pollution in the watershed. Hence monitoring the land cover change and water quality has turned into a need (Liu et al., 2006; Li.,

1996). The land cover pattern in the process of urbanization where anthropogenic activity is intense, significantly impact aquifer water quality (Rao and Chaudhary., 2019; He and Wu., 2019; Robertson et al., 2017). In order to investigate the connection among groundwater and land cover, a three-layer system was followed after He and Wu (2019) which including quantification of groundwater quality via prescribed approach, determining the land cover pattern of the selected study area, and the relationship of groundwater quality to land cover patterns.

For the first step, water quality index can productively quantify groundwater characteristics. For the second step, the land cover data is generally acquired from two sources i.e. from satellite imageries (He and Wu., 2019) or the local survey department (Cain et al., 1989). In this study, land cover patterns of the study area were collected from Landsat-8 and Sentinel-2A which were further verified in the Google earth engine. The third step determines the range of influence (He and Wu., 2019). He and Wu (2019) proposed a curved streamline searching model (CS-SLM). In this study, the artificial neural network and random forest classifier were used to look into the connection between water quality and land cover patterns.

ANN is an extensively used AI algorithm. This algorithm has been created keeping in mind the function of human neural where it uses many hidden layers to process the data then the final output comes out. However, it faces some issues such as comparatively weaker prediction power especially when the testing data is out of the range of training data when datasets are less in quantity (Khosravi et al., 2019). Henceforth random forest classifier can be considered in this regard, which is lacking in hidden layers while processing the data as a result transparent decision tree-based AI algorithm present the result, and the output is very accurate compared to other AI-based algorithms and consumes very little time (Ali et al., 2021; Bui et al., 2020).

The present study aimed to show the groundwater quality of a smart city (Aligarh, Uttar Pradesh, India) using WQI and identify the potential controlling factor in degrading the water quality. More specifically, look into the hydro geochemistry of groundwater, understand the land cover pattern, and relate the water quality with land cover pattern in the city. For the same, a total of fourteen water parameters from each of nineteen sampled stations were selected and water quality index was estimated, then three controlling land cover factors including the concentration of settlement and urban built-up, greenery coverage, and micro-level land use were considered and finally two machine learning algorithms were utilized to identify the impacts of land cover on water quality and assess the classification accuracy.

2. Study Area

Aligarh city lies at 27°88'N latitude and 78°08'E longitude, 132 km south-east of the Capital of India (Fig. 1). Aligarh city is the administrative headquarter of the Aligarh district. The city is divided into 70 wards and four zones. The total area of the city is 36 Km^2. The city is famous for its lock industry. The total population of the city is 874,408 as per the last census report, 2011.

Geomorphologically this region comes under the older alluvial plain region, and the soil type is western upland soil series. The area lies under Ganga-Yamuna alluvial plain region. Kali Nadi is the only river that passes nearby the city. There is a sharp identification of old city and new city. The older part of the city is mainly congested and haphazard in terms of settlement and population density. Aligarh comes under a flat plain region apart from few low-lying areas. The city has depression in its central part among the entire district, which gives the city saucer pan topography. As a result, there are always troubles related to waterlogging and drainage facilities.

The city has multiple land cover types, including urban built-up and settlement, vegetation, barren land, swampland, open space, and water bodies. But alone the urban built-up and settlement consists about 71% out of total areal coverage, whereas water bodies have least coverage (less than 1% out of the total area of the city). So, lower areal extension and higher population with maximum built-up coverage is an indication of poor urban quality, which may largely affect the environmental components like groundwater. The city has three layers aquifers out of which only one layer is suitable for drinking and domestic usage. The other two falls under the category of brackish water (Anwar and Aggarwal, 2014). Except for the groundwater, there is no existing alternative source of water supply available in the city. The groundwater in the city is alkaline in nature and based on TDS measure nearly 30% of the total groundwater available is not suitable for drinking purposes, but permissible only after purifying (Wasim et al., 2014).

Therefore, the present study aimed to assess the relationship between surface land cover and groundwater quality in the city in order to understand the future requisite to improve the groundwater quality throughout the city.

3. Methods And Materials

In order to fulfill the objective, various steps were followed to assess the impacts of land cover on the groundwater quality of the study area. So, initially, a land cover map of the study area was prepared. Then the sample of groundwater collection locations was selected looking towards the land cover map. Based on the general preview from the Google earth engine, it was observed that the city has haphazard growth in all directions without proper management and measures which resulted in the stress on groundwater is becoming prominent each passing day. Hence, the land cover was taken for assessing the water quality in the study area. In order to prepare the land cover of the study area, three parameters including the concentration of settlement and urban built-up (CSUB), greenery coverage (GC), and micro-level land use (MLU) were selected. Out of which, the first two parameters were prepared from Sentinel-2A satellite imagery for the year 2020 and the last parameter was derived using pixel supervised classification. For the supervised classification, a total of nine determining parameters namely agricultural land, barren land, settlement, water bodies, swamp, vegetation, agricultural fallow land, vegetation fallow land, and recreational area were considered. Keeping in mind the micro-level land use, groundwater samples were collected depending on the location of the public water supply system and their surrounding land use pattern. Most of the samples were selected so that almost the entire city can

be covered, but few areas are dominated by personal submersible; hence, trials from those private water sources have been collected. The sampling points were marked in the 'x' 'y' GPS coordinate system so that results can be put for the data visualization. Then WQI was calculated to identify the water quality of the area. To evaluate the correlation between land cover and groundwater quality, a 500m buffer was created around each sampled station to delineate water quality. For assessing the accuracy, machine learning techniques such as artificial neural network (ANN) and random forest (RF) were used in the case of both the entire city and within the buffer area. The flow chart presents the details of the methodology applied in the present study (Fig. 2).

3.1 Groundwater sample collection

For groundwater sample collection, clean 1L plastic bottles have been used, and before collecting the sample, those bottles were rinsed off with distilled water. The water samples were collected from a running water source so that no stagnant water gets compiled into the bottle, because in stagnant or storage water in the tank for example can change its bacteriological and physio-chemical properties, that's why running water from the hand pump and submersible were considered. *Total 1* shows that a total 19 groundwater sample stations were selected for collecting sample water which is mainly from bore wells that supply water in the entire city, and very few samples were collected from a submersible, where people mainly prefer their water supply. Both pre-monsoon and post-monsoon data were collected from these sample points in May 2019 and October 2019. Few parameters like DO, EC, pH, and temperature were tested on the spot and the rest of the tests were done at the laboratory. The parameters such as TDS, TSS, TH, Chloride, Calcium, Magnesium, Alkalinity, Turbidity, and COD were tested at the Environmental Engineering laboratory, Aligarh Muslim University, whereas the other parameters were tested at the Analysis laboratory, Agra. All tests were performed as per the BIS norms (Table 2).

Table 1
Groundwater sampled points and their codes

Sl. No.	Place	Coordinates	Sl. No.	Place	Coordinates
1	Double tankey coloney, Shahjamal	27°52'44.29"N 78° 3'15.02"E	11	Dodhpur chauraha	27°54'21.88"N 78° 5'7.65"E
2	Avas Vikas colony	27°53'50.23"N 78° 3'58.24"E	12	Near Student Union hall	27°54'25.85"N 78° 4'31.76"E
3	Pratibha colony	27°53'37.48"N 78° 3'21.93"E	13	Press colony	27°53'47.74"N 78° 4'47.09"E
4	Malkhan Singh Hospital	27°53'21.13"N 78° 4'15.21"E	14	Duda colony, Sootmill	27°54'21.88"N 78° 3'12.68"E
5	Jumma Masjid, Upparkot	27°52'46.42"N 78° 3'59.99"E	15	B.S.J hall	27°55'8.96"N 78° 4'5.87"E
6	Dhobi Ghat, Jamalpur	27°55'39.64"N 78° 4'44.57"E	16	J.N.M.C	27°55'2.30"N 78° 5'20.47"E
7	Shiwalik Ganga Phase IV	27°53'32.93"N 78° 6'1.06"E	17	Mulla para	27°51'46.80"N 78° 3'37.29"E
8	Vaishno Royal apartment, Ramghat road/ Surendranagar	27°52'54.12"N 78° 5'11.61"E	18	Kali Deh	27°52'6.69"N 78° 5'2.22"E
9	Gandhi Eye Hospital	27°53'21.64"N 78° 5'0.27"E	19	Patel Nagar	27°52'19.27"N 78° 4'18.01"E
10	Ambedkar Park, Jiwangarh	27°54'38.99"N 78° 5'40.53"E			

Table 2
Details about the test method of selected parameters

Sl. No	Parameters	Test methods
1	pH	pH meter
2	Turbidity	Neflometer
3	EC	Conductivity meter
4	TDS	Filtration method
5	TSS	Evaporation method
6	Alkalinity	Indicator method
7	DO	Wrinkle's method
8	Sodium	IS:3025(Part 45)-1993
9	Potassium	IS:3025(Part 45)-1994
10	Sulphate	IS:3025(Part 24)-1986
11	Carbonate	IS:3025(Part 51)-2001
12	Bicarbonate	IS:3025(Part 51)-2001
13	Magnesium	Flame AAS
14	Chloride	Spectrophotometric

3.2 Relevant data used and their sources

This study was based on both primary and secondary data. Primary data include groundwater from 19 sampled locations in two seasons, followed by laboratory tests and experiments. Apart from that, various secondary datasets were also collected for different aspects. For the purpose of applying machine learning techniques, the land cover map was prepared from Google earth and verified using Landsat 8 data. The concentration of settlement and urban built-up, and greenery coverage of the city was prepared using Sentinel-2A imageries. The details about the data collection and use have been shown in the following table (Table 3).

Table 3
Relevant data and their sources

Layers	Sources	Format
Groundwater parameters	Collection from sampled stations by authors	Numeric data
Micro-level land use	Google earth, Landsat 8, 2021 (USGS earth explorer),	Vector data from Google earth, Raster data from Landsat 8
The concentration of settlement and urban built-up	Sentinel-2A (https://apps.sentinel-hub.com/)	Raster
Greenery coverage	Sentinel-2A (https://apps.sentinel-hub.com/)	Raster

3.3 Selection of chemical parameters

Based on various literature, it has been observed that pre-monsoon and post-monsoon seasons have sufficient influence on the balance of chemical parameters on the groundwater quality (Rao., 2017). Both dry season and wet season have a direct influence on water stress and contamination along with various chemicals present on rock or infiltration (Haines et al., 2006). Needless to say that an ample amount of geochemical compounds are present in the groundwater, out of which few Physico-chemical parameters which may create many health problems were considered for this study, such as pH, Alkalinity, Dissolved Oxygen (DO), Total Dissolved Solids (TDS), Total Suspended Solids (TSS), Total Hardness (TH), Chloride (Cl^-), Electrical Conductivity (EC), Calcium (Ca), Magnesium (Mg), Turbidity, Sulphate (SO_4^{2-}), Bicarbonate (HCO_3^-) and Sodium (Na).

pH is one of the important parameters which determines whether the water is acidic or alkaline. Extreme exposure of pH can give rises to eye irritation and mucous skin (Ali and Ahmad., 2020). *TDS* refers to the presence of organic and inorganic substances in water. This brings certain necessary matters with it. Negative TDS is unhealthy but beyond the permissible limit can intrigue some health problems (Burton and Cornhill., 1977; Schroeder., 1960; Schroeder., 1966). *TSS* may contain siltation and various decomposed matter in water. As per BIS and WHO guidelines, the presence of TSS is not permissible in drinking water. The presence of TSS might bring out various unwanted microorganisms which can trigger Nausea, Diarrhea, headaches, etc. (Kjelland et al., 2015). *Turbidity* has not any specific health implications but it can act as a breeding ground for microbial development. High turbidity makes it difficult to wash away the impurities from water by chlorine (Baghvand et al., 2010). *Chloride* does not have an active effect on human health but those who already have some previous health issues related to sodium chloride metabolism may get affected. Prolonged ingestion of chloride for ages might bring concerns about human health. Alkalinity in drinking water is also an important parameter where crossing the permissible limit of *Al* might bring out skin irritation and gastrointestinal diseases along with vomiting and nausea besides their health impact (Wynn et al., 2010). The presence of *DO* in water is a sign of good

water quality, anyhow beyond the permissible limit turns the water into a breeding ground for the bacteriological phenomenon (Bhatia et al., 2015). *Total hardness* is an important element in water. It brings benefits with it but beyond the permissible limit, makes the water soapy, pulses take more time to get boiled than usual and water pipes become narrow by its siltation. The health impact of *TH* is controversial among researchers, none the less it has some share in Kidney stones, cardiovascular diseases, digestion, etc. (Sengupta, P., 2013). *Calcium* is one of the important elements needed for our body. A good amount of Calcium intake is needed on our consumption and lack of Calcium can cause Hypocalcaemia, muscle cramp, dry skin, etc. (Pravina et al., 2013). *Magnesium* is another important geo-chemical present in water. The deficiency of magnesium can cause Hypomagnesemia, hypertension, Osteoporosis, headaches, etc. (Watson et al., 2012; Al Alawi et al., 2018). *EC* determines the presence of minerals in the water. High EC means a high presence of minerals that might ruin the taste of water (Meride and Ayenew., 2016). *Sulphate* has a huge laxative impact on pregnant women and infants are more prone to diarrhea when the water has an undesirable amount of sulphate. It degrades the taste of water as well (Heizer et al., 1997). *Bicarbonate* is a dependent ion in groundwater; bicarbonate joins with sodium most of the time. Exceeding the permissible limit of bicarbonate might result in hypernatraemia, rebound alkalosis. *Sodium* presence in water is a healthy sign. It can mitigate kidney damage, headache, hypertension, etc. but an overdose of it can inference into heart diseases, blood pressure etc. The spatial distribution of selected groundwater parameters in both seasons has been shown in the following figure (Fig. 3 and 4).

Land cover plays an important role in the sustainable growth and development of the city. the entire Aligarh district has only 1.83% area under forest cover in 2019 assessment as per the forest survey of India, whereas the Aligarh city has about 11.76% area under vegetation cover. Various studies showed that land use land cover has an impact on hydrometeorology, surface runoff, aquifer level, and water quality (Baidya et al., 2002; Sajikumar et al., 2015; Schilling et al., 2008). Hence, it can be said that it is an intense need to observe the relationship between groundwater quality and land cover pattern.

In order to study groundwater quality, several possible approaches along with their classification and assessment are significant (Su et al., 2019; Tian and Wu., 2019). In case of the present study, the impact of land cover on groundwater quality consists of both supervised and unsupervised classification. The supervised classification includes information related to groundwater quality and quantifies its standard, whereas the unsupervised classification is used to present the natural pattern of water quality. The supervised classification includes WQI (Chaurasia et al., 2018), ANN (Gupta et al., 2019), RF classifier (Grbčić et al., 2020), and unsupervised classification includes a self-organized map (Liu et al., 2018).

3.4 Calculation of WQI

Groundwater is considered the most potent source of freshwater which is present in the least polluted form because it is least interfered with by humans (Wagh et.al, 2019). It is estimated that around one-third world's population depends on groundwater, among them Aligarh city is a peculiar example where no other alternate source is available for drinking water. Here public's only source of domestic and drinking water is groundwater. Hence, checking its status is essential. WQI is a method which is widely

used for over a decade for the assessment of water quality (Gupta et al., 2017; Sadat-Noori et al., 2014; Şener et al., 2017; Mishra and Patel., 2003; Avvannavar and Shrihari 2008; Vasanthavigar et al., 2010). There are several advantages of WQI including it is relatively easy to apply, it can easily add new variables, it has a precise range of results, etc. (Walsh and Wheeler., 2013). So, there is no doubt about its reliability and simplified method. Here in this study, the rank weight method of WQI was used. For the assessment of the parameters, the Bureau of Indian standard was considered as the standard parameters, and those standards which are not described in BIS, World Health Organization standard has been selected for them. Among them, Total dissolved solids (TSS) are not mentioned on both BIS and WHO guidelines; hence NEMA (Kenya) standard has been selected for that. A total of 14 geochemical parameters were taken into consideration. A relative weight from 1 to 4 out of five has been assigned based on their importance and previous literature reviews (Batabyal and Chakraborty., 2015; Vasanthavigar et al., 2010). Turbidity has given a minimum value of 1 since it is a pretty less important parameter and a maximum value of 4 has been given to pH, TDS, and Sulphate. After that relative weight has been given to each parameter by the following equation (Eq. 1):

$$W_i = \frac{Rn}{\sum_{i=1}^{n} Rn}$$

1

Where relative weight = W_i, rank of each parameter = R, number of parameter $=n$

After placing calculated relative weight (W_i), a quality ranking scale (Q_i) was assigned for every single parameter by dividing its concentration in each water sample by its standard as per the Bureau of Indian Standard (BIS, 1991), and then the result was multiplied by 100 (*Eq. 2*):

$$Q_i = \frac{Ci}{Si} \times 100 \text{ (2)}$$

Where quality rating = Q_i, Concentration of each chemical parameters in water sample = Ci, Indian drinking water standard for the mentioned parameters $= Si$

The next step is to determine the sub-index (Sl_i) of each Physico-chemical parameter of water samples by using the following equation (Eq. 3):

$$Sl_i = Wi * Q_i$$

3

where sub-index of ith parameter = Sl_i, the relative weight of ith parameter = Wi

Finally, the WQI was obtained by the following equation (Eq. 4):

$$WQI = \sum Sli$$

4

After calculating the water quality index, the value was divided into five categories ranging from excellent to unsuitable for drinking.

3.5 Mapping of land cover

The land cover considered in the present study consists of three different raster layers. These are the concentration of settlement and urban built-up, greenery coverage, and micro-level land use. The data of micro-level land use were obtained from the real-time Google Earth engine for the year 2020 and the supervised image classification was used for mapping. Here, a total of nine pre-identified signatures were considered for classification purposes. The signature categories includes settlement, water body, vegetation, barren land, recreational area, fallow land, agricultural land, swamp, and open space. On the other hand, the concentration of settlement and urban built-up (CSUB) and greenery coverage (GC) were extracted from Sentinel-2A imageries (https://apps.sentinel-hub.com/eo-browser). The CSUB was derived from the band combination of 12 (Shortwave infrared-1), 11 (Shortwave infrared-2), and 4 (Red). This composite is generally used to visualize settlement and urban built-up areas more clearly. On the other hand, the GC was derived using the following equation (Eq. 5):

$$GS = \frac{(B8 - B4)}{(B8 + B4)}$$

5

Where $B8$ is the Near Infrared band, and $B4$ is the Red band. The value range of GC is -1 to 1. A negative value corresponds to water, values close to 0 correspond to barren areas of rock, sand, or snow. Positive values represent vegetation (nearly 0.2 to 0.4), whereas a high value indicates tropical forest.

3.6 Models used

The selected groundwater parameters were mapped first, followed by mapping of water quality index, and then land cover maps were generated as mentioned above. The WQI and land cover of the whole city and 500m buffer regions around the sampled stations for both pre-monsoon and post-monsoon were mapped. To know the relationship between WQI and land cover, machine learning algorithms were applied. Based on the model's classification accuracy, the spatial correlation between WQI and land cover was determined in this study.

3.6.1 Random forest (RF)

Random forest is a tree rooted classifier, an improvised version of the bagging-based method coined by Breiman (2001). It is the combination of bagging tree and multivariate data which turns it into a way better version tool for pattern recognition of multivariate and large scale data tree creation (Liaw and Wiener, 2002). It produces a huge number of trees by using bootstrapping, it provides numerous metrics which in turn helps in interpretation (Prasad et al., 2006; Baudron et al., 2013). Although RF has its cross-validation (Breiman, 2001; Efron and Tibshirani, 1997), it is recommended to check the sensitivity of the

model so that maximum accuracy can be obtained (Matthew, 2011). Random forest works in four basic steps (Ali et al., 2021)

1. Selection of sample feature 'k' randomly from the total sample 'm', where 'k' < 'm',

2. Calculation of the node tree 'd' by applying the befitting split point amidst the selected features 'k',
3. Again applying the best split point 'd' into daughter nodes tree 'dn' then
4. Reiteration of the above-mentioned steps till the Ith number of nodes tree is brought off.

The RF classifier determines a paramount number of nodes to its output class (Bonissone et al., 2010). Therefore, for an input data 'x' has computed its output 'y' from the highest ensemble which is exhibited in the equation (Eq. 6):

$$Y = (x) = max \left[\sum_{k} I(t) \right]$$

6

Where *I(t)* is an indicator function marked off as:

$$I(t) \begin{cases} 1, t = 'YES' \\ 0, t = 'NO' \end{cases}$$

In the case of the present study, the good water quality and poor water quality locations were indicated by 'YES' and 'NO', respectively.

3.6.2 Artificial neural network (ANN)

The artificial neural network is a technique that follows the role of a neuron as in the human brain (Diamantopoulou et al., 2005; Wu et al., 2014), which is faster and reliable. Like neurons, the artificial neural network works in different layers where non-linear data is processed and transmitted from one layer to another (Isiyaka et al., 2019). Here the whole process works in three layers, that is input layer, the hidden layer, and the output layer (Diamantopoulou et al., 2005). For this process both training and testing data are required, where training data includes weights of the variables and adjust them with the help of the iterative method (Isiyaka et al., 2019). In this study, a multi-layer perceptron feed-forward method of ANN was used with a back propagation algorithm to identify the most polluting contributor in groundwater quality, and it model every single individual's percentage of contribution to the pollution.

In this regard, two input combination models were produced so that the most statistically significant API with high accuracy can be achieved (Alagha et al., 2014; Sarkar and Pandey, 2015). Here, ANN has been applied to both the whole city area and within the buffer region. Total data has been categorized into training (80%) and testing (20%) datasets. To test the Multi-layer perceptron ANN model, route means square error and coefficient of determinants was used by following the equations (Eqs. 7 and 8):

$$R^2 = 1 - \frac{\sum (x_{i-y_i})^2}{\sum y_i^2 - \frac{\sum y_i^2}{n}}$$

7

$$RMSC = \sqrt{\frac{1}{n} \sum_{i=1}^{i=n} (x_i - y_i)}$$

8

Where x_i = the observed data, {y}_{i} = the predicted data, and N = the total number of observations.

3.6.3 Inventory data preparation and data resampling

Inventory data preparation is cardinal in order to execute machine learning techniques and also for the validation of the models. To pertain random forest and artificial neural network, a dataset of various water quality determining factors has been taken into consideration like micro-level land use of the study area, concentration of settlement and urban built-up, greenery coverage, and WQI of the area. Using the GIS, 100 points in the category of good water quality points together with 100 poor water quality points were digitized within the study area. The total good and poor water quality sample points were subsequently categorized into training (80%) and testing (20%) data. Thus, for the purpose of the whole city, a total number of 80 good and 80 poor water quality point data were included in training data, meanwhile, the rest of the 20 good and 20 poor water quality data were allocated for the validation of the models. Analogously, for the demarcation of water quality within buffer region total number of 140 point data was used out of a total 200. From the 140 point data, 80% training and 20% testing data were utilized where training data sets counted as a total number of 114 points and testing data counts as 28 points (Table 4).

Table 4
Details of training and testing dataset

Pre-monsoon				Post-monsoon			
Within whole city limit		Within buffer limit		Within whole city limit		Within buffer limit	
Training	Testing	Training	Testing	Training	Testing	Training	Testing
160	40	114	28	160	40	114	28

3.6.4 Feature selection

With reference to the machine learning algorithm selection of features is a supreme step of model building (Booker and Snelder, 2012). Feature selection is required because of evaluating the relevance of selected factors for model building. In the present study to determine the factor's importance, the ReliefF ranking evaluator was used. ReliefF ranking is the probabilistic approach used for data classification

which monitors conditional reliance and discriminative power of identified factors (Belgiu and Drăguţ, 2016). Increased rank expresses more significance and '0' expresses no or irrelevant factor for modeling.

3.6.5 Performance evaluation of the models

3.6.5.1 Receiver operating characteristics (ROC) curve

The ROC curve is a widely accepted method that has been widely used in various fields including geospatial analysis for the validation of the model in which ROC curve exhibits the trade-off between specificity and sensitivity (Chen et al., 2020). The far off the curve from the ROC space, the more accurate the test. The ROC doesn't hang on the class distribution. One common approach to calculate the ROC uses the area under the curve (AUC) for numeric evaluation where specificity and sensitivity are being placed on 'x' and 'y' axis consequently. A greater value of AUC indicates higher accuracy of the result.

3.6.5.2 Arithmetic measures

Besides the ROC curve, various arithmetic measures were also carried out for the prediction of the accuracy of the model. For this purpose, root mean square error (RMSE), Kappa index (K), sensitivity, specificity, and overall accuracy (OAC) were reckoned to measure the accuracy of random forest and ANN models used in the present analysis. The below mentioned Eqs. (9) - (14) were useful for the statistical measures (Khosravi et al., 2019; Ali et al., 2020). Mean Absolute error (MAE) is the average between actual observation and prediction over the test sample of absolute difference. Route Mean Square Error (RMSE) is the unadulterated measure of fit where minimum values of RMSE designate better fit. Kappa index helps out to measure inter reliability among the variables.

$$MAE =\frac{1}{n}\sum _{i=1}^{i=n}{ X}_{ei} - {X}_{oi}$$
9

$$RMSC=\sqrt{\frac{1}{n} \sum _{i=1}^{i=n}\left({x}_{i}- {y}_{i}\right)^2}$$
10

$$\left.K =\left(\frac{{W}_{c}- {W}_{exp}}{1 - {W}_{exp}}\right.\right)$$
11

$$Sensitivity = \left.\left(\frac{TP}{TP+FN}\right.\right)$$
12

$$Specificity = \left.\left(\frac{TN}{TN+FP}\right.\right)$$
13

$$OAC =\left.\left(\frac{TP+TN}{TP+TN+FP+FN}\right.\right)$$
14

Where TP is the true positive, TN is the true negative and both TP and TN represent the number of pixels correctly classified; FP is the false positive, FN is the false negative and both FP and FN represent numbers of pixels that are incorrectly classified, ${W}_{c}$ is the number of pixels that are correctly classified as good water quality and poor water quality, ${W}_{exp}$ is the expected agreement's value, ${X}_{ei}$ is the predicted value, ${X}_{oi}$ is the observed value, and n is the number of datasets.

4. Results

4.1 Water quality assessment parameters and their proportion

The prevailing condition of Aligarh city reflects that most of the samples score poor water quality rank as per the standard of BIS or WHO. In pre-monsoon season, pH value exceeds the permissible limit at 5 stations, every sample exceeds the permissible limit of DO, 10 stations exceeds TDS permissible limit, TH relies on between the permissible limit, where 6 stations surpassed the permissible limit of magnesium, EC, and Alkalinity exceeds its permissible limit at 18 sampling stations, 1 station crossed Sulphate limit, 16 stations crossed sodium and magnesium permissible limit where Calcium, Chloride, Turbidity, Bicarbonate are within the limit. In the post-monsoon season, some values increased while some have deceased. Only one station crossed the permissible limit of pH, again DO cross its permissible limit at every station, 9 stations surpassed TDS permissible limit, 14 stations outstripped TSS limit. Magnesium range is exceeded by 6 stations, EC and Alkalinity permissible limit has been overshadowed by 18 stations, 7 stations transcended Sodium permissible limit and TH, Calcium, Chloride, Turbidity, Sulphate, Bicarbonate is within the permissible range.

It has been observed that the areas nearby to the infiltration points from where a regular seepage water penetration is taking place, the areas where very little surface water penetration is happening or deep dumping of industrial hazardous wastes via borehole into the underground where it finds its way to the aquifer is happening, has comparatively poor water quality throughout the year and the areas where big empty spaces are tracked down or waterlogging level is higher in monsoon season, has pretty better rainwater harvesting capacity hence the water quality is better in post-monsoon season.

4.2 Land cover of Aligarh city

In the present study, the land cover was defined by three parameters, i.e. concentration of settlement and urban built-up, greenery coverage, and micro-level land use (Fig. 5). As per the image classification, it is shown that about 80% of the total city area is covered by the urban settlement and built-up area, whereas 10% areas are covered by vegetation, 7% barren land, and the rest 3% covered by other land-use types including agricultural land, open space, recreational area, water bodies, and swamp. In terms of greenery coverage, only the campus area (i.e. Aligarh Muslim University) has covered by vegetation, and other areas have the least vegetation cover. In order to train and test the machine learning model, the land cover types within 500m buffer areas were mapped (Fig. 6). The details of pixel-based correlation between land cover of the study area and water quality index have been discussed in section 4.6, as follows.

4.3 Results of water quality index (WQI)

WQI was calculated and evaluated based on the weightage given looking towards their importance in water quality assessment. It is generally recommended to examine at least four parameters to assess WQI, and in this study, a total of fourteen parameters were taken into consideration out of which, ${SO}_{4}$,

pH and TDS were given a maximum weightage of 0.1052, whereas Alkalinity, Total suspended solids, chloride, bicarbonate, and sodium were given the weightage of 0.0789. The minimum weight of 0.0526 was allotted for DO, total hardness, calcium, magnesium, and EC. All 19 sampling stations were categorized into five categories based on WQI in both pre-monsoon and post-monsoon seasons. The enumerated values of WQI range from 78.54 to 209.88 and 78.39 to 179.98 during pre-monsoon and post-monsoon, respectively (Table 5).

Table 5
Calculated water quality index for the study area

Water quality index		Types of water	Remarks
Pre-monsoon	Post-monsoon		(For drinking purposes)
78.54 - 104.80	78.39 - 98.71	Very good	Highly preferable
104.80 - 131.07	98.71 - 119.033	Good	Preferable
131.07 - 157.34	119.03 - 139.35	Moderate	Considerable
157.34 - 183.61	139.35 - 159.66	Bad	Not considerable
183.61 - 209.88	159.66 - 179.98	Very bad	Undesirable

The WQI value for each sample station has been exhibited in Table 6. For the pre-monsoon season, 26.31% of the sampled waste constitutes excellent water. Similarly, 52.63% indicates good quality, 15.78% indicates poor quality and 5.26% indicates both very poor quality and undesirable for drinking purposes, respectively. Among each category, the best water quality was observed on Soot Mill (78.54), Vaishno Royal apartment (107.22), Ambedkar Park (132.53), Patel Nagar (163.26), and Kali Deh (209.88), respectively. Likewise, pre-monsoon season; the meantime of post-monsoon signifies that 26.31% sampled as excellent water quality, followed by 36.84% good quality, 21.05% moderate quality, 5.26% poor quality, and 10.52% very poor and thus undesirable for drinking purposes, respectively. During the post-monsoon, the best water quality was found at Malkhan Singh Hospital (78.39), Dodhpur chauraha (98.82), Gandhi eye hospital (120.85), Mulla Para (144.947), and again Kali deh (179.98), respectively.

Table 6
Water quality index of each sampled station

Area Name	WQI	
	Pre-Monsoon	Post-monsoon
Avas vikas colony	127.0781	125.7346
Pratibha colony	163.2662	97.341
Malkahn singh	97.235	78.3975
Jama masjid	138.5862	161.0143
Dhobi ghat, jamalpur	120.8948	116.3239
Shiwalik ganga phase iv	126.7235	93.5071
Vaishno royal apartment, ranghat road	107.2266	100.7975
Gandhi eye hospital	104.9942	120.8509
Ambedkar park	132.534	100.7651
Dodhpur chaurha	110.6856	98.827
Near student union hall	128.2212	124.0769
Press colony	111.3021	118.4534
Soot mill	78.5401	110.1755
BSJ hall	114.5635	131.4861
JNMC	96.4625	107.7026
Mulla para	153.3599	144.9474
Kali deh	209.8892	179.9869
Patel nagar	111.1949	94.9312
Double tanky	87.8219	90.2271

When we focus on particular land cover of the study area and WQI of that place, it was found that during the pre-monsoon season, poor water quality was observed at Jama masjid, Ambedkar park and Mulla Para that might be because of Jama masjid in the city has maximum elevation, and at that level where percolation of freshwater does not occur and that area comes under the industrial hub of Lock industry where some of the previous cases reported that hazardous wastes are secretly or sometimes openly being dumped into the boreholes that no longer in use. Ambedkar Park is that region where the elevation is lowest in the entire city where sanitation and hygiene is a daily life issue, hence over-exploitation of groundwater along with the release of direct discharge into groundwater can be a cause and the area Mulla Para which is situated just beside the city solid waste department (A2Z solid waste department),

where stagnant drain water along with a heap of garbage succumb the entire regions water quality. The very poor water quality was monitored at Pratibha colony which has a dense population along with barren land which is of no use in the pre-monsoon season. Kali deh is the only region where the physical property of the water was found damaged. The area is near a lot of water bodies which has stagnant filthy water for the entire year along with a pumping station that occupies a big area where sewerage line water is being pumped out for the entire year. In case of the post-monsoon season, the overall water quality of the city was noticed improved, where maximum WQI was observed as 179.98 to 209.88 during the pre-monsoon season. But based on their arithmetic mean category, post-monsoon season shows more poor water quality than pre-monsoon season, where 73.68% sampled water quality was observed good in pre-monsoon season rather than 63.15% during post-monsoon season. Similarly, poor water quality level was also increased in post-monsoon season (36.84%) compared to pre-monsoon season (26.31%). This phenomenon may be due to the percolation of impure water from the surface to the underground (Fig. 7).

Despite having overall better water quality compared to the pre-monsoon season, poor water quality was noticed at Gandhi eye hospital, near student union hall, B.S.J. Hall, and Avas Vikas colony in the post-monsoon season. Except for Avas Vikas colony, all three paces come under the civil line sector which is a shallow laying area hence rainwater percolation rate is higher in these areas which trigger the water along with impure wastewater being fused up the groundwater quality.

4.4 Results of feature selection

In this present study, the training and testing dataset were prepared by applying the locations of good and bad water quality, and a multi-values to point extraction tool was executed in which individual points occupy a specific pixel value of every determining factor. In the training dataset, the binary value such as 1 and 0, stipulating good and bad water quality, respectively. The ReliefF evaluator was utilized in order to calculate the rank of different factors responsible for deteriorating water quality. In the pre-monsoon season, all three determining factors were applied both in the whole city area and within the buffer region and the same process goes for the post-monsoon season as well. In the pre-monsoon season, the result of ReliefF ranker shows that concentration of settlement and urban built-up ranks 0.0139, greenery coverage ranks 0.0123, whereas micro-level land use has 0 ranks during the model training within the whole study area and output rank 0 for all these three factors during the model testing. Using the training dataset within the 500m buffer region shows a rank of 0.0340 for the concentration of settlement and urban built-up, 0.0179 for greenery coverage, and again close to 0 (i.e. 0.0046) for micro-level land use, whereas the testing dataset offers a rank 0.0456 for the concentration of settlement and urban built-up, 0.0119 for greenery coverage and absolute 0 for micro-level land use. Simultaneously, in post-monsoon using the training data, the result of ReliefF evaluator shows a rank of 0.0263 for the concentration of settlement and urban built-up, 0.0522 for greenery coverage, and 0 for micro-level land use. On the other hand, using the testing dataset, the result represents a rank of 0, 0.0295, 0.0281 for the concentration of settlement and urban built-up, greenery coverage, and micro-level land use, respectively. Similarly, for buffer region, the shows that the concentration of settlement and urban built-up, greenery coverage, and

micro-level land use holds a rank of 0.0453, 0.0371, and 0 during model training and 0, 0.0089, 0.0218 during model testing, respectively.

4.5 Models applied and their comparison

In the present study, two machine learning algorithms naming artificial neural network (ANN) and random forest (RF) classifier were applied to assess the impact of land cover on groundwater quality. The statistical presumption was used to superimpose good and bad water quality contributing factors with the help of existing WQI data which records the water quality of the city for accessing and validating the results. Initially, three land cover factors were evaluated and all three selected factors were integrated to prepare the final result of each of the parameter's contribution to deteriorating water quality. The dispatch of these two models for the prediction of groundwater quality of Aligarh city was evaluated and juxtaposed by using ROC and various arithmetic measures including MAE, RMSE, and Kappa statistics were also used for determining the suitability of the model for water quality assessment. Table 7 and Table 8 show the results of applied models in the case of whole study area and within the buffer region for both seasons.

Table 7
Validation of the applied models within whole study area

	Training		Testing		Training		Testing	
	Pre-monsoon		Pre-monsoon		Post-monsoon		Post-monsoon	
	RF	ANN	RF	ANN	RF	ANN	RF	ANN
Mean absolute error	0.372	0.469	0.343	0.430	0.307	0.395	0.334	0.372
Root mean squared error	0.423	0.486	0.397	0.459	0.384	0.443	0.387	0.427
Kappa statistic	0.418	0.142	0.555	0.250	0.511	0.345	0.574	0.456
Accuracy (OAC)	0.693	0.593	0.810	0.600	0.756	0.675	0.787	0.775
ROC	0.800	0.594	0.838	0.665	0.863	0.741	0.862	0.755

Table 8
Validation of the applied models within 500m buffers

	Training		Testing		Training		Testing	
	Pre-monsoon		Pre-monsoon		Post-monsoon		Post-monsoon	
	RF	ANN	RF	ANN	RF	ANN	RF	ANN
Mean absolute error	0.0008	0.007	0.007	0.012	0.302	0.382	0.270	0.353
Root mean squared error	0.002	0.007	0.016	0.012	0.379	0.436	0.345	0.427
Kappa statistic	1	1	1	1	0.437	0.281	0.623	0.461
Accuracy (OAC)	0.995	0.991	0.982	0.964	0.774	0.696	0.803	0.785
ROC	1.000	1.000	1.000	1.000	0.850	0.728	0.909	0.758

Within the whole study area during the pre-monsoon season, the result of training data shows that the ROC is higher for RF (0.800) than the ANN (0.594). Among arithmetic measures, the MAE of RF (0.3721) is less than ANN (0.4694). In terms of RMSE, the RF has a lower value (0.4236) than ANN (0.4865), Kappa statistics (K) and OAC were estimated, where K is less in RF (0.418) than ANN (0.1422), and the OAC is higher in RF (0.6937) than the ANN (0.59375). The result of testing data reveals that the ROC curve is greater in the RF classifier (0.838) than the ANN (0.665), the RMSE has less value in the RF classifier (0.3973) compared to ANN (0.4595), whereas MAE is less in RF (0.3432) than ANN (0.4301). The K is greater in RF (0.55) than the ANN (0.25) and the OAC is also higher in RF (0.8) as compared to ANN (0.6).

Within the whole study area during the post-monsoon season, the result of training data shows that ROC is slightly higher in RF (0.863) than ANN (0.741). In arithmetic measures, the RMSE is less in RF (0.3848) compared to the ANN (0.443), and MAE is less in RF (0.3073) than ANN (0.3953). The K is higher in RF (0.5113) than ANN (0.3451) and OAC is higher as well in RF (0.75625) than ANN (0.675). Using the testing data, the result shows that the ROC is greater in RF (0.862) than ANN (0.755), the RMSE is less in RF (0.3871) than ANN (0.427) and MAE is a little less in RF (0.3341) than ANN (0.3728). The K is higher in RF (0.5745) than ANN (0.4565) and OAC is also a little higher in RF (0.7875) than ANN (0.775) model.

Alongside building the model in the whole study area, models were applied within the 500m buffer for both pre-monsoon and post-monsoon seasons. Using the training data in pre-monsoon, the model accuracy reveals that the ROC is 1.00 for both RF and ANN models. The RMSE is very less in RF (0.0028) and ANN (0.0075) as well. The MAE is negligible higher in RF (0.008) than ANN (0.0075), the K is showing the absolute result (1.000) for both RF and ANN and OAC is very close to 1 for both the models, wherein OAC of RF is 0.99 and ANN is 0.99. Similar results were found using the testing dataset in the case of ROC and K for both models, i.e., absolute 1.00. The RMSE is a little higher in RF 0.0163 than ANN 0.0124, and MAE was found almost the same differences where RF (0.0163) is slightly higher than ANN (0.123). The OAC is slightly higher in RF (0.9821) than ANN (0.9642).

Using the training data in post-monsoon, it was found that the ROC is higher in RF (0.7743) than ANN (0.6964). The RMSE is less in RF (0.3796) than ANN (0.4366), and MAE is lesser in RF (0.3029) than ANN (0.382). The K is greater in RF (0.4378) than ANN (0.2815) and OAC is also greater in RF (0.7743) than ANN (0.6964). The result found using testing data was almost similar, the ROC value of RF is very close to 1 (i.e. 0.909) whereas, in ANN, the ROC is a less value (0.758). The RMSE of RF is lesser (0.3456) than ANN (0.4278) and the MAE is less in RF (0.2702) in comparison to ANN (0.3531). The K is higher in RF (0.6237) than ANN (0.4615) and the OAC is higher in RF (0.80357) than ANN (0.78571).

From the above-mentioned discussion, it is found the random forest (RF) has a more accurate result in comparison to the artificial neural network (ANN) in all cases including within the whole study area, within the 500m buffer region, during pre-monsoon and post-monsoon season. The result also reveals that within 500m buffer of sampled stations have more accurate result in terms of the ROC values and K values very close to 1.00, very less RMSE and very high OAC value.

4.6 Assessing impact of land cover on groundwater quality

Scrutiny of nature and the pattern of land cover is directly associated with groundwater quality, henceforth, it is needed to understand the current ecological problems properly. In this present scenario, the water quality of Aligarh city exhibits pretty inferior conditions as per WQI analysis. For assessing its impact, the water quality results were tied up with the remotely sensed data and machine learning models. From the analysis of the previous section, the result shows that both models are accurate and hence, can be accepted. The pixel-based classification of land cover and water quality index using RF and ANN offer spatial assimilation and relation which may help in understanding the impact of land cover on water quality.

In the context of whole city, the result shows that there is a strong correlation between the land cover and water quality. Here it was found that during the post-monsoon season with training data, the highest correlation exists between the concentration of settlement and urban built-up (0.0340) with WQI, followed by greenery coverage (0.0179) and micro-level land use (0.00469), whereas the testing dataset shows that highest relation taken place between the concentration of settlement and urban built-up (0.0456) and WQI, followed by greenery coverage (0.0119) and micro-level land was no impact on water quality. Consequently, the training data of pre-monsoon shows a positive relation between greenery coverage (0.0123) and concentration of settlement and urban built-up (0.0139) with WQI, but micro-level land use has failed to establish any relation with it. The testing dataset also reveals that a negative correlation of all three parameters with WQI during pre-monsoon; hence it is evident that this parameter doesn't bother the water quality of the area. Except for the pre-monsoon testing dataset, the other three datasets of both seasons showed a positive correlation with the two existing factors. The concentration of settlement and urban built-up is the main factor that gets highly correlation and hence affects the water quality, followed by greenery coverage (Fig. 8).

The area within the 500m buffer shows a different result compared to the whole study area assessment. The post-monsoon training dataset within the buffer area shows a high value of the concentration of

settlement and urban built-up, subsequently greenery coverage. Unlike these two parameters, micro-level land use offers a negative value hence shows no effect on water quality. The results of the post-monsoon testing dataset show the high out-turn of greenery coverage. Here the concentration of settlement and urban built-up shows the negative output. The output of the pre-monsoon training dataset indicates a greater relation of concentration of settlement and urban built-up (0.05223) than greenery coverage (0.02634), where micro-level land use represents no relation. In the case of the pre-monsoon testing dataset, the concentration of settlement and urban built-up, greenery coverage, and micro-level land use shows almost exact positive relation (Fig. 9).

5. Discussions

Water is the essence of life, mixing of impurities makes a lot of things worse hence proper detection of the rate of deteriorated water qualities and their controlling factors needs to be done for the prevention of further contamination and preservation. Hence the applications of precise techniques are getting called for to reduce the casualty caused by humans. Initiating suitable models is the prime motive for the damage assessment and their management. With the help of machine learning algorithms, good and bad water quality areas can be easily recognized and precautions can be taken over occurring activities. Presently, diverse technique and model has been presented to set out the observation of water quality of Aligarh city both in macro and micro level, where pixel-based calculations have been performed. At present, profuse work is being done in the field of groundwater along with model building (Chen et al., 2020; Belgiu and Drăguţ., 2016). Viewed in this way, the present study associated and weight up different techniques namely WQI along with machine learning algorithms such as random forest (RF) classifier and artificial neural network (ANN) to assess the land cover induced water quality in Aligarh city.

On this regard, a mathematically valid model with a good prediction tool is needed. Recently, ensemble techniques are globally appreciated for the assessment of the hydro-geochemistry and their determining factors (Chen et al., 2020; Belgiu and Drăguţ., 2016; Tung and Yaseen., 2020; Najahet al., 2021; Kumar et al., 2020). Although a lot of works have been done in previous years, a maximum of them focused on only Physico-chemical analysis. Thus, this study is the very first attempt to evaluate satellite data along with Physico-chemical parameters of water quality to find a correlation between land cover and groundwater quality. Also, the present study focused on applying machine learning algorithms for spatial classification and establishing their relation using different statistical measures.

At the present scenario, machine learning has become the basic need for the researchers hence a lot of researchers are attracted to model building or prediction (Hameed et al., 2017; Wang et al., 2017; Najafzadeh et al., 2019; Ali et al., 2021; Alagha et al., 2014). The artificial neural network got its recognition for its well predictive performance where it works like human neural and turns big data into precise results (Isiyaka et al., 2019; Kadam et al., 2019; Diamantopoulou et al., 2005; Isiyaka et al., 2019). The random forest known as a classification tree is well recognized for its best prediction tendency (Tyralis et al., 2019; Baudron et al., 2013; Belgiu and Drăguţ., 2016; Bui et al., 2020; Chen et al., 2020). So, these two models were applied in identifying impacts of land cover on WQI in Aligarh city.

Based on the output results of these two models, It was found that that random forest classifier obtained maximum accuracy (pre-monsoon training = 69.38%, pre-monsoon testing = 80%, post-monsoon training =75.62%, post-monsoon testing = 78.75%) in comparison to artificial neural network model (pre-monsoon training = 59.38%, pre-monsoon testing = 60%, post-monsoon training = 67.5%, post-monsoon testing = 77.5%). The similar results were also assessed from buffer region, where the accuracy of random forest (pre-monsoon training = 99.55%, pre-monsoon testing = 98.21%, post-monsoon training = 77.43%, post-monsoon testing = 80.35%) showed higher compared to artificial neural network (pre-monsoon training = 99.11%, pre-monsoon testing = 96.43%, post-monsoon training = 69.64%, post-monsoon testing = 78.57%). Hence, based on the comparative analysis, it is clear that the random forest classifier provides maximum accurate results than ANN.

The zeal of this study was to detect changes in water quality with varying land cover characteristics over the study area with the help of these two applied models. The output of the models remarked that both of them can be considered for the detection of land factors on water quality, and developing precaution strategies to reduce the impacts. From the perspective of comparative analysis, the Random forest classifier turns into an ideal model since it has outsmarted the ANN model in terms of ROC-AUC, MAE, RMSE, OAC, and Kappa index.

6. Conclusion

Aligarh comes under the category of class-I cities in India and more specifically it comes under Delhi NCR (National Capital Region), hence the population stress is proportionately increasing day by day. Apart from that, the industrial history of Aligarh with Lock and brass smithy is well recognized along with many more new sectors of industries which make the management and supply of water even more stressful. In this paper, the Physico-chemical properties of groundwater were analyzed and Water Quality Index (WQI) was computed for the study area to assess the role of land cover on water quality, and the results were validated using Artificial Neural Network and Random Forest Classifier. The present assessment reveals that (i) as per the BIS norms, most of the samples were marginalized under the poor water quality category. However, the WQI shows that most of the city comes under the good water quality category but the areas near to the sewer station or solid waste management like Mulla Para or kali deh have comparatively poor water quality than other areas. On overall comparison, pre-monsoon season outsmart the post-monsoon season in terms of good water quality; (ii) based on the ReliefF rank evaluator, the concentration of settlement and urban built-up is the most important factor responsible for the deteriorating water quality in the whole city, Followed by greenery coverage, i.e. the water quality of green coverage areas were found to be good in comparison to non-vegetated areas. The results within the 500m buffer areas also showed that greenery coverage is the most determining factor followed by concentration of settlement and urban built-up; (iii) the machine learning-based analysis presented the role of greenery coverage and concentration of settlement and urban built-up in groundwater quality with higher classification accuracy (higher K value and OAC) and lower errors (lower value MAE and RMSE).

After the comparative analysis of classification results of the applied models, it was evidenced that the areas where stagnant water stays for the maximum time alongside the places where very low chances of percolation exist have comparatively low WQI than the other areas in the city. As per the analysis of these models, it is very vigilant that the concentration of settlement is a significant controlling factor in determining groundwater quality. Hence, it can be determined that these Physico-chemical data along with suitable models in this study may be advantageous for the planning of the city and other areas with similar geo-environmental conditions.

Declarations

Author's Contributions Rukhsar Anjum and Sk Ajim Ali prepared data, developed the methodology, analyzed, and wrote the original draft regarding the impact of land cover on groundwater quality. Mansoor Alam Siddiqui critically reviewed and approved the final manuscript.

Funding No fund was received from any sources

Availability of data and materials: The data that support the findings of this study are available from the corresponding author [Sk Ajim Ali, skajimali@myamu.ac.in/ skajimali.saa@gmail.com], upon reasonable request.

Compliance with ethical standards

Ethical Approval The present study ensures that objectivity and transparency are followed in this research along with acknowledged principles of ethical and professional behaviour. The present research confirms that:

Competing interests: The authors declare that they have no conflict of interest.

Research involving Human Participants and/or Animals: Human Participants or Animals were not engaged or involved in the present research.

Therefore, for this study, compliance with ethical standards is not applicable

Consent to Participate Not applicable.

Consent to Publish Not applicable.

References

1. Akhtar N, Ishak MIS, Ahmad MI, Umar K, Md Yusuff MS, Anees MT et al (2021) Modification of the Water Quality Index (WQI) Process for Simple Calculation Using the Multi-Criteria Decision-Making (MCDM) Method: A Review. Water 13(7):905

2. Al Alawi AM, Majoni SW, Falhammar H (2018) Magnesium and human health: perspectives and research directions. International journal of endocrinology, 2018

3. Alagha JS, Said MAM, Mogheir Y (2014) Modeling of nitrate concentration in groundwater using artificial intelligence approach—a case study of Gaza coastal aquifer. Environ Monit Assess 186(1):35–45

4. Ali SA, Ahmad A (2020) Suitability analysis for municipal landfill site selection using fuzzy analytic hierarchy process and geospatial technique. Environ Earth Sci 79:1–27

5. Ali SA, Parvin F, Vojteková J, Costache R, Linh NTT, Pham QB et al (2021) GIS-based landslide susceptibility modeling: A comparison between fuzzy multi-criteria and machine learning algorithms. Geosci Front 12(2):857–876

6. Anwar KM, Aggarwal V (2014) Analysis of Groundwater Quality of Aligarh City,(India): Using Water Quality Index. Current World Environment 9(3):851

7. Avvannavar SM, Shrihari S (2008) Evaluation of water quality index for drinking purposes for river Netravathi, Mangalore, South India. Environ Monit Assess 143(1):279–290

8. Baghvand A, Zand AD, Mehrdadi N, Karbassi A (2010) Optimizing coagulation process for low to high turbidity waters using aluminum and iron salts. Am J Environ Sci 6(5):442–448

9. Baidya Roy S, Avissar R (2002) Impact of land use/land cover change on regional hydrometeorology in Amazonia. Journal of Geophysical Research: Atmospheres 107(D20):LBA–4

10. Batabyal AK, Chakraborty S (2015) Hydrogeochemistry and water quality index in the assessment of groundwater quality for drinking uses. Water Environ Res 87(7):607–617

11. Baudron P, Alonso-Sarría F, García-Aróstegui JL, Cánovas-García F, Martínez-Vicente D, Moreno-Brotóns J (2013) Identifying the origin of groundwater samples in a multi-layer aquifer system with Random Forest classification. J Hydrol 499:303–315

12. Bedi S, Samal A, Ray C, Snow D (2020) Comparative evaluation of machine learning models for groundwater quality assessment. Environ Monit Assess 192(12):1–23

13. Belgiu M, Drăguţ L (2016) Random forest in remote sensing: A review of applications and future directions. ISPRS journal of photogrammetry and remote sensing 114:24–31

14. Bhatia S, Sharma K, Dahiya R, Bera T (2015) Modern applications of plant biotechnology in pharmaceutical sciences. Academic Press

15. Bondu R, Cloutier V, Rosa E, Benzaazoua M (2017) Mobility and speciation of geogenic arsenic in bedrock groundwater from the Canadian Shield in western Quebec, Canada. Sci Total Environ 574:509–519

16. Bonissone P, Cadenas JM, Garrido MC, Díaz-Valladares RA (2010) A fuzzy random forest. Int J Approximate Reasoning 51(7):729–747

17. Booker DJ, Snelder TH (2012) Comparing methods for estimating flow duration curves at ungauged sites. J Hydrol 434:78–94

18. Breiman L (2001) Random forests. Mach Learn 45(1):5–32

19. Bui DT, Khosravi K, Tiefenbacher J, Nguyen H, Kazakis N (2020) Improving prediction of water quality indices using novel hybrid machine-learning algorithms. Sci Total Environ 721:137612

20. Burton AC, Cornhill JF (1977) Correlation of cancer death rates with altitude and with the quality of water supply of the 100 largest cities in the United States. Journal of Toxicology and Environmental Health, Part A Current Issues 3(3):465–478

21. Cain D, Helsel DR, Ragone SE (1989) Preliminary evaluations of regional ground-water quality in relation to land use. Groundwater 27(2):230–244

22. Chaurasia AK, Pandey HK, Tiwari SK, Prakash R, Pandey P, Ram A (2018) Groundwater quality assessment using water quality index (WQI) in parts of Varanasi district, Uttar Pradesh, India. J Geol Soc India 92(1):76–82

23. Chen W, Li Y, Xue W, Shahabi H, Li S, Hong H (2020) at al. Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree, and random forest methods. Science of The Total Environment, 701, 134979

24. Diamantopoulou MJ, Papamichail DM, Antonopoulos VZ (2005) The use of a neural network technique for the prediction of water quality parameters. Oper Res Int Journal 5(1):115–125

25. Dutta S, Dwivedi A, Kumar MS (2018) Use of water quality index and multivariate statistical techniques for the assessment of spatial variations in water quality of a small river. Environ Monit Assess 190(12):1–17

26. Efron B, Tibshirani R (1997) Improvements on cross-validation: the 632+ bootstrap method. J Am Stat Assoc 92(438):548–560

27. El Bilali A, Taleb A, Brouziyne Y (2021) Groundwater quality forecasting using machine learning algorithms for irrigation purposes. Agric Water Manage 245:106625

28. Ferguson G, Gleeson T (2012) Vulnerability of coastal aquifers to groundwater use and climate change. Nat Clim Change 2:342–345

29. Gupta R, Singh AN, Singhal A (2019) Application of ANN for water quality index. International Journal of Machine Learning and Computing 9(5):688–693

30. Haines A, Kovats RS, Campbell-Lendrum D, Corvalán C (2006) Climate change and human health: impacts, vulnerability, and mitigation. The Lancet 367(9528):2101–2109

31. Hameed M, Sharqi SS, Yaseen ZM, Afan HA, Hussain A, Elshafie A (2017) Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in tropical region, Malaysia. Neural Comput Appl 28(1):893–905

32. He S, Wu J (2019) Relationships of groundwater quality and associated health risks with land use/land cover patterns: a case study in a loess area, northwest China. Human and Ecological Risk Assessment: An International Journal 25(1–2):354–373

33. Heizer WD, Sandler RS, Seal E Jr, Murray SC, Busby MG, Schliebe BG, Pusek SN (1997) Intestinal effects of sulfate in drinking water on normal human subjects. Dig Dis Sci 42(5):1055–1061. https://doi.org/10.1023/a:1018801522760

34. Isiyaka HA, Mustapha A, Juahir H, Phil-Eze P (2019) Water quality modelling using artificial neural network and multivariate statistical techniques. Modeling Earth Systems and Environment 5(2):583–593

35. Iskandar I, Koike K, Sendjaja P (2012) Identifying groundwater arsenic contamination mechanisms in relation to arsenic concentrations in water and host rocks. Environ Earth Sci 65:2015–2026

36. Kadam AK, Wagh VM, Muley AA, Umrikar BN, Sankhua RN (2019) Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India. Modeling Earth Systems and Environment 5(3):951–962

37. Kanga IS, Naimi M, Chikhaoui M (2020) Groundwater quality assessment using water quality index and geographic information system based in Sebou River Basin in the North-West region of Morocco. International Journal of Energy and Water Resources 4(4):347–355

38. Kawo NS, Karuppannan S (2018) Groundwater quality assessment using water quality index and GIS technique in Modjo River Basin, central Ethiopia. J Afr Earth Sc 147:300–311

39. Khosravi K, Shahabi H, Pham BT, Adamowski J, Shirzadi A, Pradhan B et al (2019) A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. J Hydrol 573:311–323

40. Kim AG, Cardone CR (2005) Scatterscore: a reconnaissance method to evaluate changes in water quality. Environ Monit Assess 111(1):277–295

41. Kjelland ME, Woodley CM, Swannack TM, Smith DL (2015) A review of the potential effects of suspended sediment on fishes: potential dredging-related physiological, behavioral, and transgenerational implications. Environment Systems and Decisions 35(3):334–350

42. Kumar S, Rajesh V, Khan N (2020) Evaluation of groundwater quality in Ramanathapuram district, using water quality index (WQI).Modeling Earth Systems and Environment,1–11

43. Li X (1996) A review of the international researches on land use/land cover change, vol 51. ACTA GEOGRAPHICA SINICA-CHINESE EDITION-, pp 558–565

44. Liaw A, Wiener M (2002) Classification and regression by randomforest. R news 2(3):18–22

45. Liu J, Shen Z, Chen L (2018) Assessing how spatial variations of land use pattern affect water quality across a typical urbanized watershed in Beijing, China. Landscape and Urban Planning 176:51–63

46. Liu RM, Yang ZF, Ding XW, Shen ZY, Wu X, Liu F (2006) Effect of land use/cover change on pollution load of non-point source in upper reach of Yangtze River Basin. Huan jing ke xue=. Huanjing Kexue 27(12):2407–2414

47. Lumb A, Halliwell D, Sharma T (2006) Application of CCME Water Quality Index to monitor water quality: A case study of the Mackenzie River basin, Canada. Environ Monit Assess 113(1):411–429

48. Machiwal D, Cloutier V, Güler C, Kazakis N (2018) A review of GIS-integrated statistical techniques for groundwater quality evaluation and protection. Environ Earth Sci 77(19):1–30

49. Matthew W (2011) Bias of the Random Forest out-of-bag (OOB) error for certain input parameters. Open Journal of Statistics, 2011

50. Meride Y, Ayenew B (2016) Drinking water quality assessment and its effects on residents health in Wondo genet campus, Ethiopia. Environmental Systems Research 5(1):1–7

51. Mishra PC, Patel RK (2003) Study of the pollution load in the drinking water of Rairangpur, a small tribal dominated town of North Orissa. Aquatic Ecosystems, p 379

52. Molina M, Aburto FN, Calderan RL, Cazanga M, Escudey M (2009) Trace element composition of selected fertilizers used in Chile: phosphorus fertilizers as a source of long-term soil contamination. Soil Sediment Contam 18:497–511

53. Najafzadeh M, Ghaemi A, Emamgholizadeh S (2019) Prediction of water quality parameters using evolutionary computing-based formulations. Int J Environ Sci Technol 16(10):6377–6396

54. Najah A, Teo FY, Chow MF, Huang YF, Latif SD, Abdullah S, El-Shafie A (2021) Surface water quality status and prediction during movement control operation order under COVID-19 pandemic: Case studies in Malaysia.International Journal of Environmental Science and Technology,1–10

55. Paradis D, Vigneault H, Lefebvre R, Savard MM, Ballard J-M, Qian B (2016) Groundwater nitrate concentration evolution under climate change and agricultural adaptation scenarios: Prince Edward Island, Canada. Earth Syst Dyn 7(1):183–202

56. Prasad AM, Iverson LR, Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9(2):181–199

57. Pravina P, Sayaji D, Avinash M (2013) Calcium and its role in human body. International Journal of Research in Pharmaceutical and Biomedical Sciences 4(2):659–668

58. Rabeiy RE (2018) Assessment and modeling of groundwater quality using WQI and GIS in Upper Egypt area. Environ Sci Pollut Res 25(31):30808–30817

59. Rao NS (2017) Controlling factors of fluoride in groundwater in a part of South India. Arab J Geosci 10(23):1–15

60. Rao NS, Chaudhary M (2019) Hydrogeochemical processes regulating the spatial distribution of groundwater contamination, using pollution index of groundwater (PIG) and hierarchical cluster analysis (HCA): a case study. Groundwater for Sustainable Development 9:100238

61. Rezaie-Balf M, Attar NF, Mohammadzadeh A, Murti MA, Ahmed AN, Fai CM, El-Shafie A (2020) Physicochemical parameters data assimilation for efficient improvement of water quality index prediction: Comparative assessment of a noise suppression hybridization approach. J Clean Prod 271:122576

62. Robertson WM, Böhlke JK, Sharp JM Jr (2017) Response of deep groundwater to land use change in desert basins of the Trans-Pecos region, Texas, USA: Effects on infiltration, recharge, and nitrogen fluxes. Hydrol Process 31(13):2349–2364

63. Rufino F, Busico G, Cuoco E, Darrah TH, Tedesco D (2019) Evaluating the suitability of urban groundwater resources for drinking water and irrigation purposes: an integrated approach in the

Agro-Aversano area of Southern Italy. Environ. Monit. Assess 191., 768. https://doi.org/10.1007/s10661-019-7978-y

64. Sadat-Noori SM, Ebrahimi K, Liaghat AM (2014) Groundwater quality assessment using the Water Quality Index and GIS in Saveh-Nobaran aquifer. Iran Environmental Earth Sciences 71(9):3827–3843

65. Sajikumar N, Remya RS (2015) Impact of land cover and land use change on runoff characteristics. J Environ Manage 161:460–468

66. Sarkar A, Pandey P (2015) River water quality modelling using artificial neural network technique. Aquatic procedia 4:1070–1077

67. Schilling KE, Jha MK, Zhang YK, Gassman PW, Wolter CF (2008) Impact of land use and land cover change on the water balance of a large agricultural watershed: Historical effects and future directions. Water Resources Research, 44(7)

68. Schroeder HA (1966) Municipal drinking water and cardiovascular death rates. JAMA 195(2):81–85. https://doi.org/10.1001/jama.195.2.81

69. Şener Ş, Şener E, Davraz A (2017) Evaluation of water quality using water quality index (WQI) method and GIS in Aksu River (SW-Turkey). Sci Total Environ 584:131–144

70. Sengupta P (2013) Potential health impacts of hard water. International journal of preventive medicine 4(8):866

71. Shaker R, Tofan L, Bucur M, Costache S, Sava D, Ehlinger T (2010) Land coverand landscape as predictors of groundwater contamination: a neural-network modelling approach applied to Dobrogea. Romania Journal of environmental protection and ecology 11(1):337–348

72. Su F, Wu J, He S (2019) Set pair analysis-Markov chain model for groundwater quality assessment and prediction: A case study of Xi'an city, China. Human and Ecological Risk Assessment: An International Journal 25(1–2):158–175

73. Szewrański S, Chruściński J, Van Hoof J, Kazak JK, Świąder M, Tokarczyk-Dorociak K, Żmuda R (2018) A location intelligence system for the assessment of pluvial flooding risk and the identification of storm water pollutant sources from roads in suburbanised areas. Water 10(6):746

74. Tian R, Wu J (2019) Groundwater quality appraisal by improved set pair analysis with game theory weightage and health risk estimation of contaminants for Xuecha drinking water source in a loess area in Northwest China. Human and Ecological Risk Assessment: An International Journal 25(1–2):132–157

75. Tung TM, Yaseen ZM (2020) A survey on river water quality modelling using artificial intelligence models: 2000–2020. J Hydrol 585:124670

76. Tyagi S, Sharma B, Singh P, Dobhal R (2013) Water quality assessment in terms of water quality index. American Journal of water resources 1(3):34–38

77. Tyralis H, Papacharalampous G, Langousis A (2019) A brief review of random forests for water scientists and practitioners and their recent history in water resources. Water 11(5):910

78. Vasanthavigar M, Srinivasamoorthy K, Vijayaragavan K, Ganthi RR, Chidambaram S, Anandhan P et al (2010) Application of water quality index for groundwater quality assessment: Thirumanimuttar sub-basin, Tamilnadu, India. Environ Monit Assess 171(1):595–609

79. Wagh VM, Mukate SV, Panaskar DB, Muley AA, Sahu UL (2019) Study of groundwater hydrochemistry and drinking suitability through Water Quality Index (WQI) modelling in Kadava river basin, India. SN Applied Sciences 1(10):1–16

80. Walsh PJ, Wheeler WJ (2013) Water quality indices and benefit-cost analysis. Journal of Benefit-Cost Analysis 4(1):81–105

81. Wang X, Zhang F, Ding J (2017) Evaluation of water quality based on a machine learning algorithm and water quality index for the Ebinur Lake Watershed. China Scientific reports 7(1):1–18

82. Wasim SM, Khurshid S, Shah ZA, Raghuvanshi D (2014) Groundwater Quality in Parts of Central Ganga Basin, Aligarh City, Uttar Pradesh, India. Proc Indian Natn Sci Acad (Vol 80(1):123–142

83. Watson RR, Preedy VR, Zibadi S (eds) (2012) Magnesium in human health and disease. Springer

84. Wu W, Dandy GC, Maier HR (2014) Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling, vol 54. Environmental Modelling & Software, pp 108–127

85. Wu Z, Zhang D, Cai Y, Wang X, Zhang L, Chen Y (2017) Water quality assessment based on the water quality index method in Lake Poyang: the largest freshwater lake in China. Sci Rep 7(1):1–10

86. Wynn E, Krieg MA, Lanham-New SA, Burckhardt P (2010) Postgraduate symposium positive influence of nutritional alkalinity on bone health: Conference on 'over-and undernutrition: challenges and approaches'. Proceedings of the Nutrition Society, 69(1), 166-173

87. Yan F, Qiao D, Qian B, Ma L, Xing X, Zhang Y, Wang X (2016) Improvement of CCME WQI using grey relational method. J Hydrol 543:316–323

88. Yaseen ZM, Ramal MM, Diop L, Jaafar O, Demir V, Kisi O (2018) Hybrid adaptive neuro-fuzzu models for water quality index estimation. Water Resour Manag 32:2227–2245

89. Yisa J, Tijani JO, Oyibo OM (2012) Underground water assessment using water quality index

90. Zhang Q, Li Z (2019) Development of an interval quadratic programming water quality 1213 management model and its solution algorithms. J Clean Prod 119319:1214. https://doi.org/10.1016/j.jclepro.2019.119319
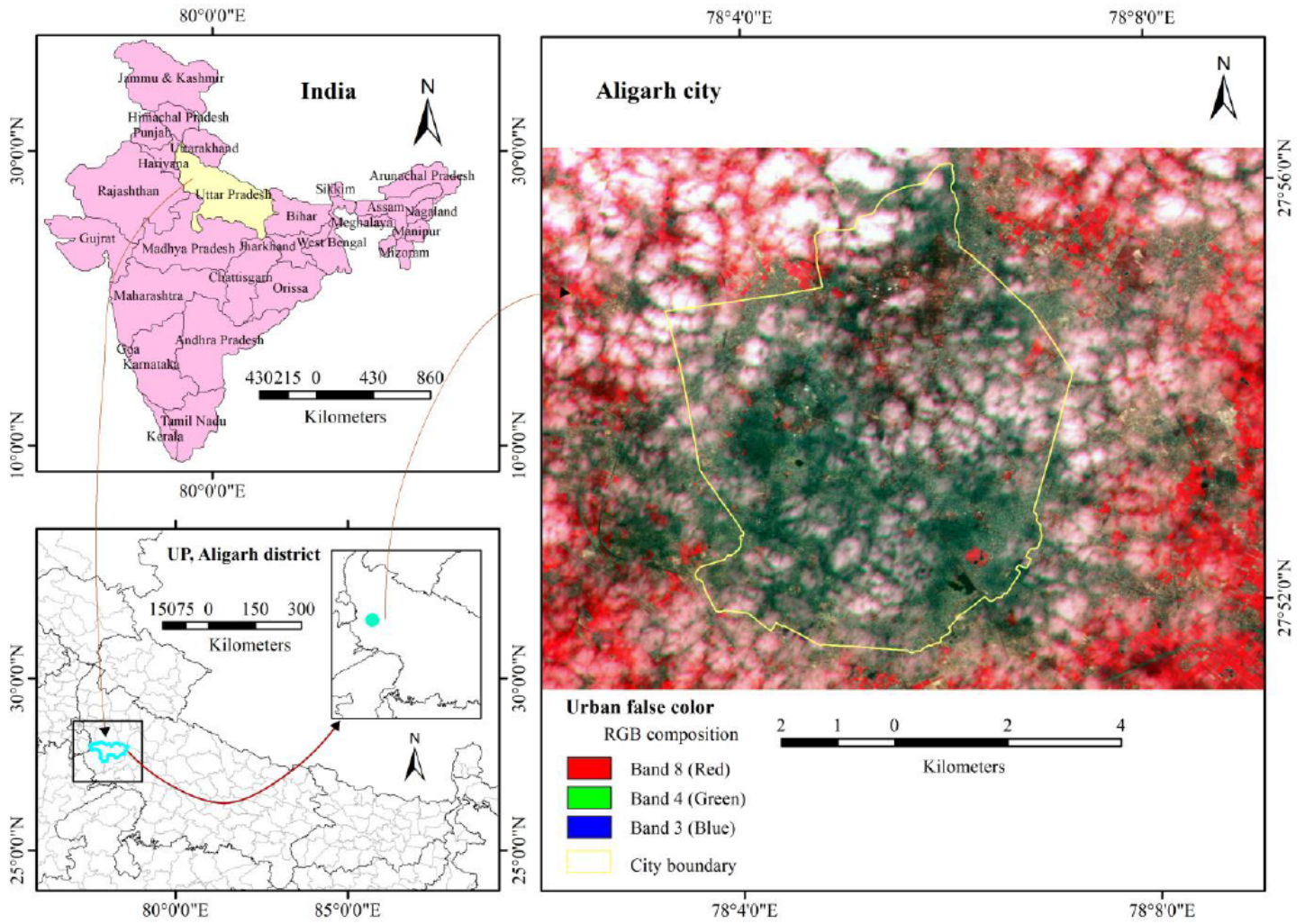
Figures

Figure 1

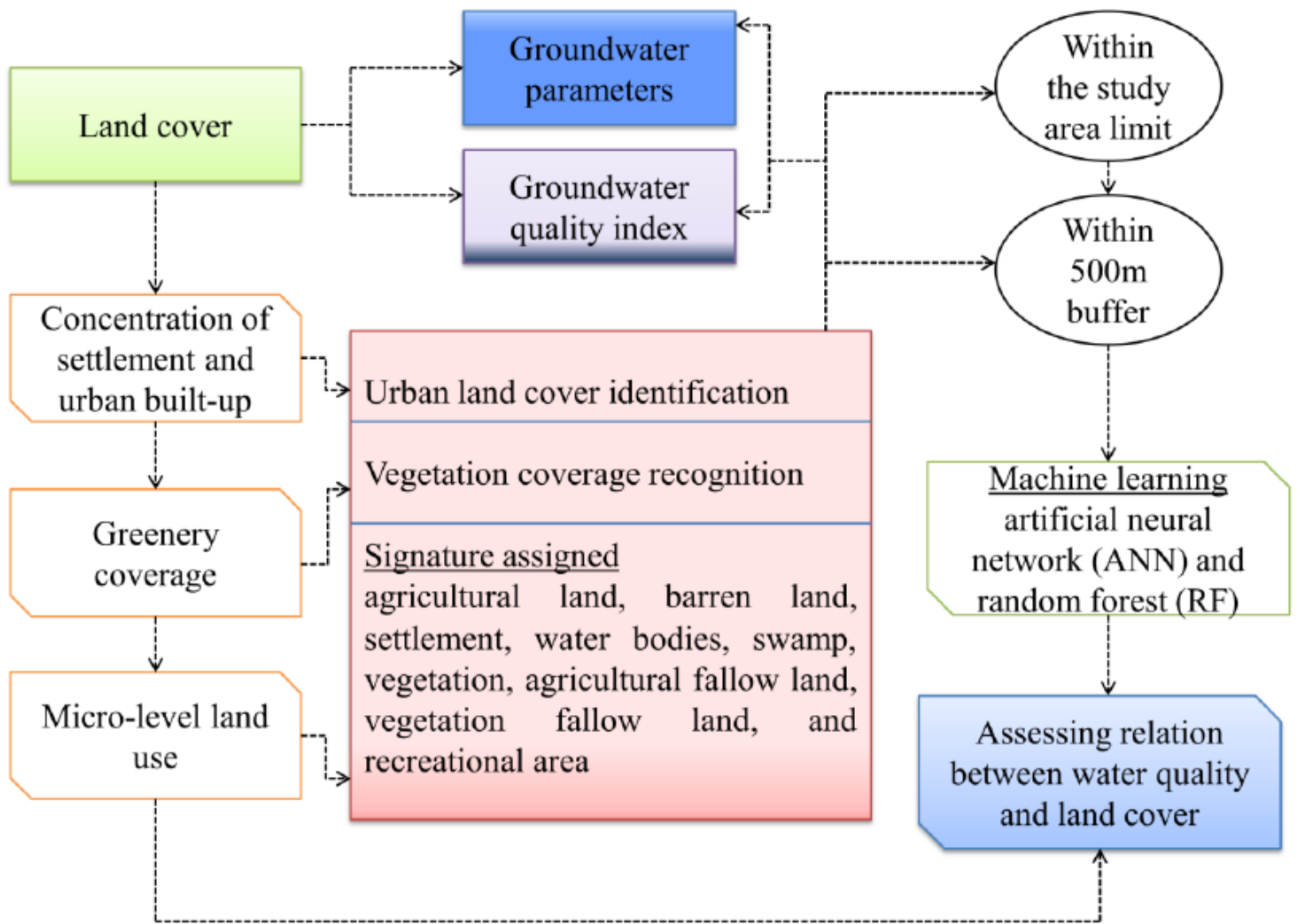Location of the study area

Figure 2

Methodological flow chart, presenting a graphical view of the methods used in this study
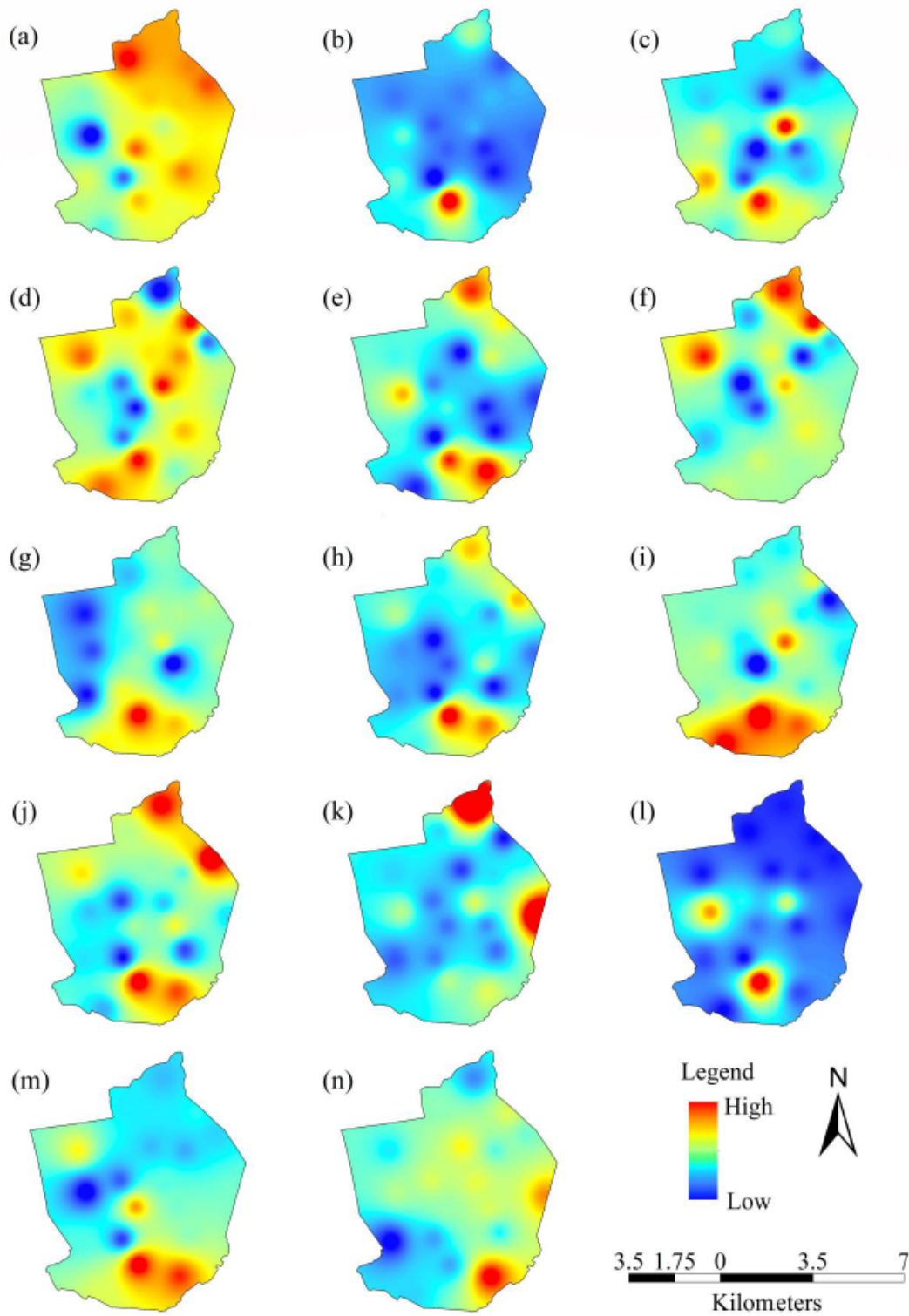
Figure 3

Physico-chemical parameters of groundwater during pre-monsoon (a) pH (b) TDS (c) TSS (d) Turbidity (e) Chloride (f) Alkalinity (g) DO (h) Total hardness (i) Calcium (j) Magnesium (k) EC (l) Sulphate (m)Bicarbonate (n) Sodium
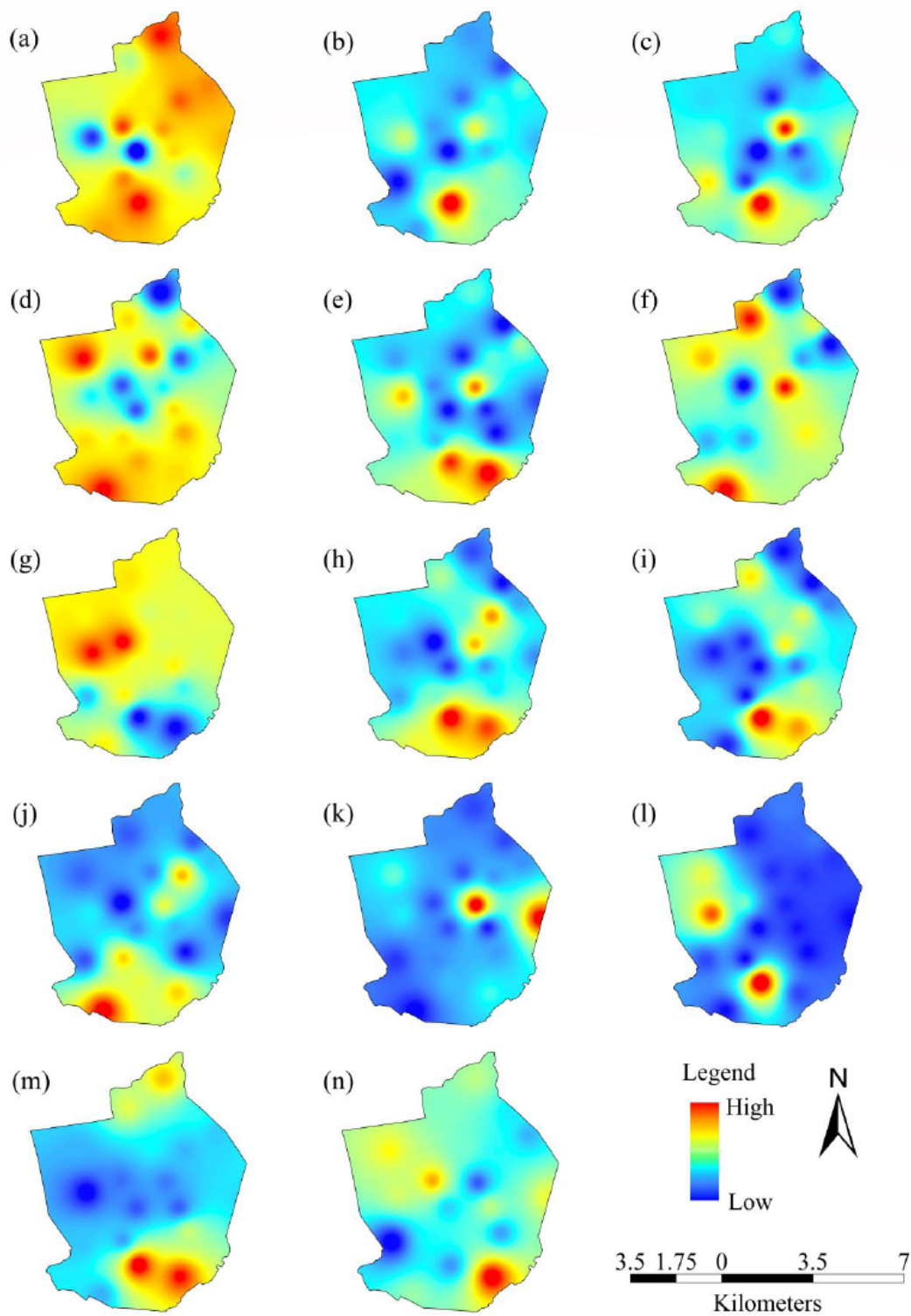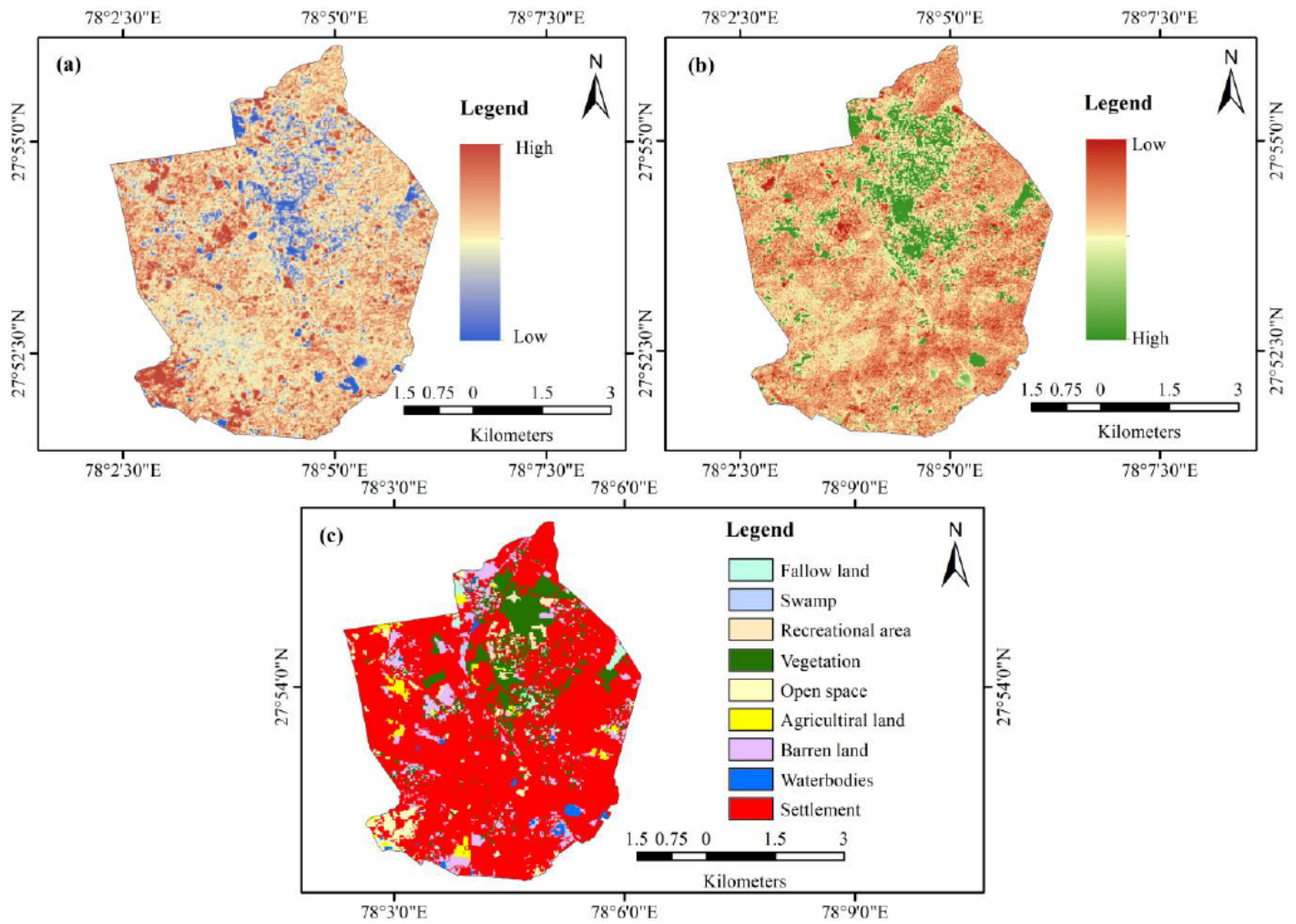
Figure 4

Physico-chemical parameters of groundwater during post-monsoon (a) pH (b) TDS (c) TSS (d) Turbidity (e) Chloride (f) Alkalinity (g) DO (h) Total hardness (i) Calcium (j) Magnesium (k) EC (l) Sulphate (m)Bicarbonate (n) Sodium

Figure 5

Land cover of the study area (a) Concentration of settlement and urban built-up, (b) Greenery coverage, (c) Micro-level land use
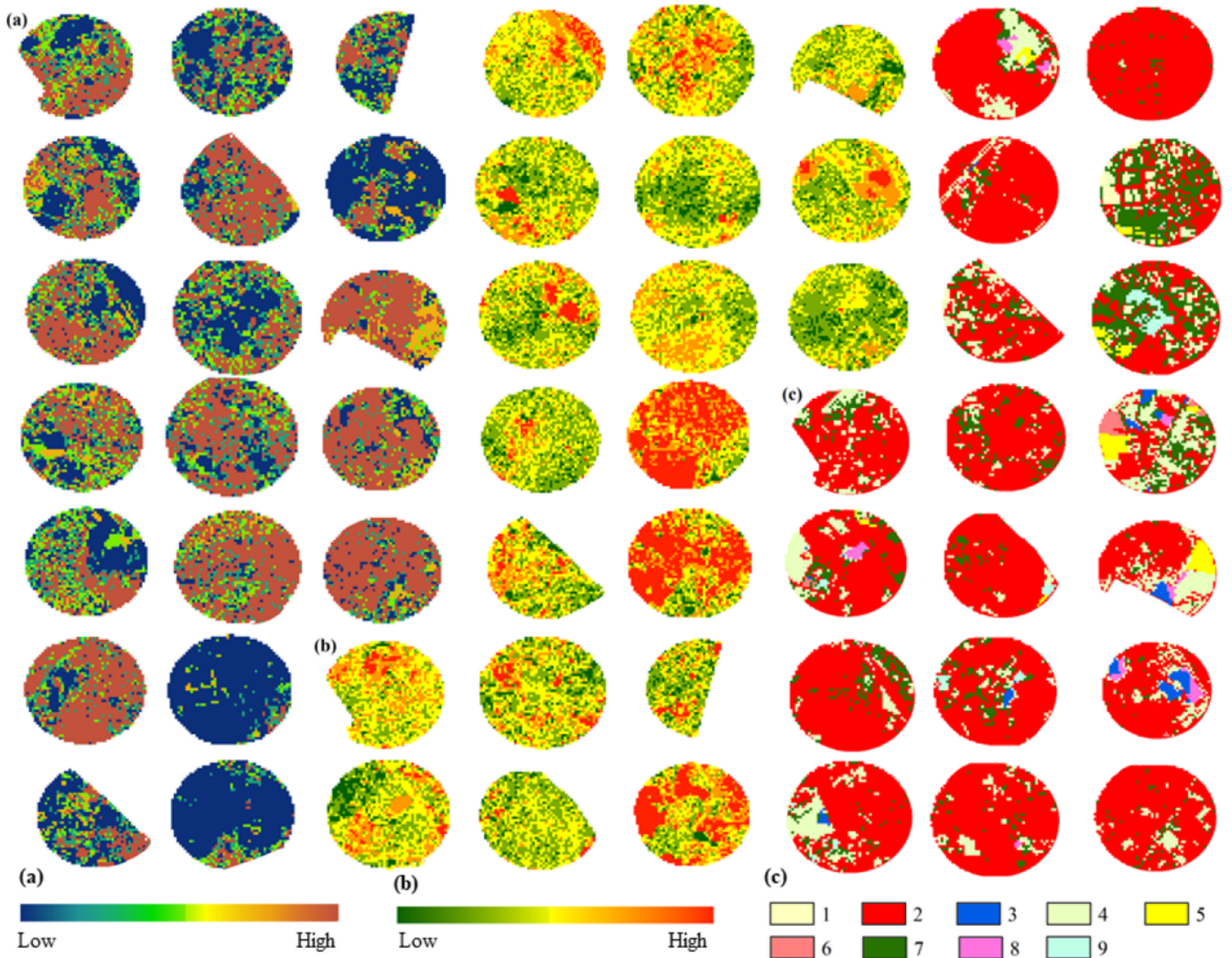
Figure 6

Land cover of the study area within 500m buffers of 19 sampled stations (a) Concentration of settlement and urban built-up, (b) Greenery coverage, (c) Micro-level land use (*Note:* 1 = Recreational areas; 2 = Settlement, 3 = Waterbodies; 4 = Barren land; 5 = Agricultural land; 6 = Open space; 7 = Vegetation; 8 =Swamp; 9 = Fallow land)
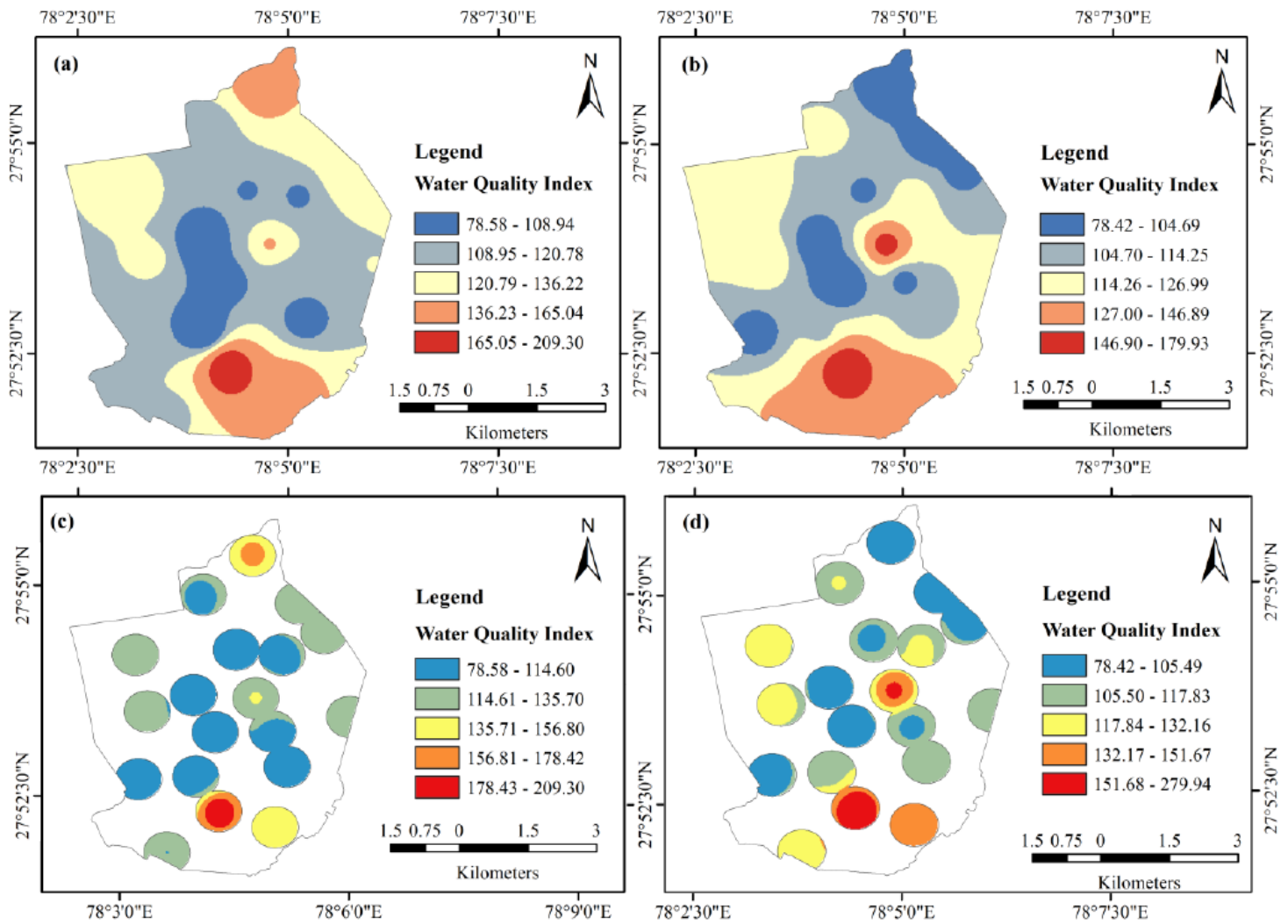
Figure 7

Groundwater quality of the Aligarh city (a) pre-monsoon (b) post-monsoon (c) pre-monsoon within 500m buffer from sampled stations (d) post-monsoon within 500m buffer from sampled stations
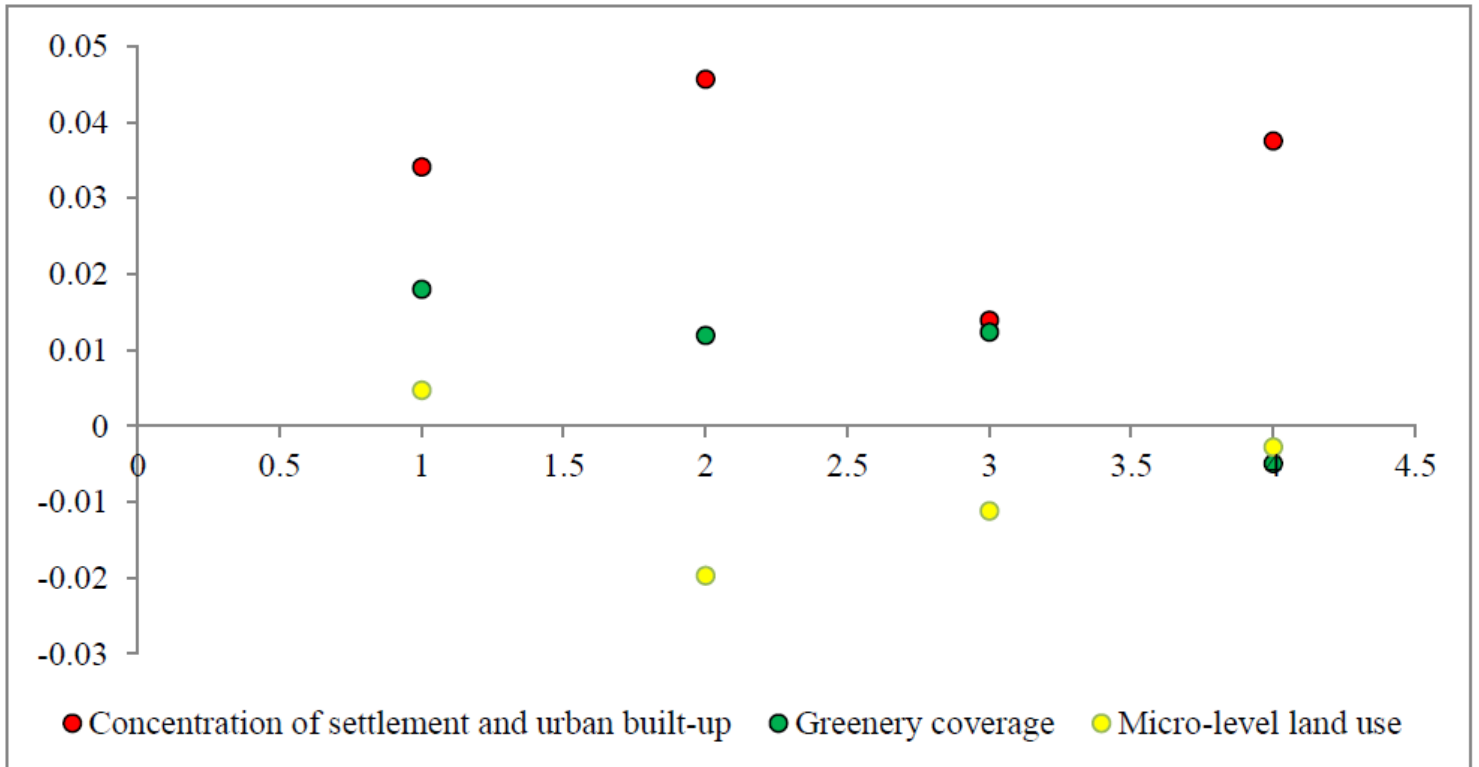
Figure 8

Correlation between land cover and water quality index within the whole study area in pre-monsoon and post-monsoon season
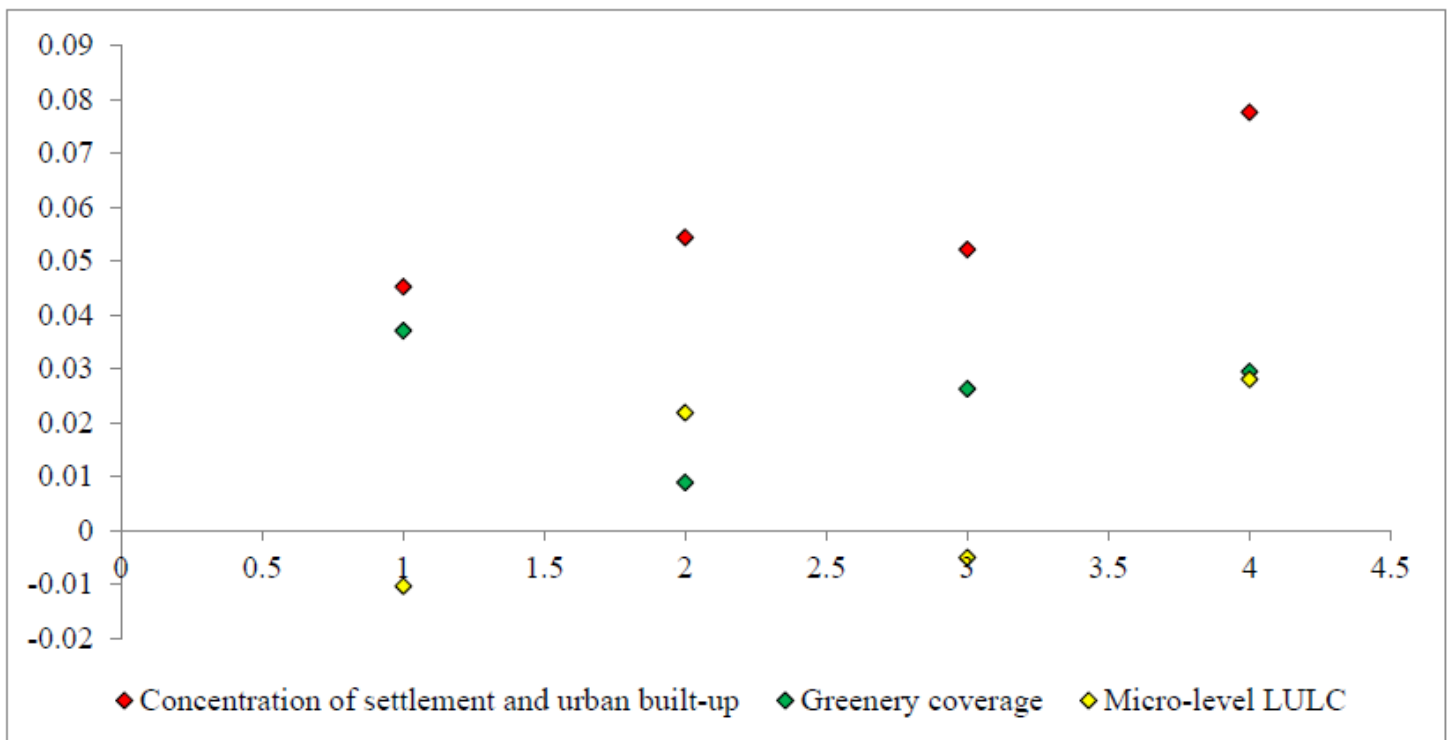

Figure 9

Correlation between land cover and water quality index within the buffer areas in pre-monsoon and post-monsoon season