Water Supply

Assessment of correlation amongst physico-chemical, topographical, geological, lithological and soil type parameters for measuring water quality of Rawal watershed using remote sensing

Mehreen Ahmed, Rafia Mumtaz*, Shahbaz Baig and Syed Muhammad Hassan Zaidi

School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad, Pakistan
*Corresponding author. E-mail: rafia.mumtaz@seecs.edu.pk

MA, 0000-0003-0996-3156; RM, 0000-0002-0966-3957; SB, 0000-0001-5940-0752

ABSTRACT

The quality of water is traditionally assessed by the collection of physico-chemical parameters, i.e., pH, turbidity, dissolved oxygen of the water bodies. However, the variations in environmental factors may have an impact on the quality of water, as changes in these attributes may affect the water bodies. These factors include the topographical, geological, lithological and soil type parameters of the watershed. In this study, the relationship amongst the physico-chemical, topographical, geological, lithological and soil type parameters of the Rawal watershed was evaluated. The parameters included in the present study could be classified as follows: (a) water quality parameters (b) topographical parameters, (c) geological parameters, (d) lithological parameters, and (e) soil type parameters. Water quality parameters consisted of dissolved oxygen, pH, turbidity and temperature. The topographical parameters include the slope and aspect of the watershed while the lithological, geological and soil type parameters include the lithology, geology and soil type of the watershed. Pearson's correlation was used to determine the relationship amongst these different parameters. The results have revealed that the correlations of the topographical, lithological, geological parameters with the water quality parameters in the Rawal watershed for the monsoon seasons of June to August mostly have the same trend. Throughout the four-year time period, turbidity and temperature parameters had positive correlations with soil type (ranging 0.03–0.24), but however had weak correlation with geological and lithological parameters. Dissolved oxygen did not show any relationship with topographical and lithological parameters. The results for pH show that it has weak to fair positive correlations with topographical parameters. However, this analysis is based on the Landsat 8 images extracted for the monsoon seasons of the years 2017–2020, and to examine a more prominent impact of geographical or environmental factors on the physico-chemical features, a large dataset should be considered.

Key words: correlation, lithological parameters, physico-chemical parameters, topographic parameters, water quality monitoring

HIGHLIGHTS

- Pearson's correlation is applied to physico-chemical, topographical, geological, lithological and soil type parameters.
- pH, DO, temperature and turbidity are derived from Landsat-8 data.
- Slope, aspect, lithology, geology and soil type are derived from DEM.
- Turbidity and temperature show fair, whereas DO and pH show weak correlations.
- The environmental features collected play a vital role in the determination of the water quality.

INTRODUCTION

Water is a primary requisite that is essential for the sustenance of life on Earth. Despite its significance, the quality of water is deteriorating constantly through the occurrence of anthropogenic activities like construction, improper sewage disposal and poor agricultural activities (Hamzaoui-Azaza et al. 2011; Pazand & Hezarkhani 2012). Environmental factors like soil erosion can also cause change in the physical landscape which may affect the water bodies (Issaka & Ashraf 2017). Traditionally, physico-chemical water parameters like turbidity, pH, and conductivity are used for estimating the water quality. However, these factors alone are not enough to determine the quality of water. Environmental variables that include topographical

parameters including slope, aspect and geological, lithological and soil type parameters like lithology, geology and soil type may have a direct/indirect impact in determining the overall water quality.

In Pakistan, water deterioration is a serious health concern with the increase in urbanization. Pakistan still relies on traditional methods and tools for collecting water samples manually based on human intervention (Qadir *et al.* 2008; Khan 2010; Nazeer *et al.* 2014; Bhatti *et al.* 2018). This is a tedious and complex task which requires manual labor and is dependent on availability and access to sample collection sites. For this reason, advancements in technology can provide better ways for data acquisition. The use of remote sensing for assessing the water quality has been proposed in many studies (Chen *et al.* 2008; Hsu *et al.* 2009; Wen & Yang 2010). All physico-chemical parameters change the optical properties of the water bodies which makes it easier to detect through the development of remote sensing techniques (Fichot *et al.* 2016). Moreover, considering other hydrological and topographical factors for assessment of water quality can prove to be an essential part, as changes in these attributes may affect the water bodies. Thus, building an enhanced water quality estimation model that overcomes the limitations of the traditional systems, is necessary.

The area selected for this research is Rawal watershed. The parameters included in the present study could be classified as follows: (a) **water quality parameters,** (b) **topographical parameters**, (c) **geological parameters**, (d) **lithological parameters**, and (e) **soil type parameters**. Four physico-chemical parameters, i.e., pH, turbidity, temperature and dissolved oxygen of water are extracted from Landsat 8 (Collection 1 Level 1(C1 L1)) images with Rawal watershed as selected area for the monsoon seasons of the years of 2017–2020. Digital Elevation Model (DEM) is created with Shuttle Radar Topography Mission (SRTM) data to extract topographical, geological, lithological and soil type parameters that include slope, aspect, geology, lithology and soil type. Besides that, to analyse the relationship between these different categories of parameters, Pearson's correlation is performed on the extracted parameters. This study is limited to the satellite imagery collected for the monsoon seasons of the years of 2017–2020. This time period was selected due to the presence of seasonal variations for the selected study area and the availability of satellite images for the chosen parameters that covered the area. The study reveals that there is a significant positive relationship between the physico-chemical parameters and other topographical, geological, lithological and the soil type parameters. However, the lack of any significant results with other features may be dependent on the limitations of this study. Hence, to examine a more prominent impact of geographical or environmental factors on the physico-chemical features, a large dataset should be considered. The major contributions of this study are (1) the extraction of topographical, geological, lithological and the soil type parameters for Rawal watershed from the SRTM DEM, (2) the extraction of physico-chemical parameters using Landsat 8 images for Rawal watershed for the monsoon seasons of 2017–2020, (3) the comparison of all the extracted parameters to establish the importance of other environmental factors for water quality assessment, using the Pearson's correlation method.

The paper is organized, as follows. Section 2 covers the related work pertaining to the remote sensing techniques used for calculation of physico-chemical and other water quality parameters. In the subsequent section, the proposed methodology for the extraction of topographical, geological, lithological and the soil type parameters and physico-chemical water quality is discussed. The results of the applied Pearson's correlation are discussed in Section 4. Section 5 concludes the paper.

RELATED WORK

Remote sensing techniques have been used for extracting the water quality parameters from satellite imagery, as these techniques show an optical variation in the water bodies. Many studies have extracted the physico-chemical parameters using the remote sensing techniques due to the availability of satellite images and the investigation of various estimation algorithms devised for calculating the water quality parameters. In 2015, a model was developed (Abdullah *et al.* 2017) to extract temperature using Landsat 8 images retrieved for Dokan lake, Iraq. The developed model achieved a poor accuracy for images extracted for different seasons. In 2015 (Lim & Choi 2015), turbidity was calculated using a model developed by using Bands 2–5 of Landsat 8. Their model showed that Landsat Bands 2–5 were related to the suspended sediments levels. Kapalanga (2015) extracted turbidity with Landsat 8 images over Olushandja Dam, Namibia, with an accuracy of 0.98 R^2. Similarly other studies like Dogliotti *et al.* (2015) tested wavelength bands (645 and 859 nm) to extract turbidity parameter. A ratio between the blue (Band 2) and red bands (Band 4) was proposed by Theologou *et al.* (2015) to retrieve dissolved oxygen with Landsat 8 images of a lake located in Greece and achieved an accuracy of 0.80 R^2. In 2015 (Bonansea *et al.* 2015), Landsat TM and Landsat ETM+ data were compared with field data for a lake in Argentina. Correlations amongst Landsat bands and the water quality were observed using regression analysis. The results revealed that Band 3 of Landsat images has a

prominent effect on the prediction of water quality. In 2017 (Markogianni *et al.* 2017), Landsat 8 images for the years 2013 and 2014 were retrieved for a lake in Greece along with 22 water samples collected for August 2014. Correlation analysis was performed between Landsat 8 bands combinations. The study concluded that concentrations of ammonium and chlorophyll-a (Chl-a) were more precise than nitrate, nitrite, phosphate and total nitrogen concentrations. In 2018 (Liu *et al.* 2018), Chl-a was extracted from images of Sentinel-2A (S2A) acquired on October 7, 2016 for the study area of Harsha Lake, and an ensemble approach was applied that showed a 20.6% improvement over the other methods. A recent study (Sharaf El Din 2020) used five Landsat 8 images for the years 2015 and 2016 to observe the turbidity and total suspended solids parameters. A comparison was made between the manually collected samples with the Landsat 8 data points using a stepwise regression technique to reveal a positive correlation amongst the parameters.

Along with these traditionally used physico-chemical features that are extracted by applying remote sensing techniques, other environmental factors can also be derived using the satellite images. These include topographical parameters that can be extracted from remote sensing techniques using SRTM DEM data. A study in 2014 (Beeson *et al.* 2014), extracted slope and other topographic data from DEM and concluded that special attention should be made in the selection of spatial resolution and input source as they kept changing due to the advancements in remote sensing and can prove critical in water quality models. In 2019 (Oyedotun 2019), land use changes were examined for Chaohu Lake using Landsat MSS and OLI/TIRS images of 1979–2015. The results showed a 25% increase in built up area causing degradation of the basin due to improper land use activities. A study (Oyedotun 2020) extracted hydrological parameters like drainage network for Chaohu Lake using SRTM DEM data created with Landsat MSS and OLI/TIRS images for the time period of 1979–2015. The study focused on analysing the dynamics of the different streams by extracting the changes in land use patterns.

The literature has revealed that, overall, the studies have used Landsat ETM+, Landsat 7, Landsat 8, Landsat 5 Thematic Mapper ™, Sentinel 2, Sentinel 3 and the Terra Moderate Resolution Imaging Spectroradiometer (MODIS) as satellite sensors for extracting water quality parameters. The use of remote sensing techniques for the calculation of the physico-chemical water quality parameters is a much faster and cheaper method that gets rid of the transport and other labor costs involved in the process. However, the lack of accuracy for the remote sensing techniques quality can be a major concern, and needs to be investigated. In addition, the remote sensing techniques have only been used for the calculation of physico-chemical parameters and none of the studies have used these techniques for the extraction of geological, lithological and soil type parameters to assess the water quality. Moreover, there is no evidence of any comparisons made between the topographical, geological, lithological and soil type parameters with the most commonly used parameters i.e. physico-chemical properties for water quality assessment. However, these studies have discussed the importance of the topographical and other environmental parameters in the assessment of water quality and their role needs to be explored.

METHODOLOGY

The methodology employed for analysing the relationship among different physico-chemical, topographical, geological, lithological and soil type parameters for Rawal watershed (see Figure 1) is described in the subsequent subsections and depicted in Figure 2.

Study area

Rawal watershed (Ali *et al.* 2013), as seen in Figure 1, begins at a lake which is located in the capital city of Islamabad at a longitude of 73°7′E, latitude of 33°42′N and altitude of 1,800 m. The lake receives water from four major streams and 43 small streams. The watershed area is surrounded by highly populated places including Bhara Kahu, Bani Gala, Malpur and Noor-pur Shahan. Anthropogenic activities like construction, solid waste disposal, untreated sewage in the tributaries and disposal of poultry waste contribute to the deterioration of water quality. Other natural events like local spring discharges and untreated sewage water occasionally fall in the watershed area during the monsoon season, leading to sedimentation which has decreased the storage capacity from 58,581,753 m^3 to 38,237,937 m^3.

Data acquisition

The data were acquired from the archive *United States Geological Survey (USGS)* (https://earthexplorer.usgs.gov/) in the form of multispectral satellite images, which were comprised of Landsat 8 *(C1 L1)* satellite data for the monsoon seasons in the years 2017–2020 and SRTM data were used for the creation of DEM. The Landsat images were observed for the monsoon seasons, i.e., June to September for the years 2017–2020, which were comprised of a total of 16 satellite images that
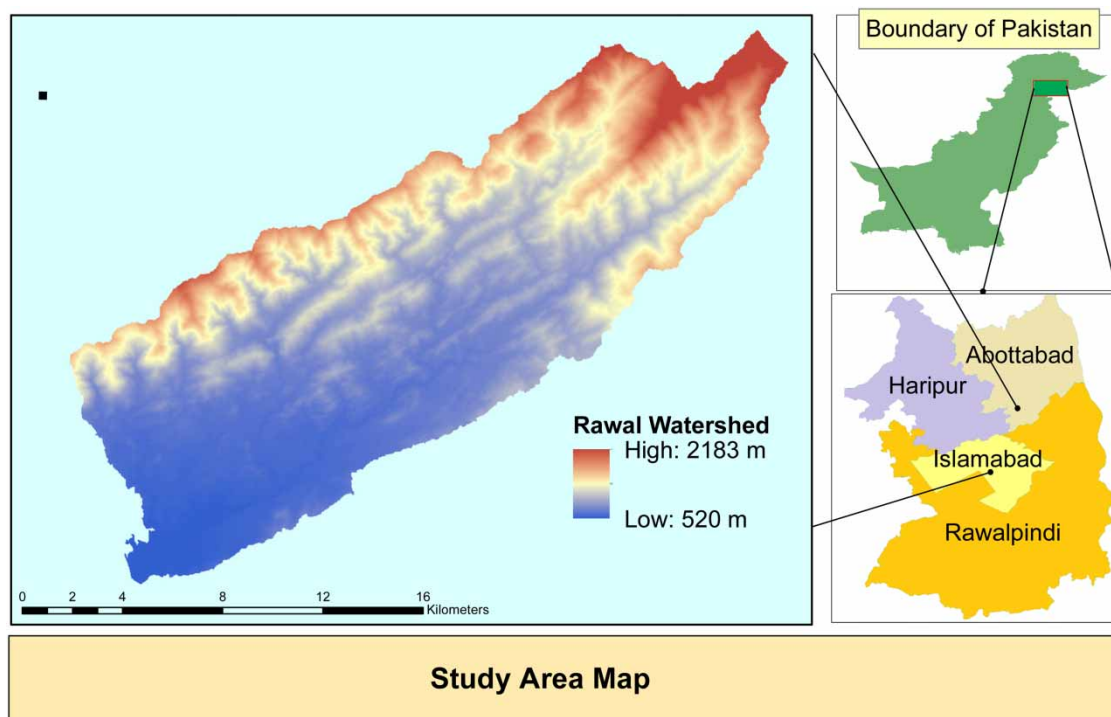
Figure 1 | Study area map showing Rawal watershed.

covered the Rawal watershed. These satellite images were used to perform band calculations to acquire water quality parameters for the watershed including *'turbidity'*, *'pH'*, *'Dissolved Oxygen (DO)'* and *'Land Surface Temperature (LST)'*. The topographical, geological, lithological and soil type parameters including *'slope'*, *'aspect'*, *'geology'*, *lithology'* and *'soil type'* were extracted from DEM based on the sources listed in Table 1. Here, 100 data/sample points were extracted from each satellite image after calculation of the water quality parameters, i.e., 1,600 data points in total. These data points were then correlated with the topographical, geological, lithological and soil type parameters. The monsoon seasons of the years 2017–2020 were selected as the appropriate time period for data acquisition as it contained the satellite imagery for the selected study area. Moreover, these years showed the most seasonal variations that would provide better observations regarding the impact of environmental factors on the quality of water.

Topographical, geological, lithological and soil type parameters extraction from DEM

To extract the topographical, geological, lithological and soil type parameters, the watershed DEM was produced using the method known as Watershed Delineation (Bajjali 2018). The SRTM DEM tiles of 30 m for the area of interest were downloaded from the USGS Earth Explorer repository (https://earthexplorer.usgs.gov/). A single raster was created by mosaicking the individual tiles downloaded from the repository. ArcGIS software was used to perform the water delineation using the Spatial Analyst extension and hydrology tools. There are many steps involved in the creation of the Rawal watershed DEM, which are summarized in Figure 3. After the creation of the watershed DEM, the parameters were extracted. Each parameter is discussed in detail as follows:

Slope

The watershed DEM for the study area can be seen in Figure 4. Slope defines the steepness of a surface and can play a key role in the outlining of the potential water zones. The slope classification was performed by using the slope extension in ArcGIS software on the watershed DEM (Jian-hua 2011). The output was a slope classification raster which gave the slope classes in percentage (%). The slope attribute extracted for the observed study area was classified into three classes: (i) 0–10, (ii) 10–22, (iii) 22–65 as seen in Figure 5.
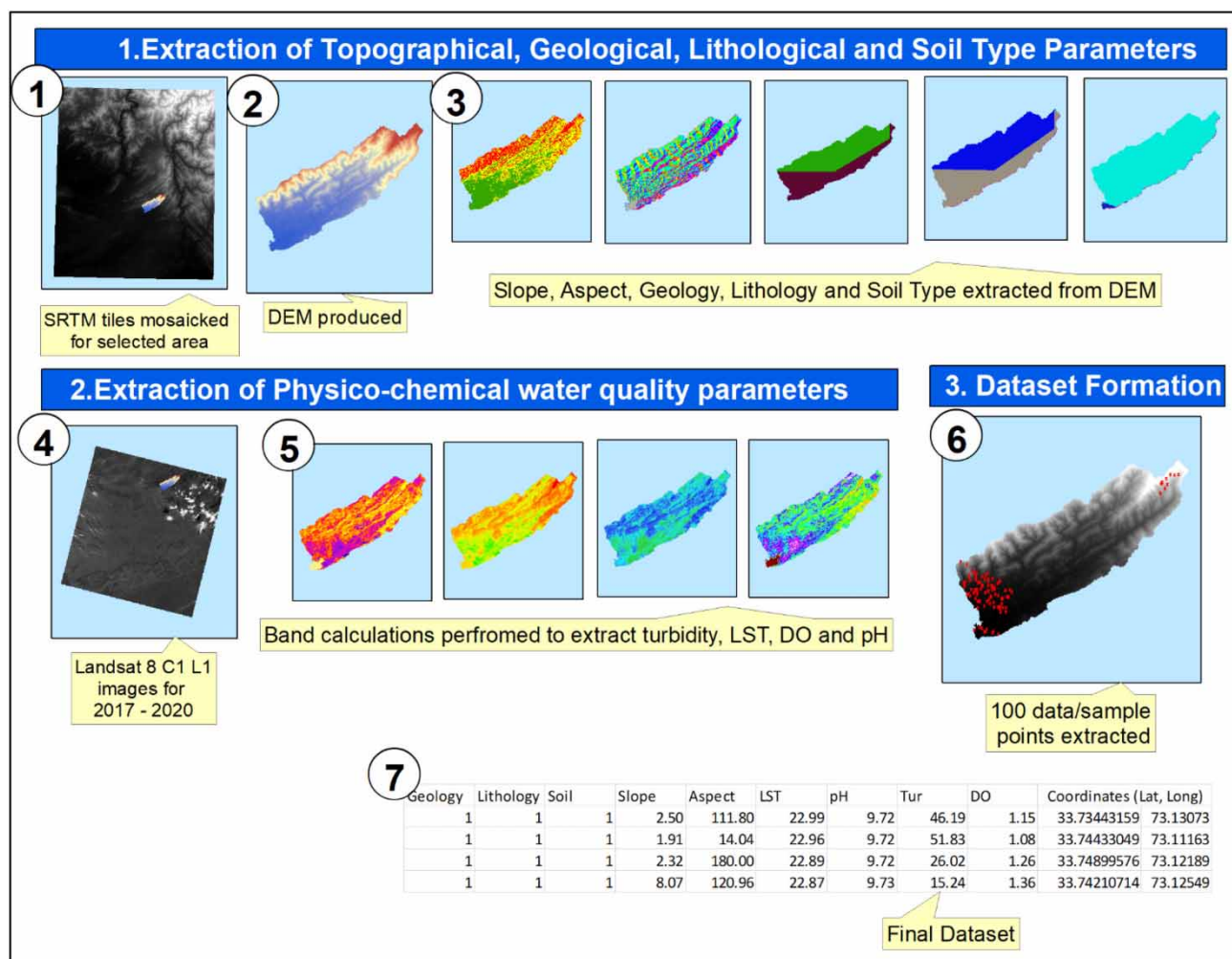
Figure 2 | High level architecture of the methodology.

Aspect

Aspect is measured clockwise in degrees (°) and is the representation of the orientation of slope. The aspect parameter can indicate the compass direction to which the topography faces (Magesh *et al.* 2013). The aspect parameter was derived from the watershed DEM using the aspect extension in ArcGIS software. The aspect was distributed in ten classes namely: (i) Flat (−1), (ii) North (0–22.5), (iii) Northeast (22.5–67.5), (iv) East (67.5–112.5), (v) Southeast (112.5–157.5), (vi) South (157.5–202.5), (vii) Southwest (202.5–247.5), (viii) West (247.5–292.5), (ix) Northwest (292.5–337.5) and (x) North (337.5–360), that can be seen in Figure 5.

Soil type

There are four different types of soil: (i) clay, (ii) sandy and gravel, (iii) silt and (iv) loam soils. Pakistan has a diverse landscape due to its unique climatic characteristics. The Digital Soil Map of the world was downloaded in the form of shape files from an online resource freely accessible (https://worldmap.harvard.edu/data/geonode:DSMW_RdY). The shape file was loaded in ArcGIS software and the extract tool was used to obtain the soil type classification for the area of interest. The soil types for the Rawal watershed are classified as (i) Be – Eutric Cambisols and (ii) Rc – Calcaric Regosols.

Lithology

Lithology is the general physical characteristic or the composition of rocks, like limestone or sandstone. The lithological data of the world were downloaded in the form of shape files from an online resource freely accessible

Table 1 | Physico-chemical, topographical, geological, lithological and soil type parameters and their respective sources

Category	Type of data	Sources
Physico-chemical parameters	*Turbidity* *pH* *DO* *LST*	Landsat-8 satellite (https://earthexplorer.usgs.gov/)
Topographical, geological, lithological and soil type parameters	*Slope*	Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) (https://earthexplorer.usgs.gov/)
	Aspect	Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) (https://earthexplorer.usgs.gov/)
	Soil Type	Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) Digital Soil Map (https://worldmap.harvard. edu/data/geonode:DSMW_RdY)
	Geology and Lithology	Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) (http://geotypes.net/downloads.html)
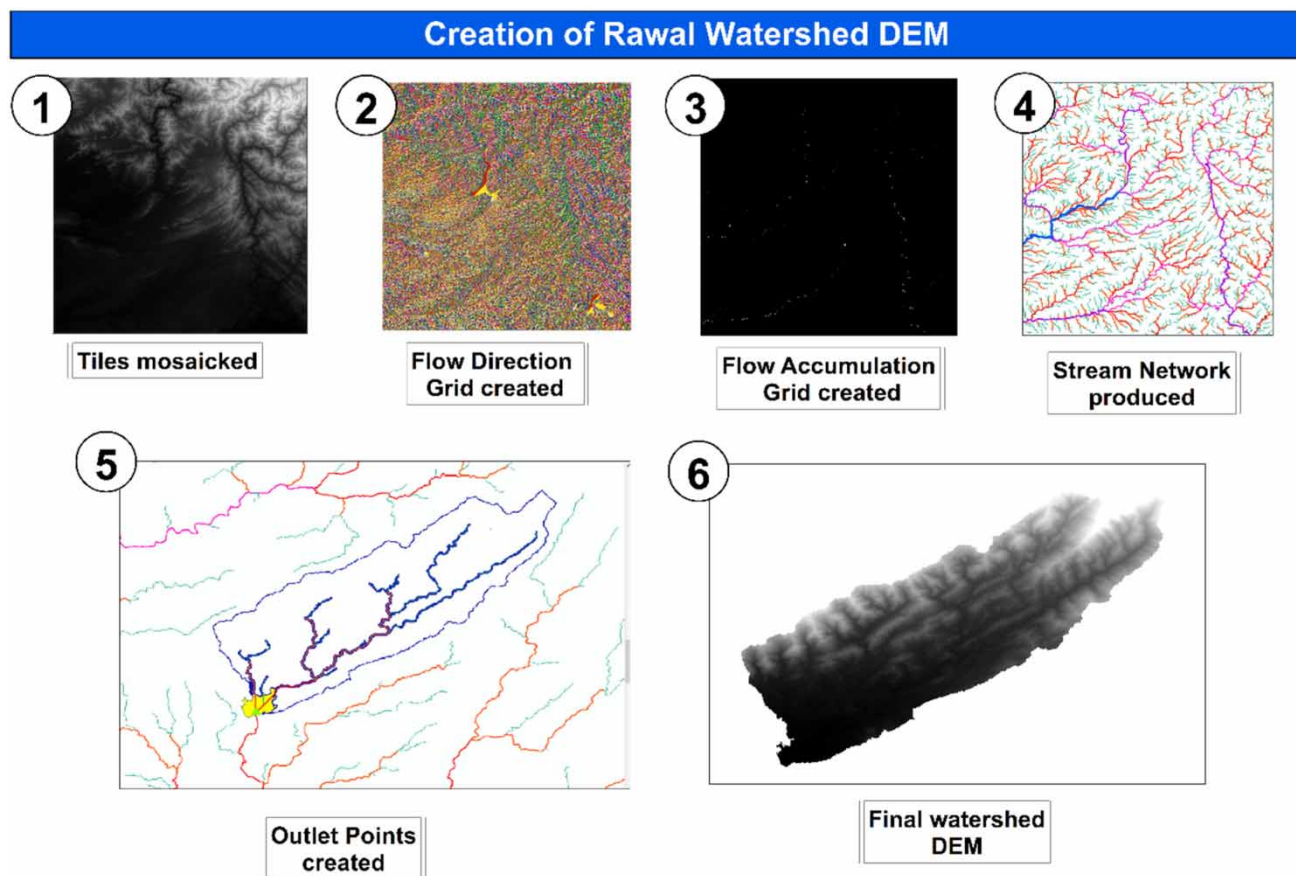
Figure 3 | Watershed delineation.

(http://geotypes.net/downloads.html). The shape file was loaded in ArcGIS software and the extract tool was used to obtain the lithological classification for the selected area. Lithology for Rawal watershed has (i) Siliciclastic Sedimentary Consolidated (Ss) and (ii) Mixed Sedimentary Consolidated (Sm) rocks.
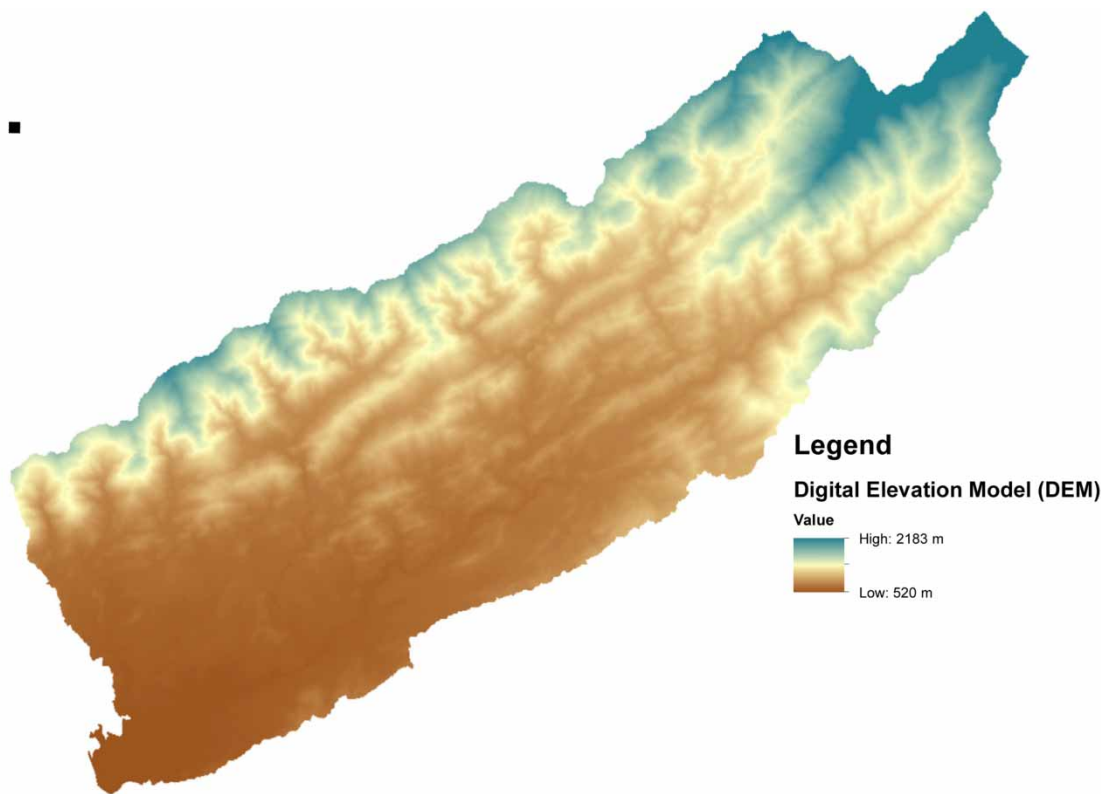
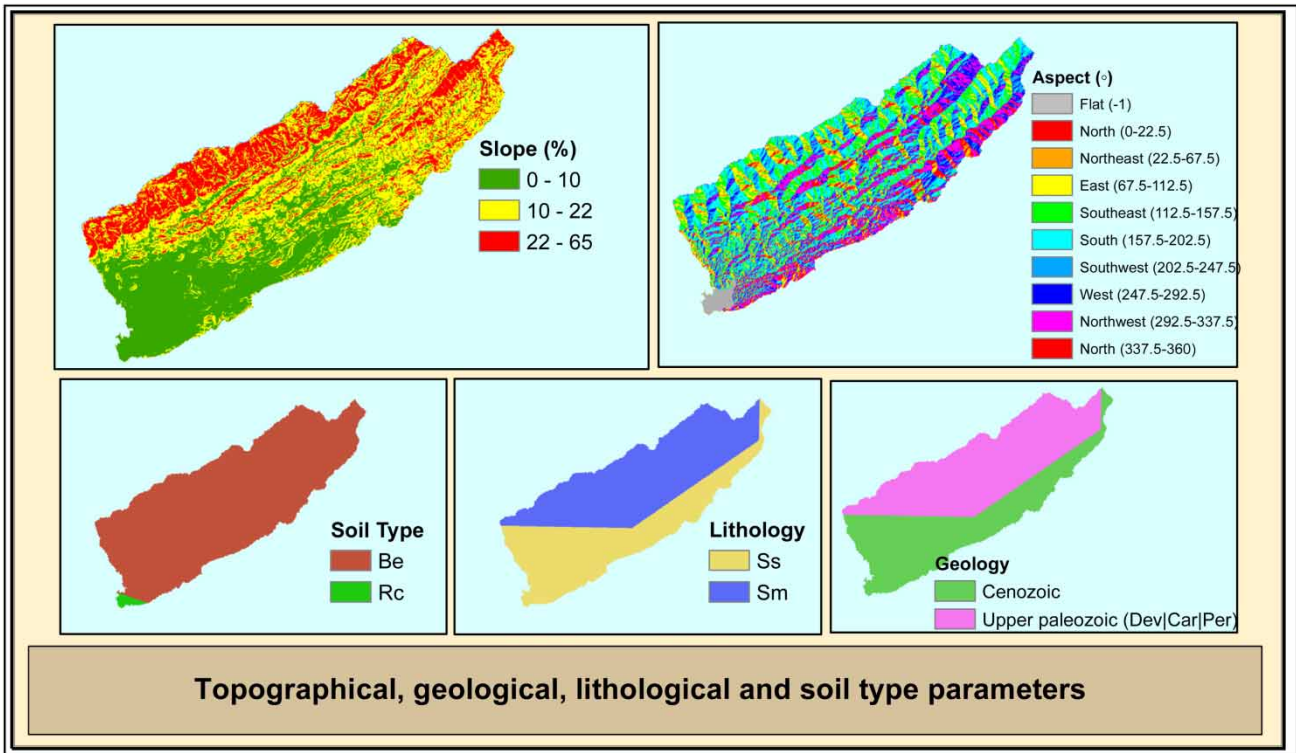**Figure 4** | Digital Elevation Model (DEM) for Rawal watershed.

Figure 5 | Topographical, geological, lithological and soil type parameters extracted from DEM.

Geology

Geology refers to the evolution of natural minerals and the energy resources. The geological data of the world were retrieved in the form of shape files from an online resource freely accessible (http://geotypes.net/downloads.html). ArcGIS software was used to extract the geological classification for the selected area. The geological formations of the study area are classified as follows: (i) Cenozoic and (ii) Upper Paleozoic (Dev|Car|Per).

The slope, aspect, soil type, lithology and geology of the study area are shown in Figure 5.

Physico-chemical parameters extraction from remote sensing

The physical and geographical data covered by satellites over a large scale can prove to be an essential tool towards understanding the water quality characteristics. Landsat 8 (C1 L1) satellite data for a four-year time period, i.e., 2017–2020, was downloaded from the USGS Earth Explorer repository (https://earthexplorer.usgs.gov/). These satellite images contain 11 bands and are high quality Landsat scenes. In total, 16 images were retrieved for the monsoon months (June, July, August and September) for the four-year time period (2017–2020).

In remote sensing, estimation algorithms are applied on satellite images to extract the parameters. These algorithms require the conversion of the Digital Numbers *(DN)* or pixels in Top Of Atmosphere *(TOA)* reflectance. This was done for Landsat 8 images with Equation (1), provided by USGS. After the calculation of the TOA, many algorithms can be applied to analyse the pH, turbidity etc. For this study, the adapted equations for calculating the physico-chemical parameters are given in Table 2. ArcGIS software was used to perform band math on the Landsat images. The calculated parameters were compared to the original manual values obtained for the study area to determine the accuracy of the methodology. The details of each parameter calculation are discussed as follows:

Turbidity

The estimation algorithm adopted for calculating the turbidity parameter was proposed by Kapalanga (2015). The algorithm was selected on the basis of its close estimation of the turbidity parameter as observed for the manually retrieved values of turbidity in the study area. Equation (2) (given in Table 2) shows the turbidity calculation which is based on the combination of TOA of Bands 2, 3, 4 and 5, which can be acquired by applying Equation (1) on the bands. For example, REFLECTANCE_ADD_BAND_2: -0.10000000149011612 and REFLECTANCE_MULT_BAND_2: 0.000019999999494757503, then TOA reflectance for Band 2 is calculated by putting the values in Equation (1) as follows:

$$R_2 = \ 0.000019999999494757503 *' Band\ 2' + (-0.10000000149011612\ )$$

R_3, R_4, R_5 can also be calculated in a similar way. The Kapalanga estimation algorithm has a coefficient of determination (R^2) of 0.02 when observed with the ground truth values obtained for the study area.

Dissolved oxygen

Dissolved oxygen (DO) was calculated with the estimation algorithm proposed by Theologou *et al.* (2015), and Khalil *et al.* (2016). They used a combination of TOA reflectance of Bands 2 and 4, which can be seen in Equation (3). The TOA reflectance of Bands 2 and 4 was calculated by applying Equation (1) on the bands. The Theologou and Khalil estimation algorithm has a R^2 of *0.39* when compared with the ground truth values for the study area.

pH

Khattab & Merkel (2014) based their estimation algorithm on the short wavelength infrared Band 6 to calculate the pH parameter. The TOA reflectance of Band 6 is calculated by applying Equation (1). Then, Equation (4) was used to calculate the pH for the study area. The Khattab and Merkel estimation algorithm gave an R^2 of 0.8 when observed with the ground truth values. The Khattab and Merkel method has been used in many studies for calculation of pH using Landsat 8 images.

Temperature

The Landsat 8 images contain Thermal Infrared *(TIRS)* Bands 10 and 11, that are used to calculate the LST. The reflectance of the Earth's surfaces is measured in DN by the satellite sensors. The LST calculation includes six steps, i.e. the application of Equations (5)–(10). These steps are given as follows:

Table 2 | Calculation for turbidity, pH, DO, LST using Landsat 8 images

Parameters	Equations		Reference
Turbidity	$R_x = M_P * Q_{cal} + A_p$	(1)	Kapalanga (2015)
	$15.31856 - 956.806\,(R_2) - 747.376\,(R_3) + 1742.455\,(R_4) + 165.173\,(R_5)$	(2)	
DO	$(R_2)/(R_4)$	(3)	Theologou *et al.* (2015), Khalil *et al.* (2016)
pH	$9.738 - 0.084\,(R_6)$	(4)	Khattab & Merkel (2014)
LST	$L_{10\lambda} = M_L * Q_{cal} + A_L$	(5)	Avdan & Jovanovska (2016)
	$BT = \dfrac{K_2}{\ln\left(1 + \dfrac{K_1}{L_{10\lambda}}\right)}$	(6)	
	$NDVI = \dfrac{NIR - VIS}{NIR + VIS}$	(7)	
	$P_v = \left(\dfrac{NDVI - NDVI_{MIN}}{NDVI_{MAX} - NDVI_{MIN}}\right)^2$	(8)	
	$\varepsilon = 0.004 * P_v + 0.986$	(9)	
	$LST = \dfrac{BT}{\left(1 + \left(\lambda * \dfrac{BT}{\rho}\right) * (\ln(\varepsilon))\right)} - 273 \cdot 15$	(10)	

Here $R_x = $ *TOA reflectance for* Band number x, $M_P = $ REFLECTANCE_MULT_BAND_x, $Q_{cal} = $ Standard pixels of Band x or DN of Band x, $A_P = $ REFLECTANCE_ADD_BAND_x where x is the Band 2, 3, 4, 5 and 6, respectively.

$L_{10\lambda} = $ TOA spectral radiance for Band 10, $M_L = $ RADIANCE_MULT_BAND_10, $A_L = $ RADIANCE_ADD_BAND_10, $K_1 = $ K1_CONSTANT_BAND_10, $K_2 = $ K2_CONSTANT_BAND_10, $NIR = $ Band 5, $VIS = $ Band 4.

$\lambda = $ wavelength of emitted radiance for Band 10 (10.895 μm), $\rho = h * \dfrac{c}{\sigma} = 1.438 * 10^{-2}$ m.K. Here: $\sigma = $ Boltzmann constant ($1.38 * 10^{23}$ J/K), $h = $ Planck's constant (6.626×10^{34} J.s) and $c = $ velocity of light ($2.998 * 10^8$ m/s).

1. TOA spectral radiance ($L_{10\lambda}$): The first step is to convert the DN of Band 10 to $L_{10\lambda}$ using Equation (5) mentioned in Table 2. RADIANCE_MULT_BAND_10 and RADIANCE_ADD_BAND_10 properties of the image are used for calculating the $L_{10\lambda}$.
2. Brightness Temperature (*BT*): The next step is to calculate the BT, which is the uncorrected temperature recorded by satellites. BT is calculated with parameters: $L_{10\lambda}$, K1_CONSTANT_BAND_10, K2_CONSTANT_BAND_10 and applying Equation (6).
3. Normalized Difference Vegetation Index (*NDVI*): Subsequently, the normalized difference of the Near Infrared Region (*NIR*), i.e., Band 5 and Visible Red Region (*VIS*) i.e., Band 4, using Equation (7), give the NDVI.
4. Proportion of vegetation (P_v): The minimum and maximum values of NDVI give the P_v value, using Equation (8).
5. Emissivity (ε): The ε is then calculated using the Equation (9). Here, 0.986 corresponds to the correction value of the equation.
6. LST: Finally, LST can now be retrieved by applying Equation (10).

RESULTS AND DISCUSSION

Data extracted from the Landsat imagery and SRTM data are discussed in detail in this section. In addition, the correlations between the two different datasets are also explained in detail.

The topographical, geological, lithological and soil type parameters retrieved for the selected area were observed. The slope attribute extracted for the observed study area is classified into three classes: (i) 0–10, (ii) 10–22, (iii) 22–65. The aspect is distributed in ten classes, namely: (i) Flat (−1), (ii) North (0–22.5), (iii) Northeast (22.5–67.5), (iv) East (67.5–112.5), (v) Southeast (112.5–157.5), (vi) South (157.5–202.5), (vii) Southwest (202.5–247.5), (viii) West (247.5–292.5), (ix) Northwest (292.5–337.5) and (x) North (337.5–360). Geology and lithology are classified in two classes each. Lithology for Rawal watershed has (i) Siliciclastic Sedimentary Consolidated (Ss) and (ii) Mixed Sedimentary Consolidated (Sm) rocks. Similarly, geology has

two classes referred as (i) Cenozoic and (ii) Upper Paleozoic (Dev|Car|Per). Moreover, the soil type presence in the Rawal watershed is classified as (i) Be – Eutric Cambisols, (ii) Rc – Calcaric Regosols.

The values observed for the monsoon months of June to September, 2017 show that 90% of the data consists of the geology parameter classified as Cenozoic, the lithology parameter is classified as Ss rocks and Be – Eutric Cambisols soil type. The slope parameter lies in the first and second class, i.e., (0–10) and (10–22). The aspect parameter lies in the fifth and sixth class, i.e., southeast and south classes. The LST parameter ranges from 10 °C to 40 °C. The turbidity parameter ranges from 13 NTU to 300 NTU, while the pH parameter lies in the 9.7 range. The values for DO range from 0.9 to 2 mg/l. A similar trend is observed throughout the years 2018–2020. Table 3 shows the parameters extracted from specific coordinates from the satellite image retrieved on 23 July, 2017. This table shows only 20 sample points out of the 100 total extracted points for the satellite image that was retrieved on 23 July, 2017. The physico-chemical water quality parameters were extracted through the band calculations that are mentioned in Table 2, and the topographical, geological, lithological and soil type parameters were extracted from the DEM. Later on, the data/sample points were extracted from the calculated parameters. The parameters and their respective coordinates can be seen in Table 3.

The physico-chemical parameters extracted for the month of July, 2019 using equations mentioned in Table 2 can be seen in Figure 6. The LST for Rawal watershed for the month of July lies between 3 °C and 17 °C. The highest values observed for turbidity, pH and DO in this month are 174 NTU, 9.7 and 1.73 mg/l, respectively. Figure 6 also shows the physico-chemical parameters extracted for the month of September, 2020. For this month, the values for the LST, turbidity and DO are high compared to the ones seen for the month of July. The LST lies between 21 °C and 37 °C. The highest values observed for turbidity, pH and DO in this month are 179 NTU, 9.73 and 2.2 mg/l respectively.

Figures 7 and 8 display the Pearson's correlation matrices used to examine the relationship between physico-chemical, topographical, geological, lithological and soil type parameters of Rawal watershed for the monsoon seasons in 2017–2020. The results reveal that the correlations of topographical, geological, lithological and soil type with water quality

Table 3 | Parameters calculated from satellite image retrieved on 23 July, 2017 for the shown coordinates

| Geology (1: Cenozoic, 2: Upper Paleozoic (Dev|Car|Per)	Lithology (1: Ss, 2: Sm)	Soil type (1: Be, 2: Rc)	Slope (%)	Aspect (°)	LST (°C)	pH (–)	Turbidity (NTU)	DO (mg/l)	Coordinates (Lat, Long)	
1	1	1	2.50	111.80	22.99	9.72	46.19	1.15	33.73443159	73.13073255
1	1	1	1.91	14.04	22.96	9.72	51.83	1.08	33.74433049	73.11162801
1	1	1	2.32	180.00	22.89	9.72	26.02	1.26	33.74899576	73.12188865
1	1	1	8.07	120.96	22.87	9.73	15.24	1.36	33.74210714	73.12549289
1	1	1	2.65	195.26	22.86	9.72	37.29	1.18	33.74361493	73.12977771
1	1	1	3.35	146.31	22.76	9.72	51.52	1.08	33.74590088	73.11407627
1	1	1	1.47	161.57	22.70	9.72	37.64	1.22	33.74324617	73.12924601
1	1	1	0.00	0.00	22.67	9.72	45.15	1.11	33.73921368	73.10137844
1	1	1	2.65	217.88	22.65	9.72	30.01	1.21	33.74498387	73.10718806
1	1	1	4.28	347.47	22.62	9.73	8.09	1.45	33.75051672	73.12109984
1	1	1	2.56	275.19	22.60	9.72	36.25	1.22	33.74561934	73.1192637
1	1	1	5.83	353.16	22.58	9.73	12.58	1.40	33.75042329	73.12088598
1	1	1	10.57	182.49	22.41	9.73	11.29	1.41	33.75371586	73.11800734
1	1	1	10.42	131.42	22.40	9.73	13.89	1.39	33.74965462	73.12964656
1	1	1	10.08	335.77	22.32	9.73	11.50	1.41	33.75269134	73.1158707
1	1	1	5.72	68.63	22.17	9.72	40.77	1.17	33.73996621	73.11517762
2	2	1	17.74	171.25	22.13	9.72	24.70	1.36	33.87557812	73.37404167
1	1	1	7.52	259.38	22.07	9.72	42.71	1.15	33.74133262	73.1101826
1	1	1	1.86	90.00	22.04	9.72	21.87	1.45	33.73663426	73.13327514
1	1	1	3.83	75.96	22.03	9.72	33.71	1.26	33.74708813	73.09742681
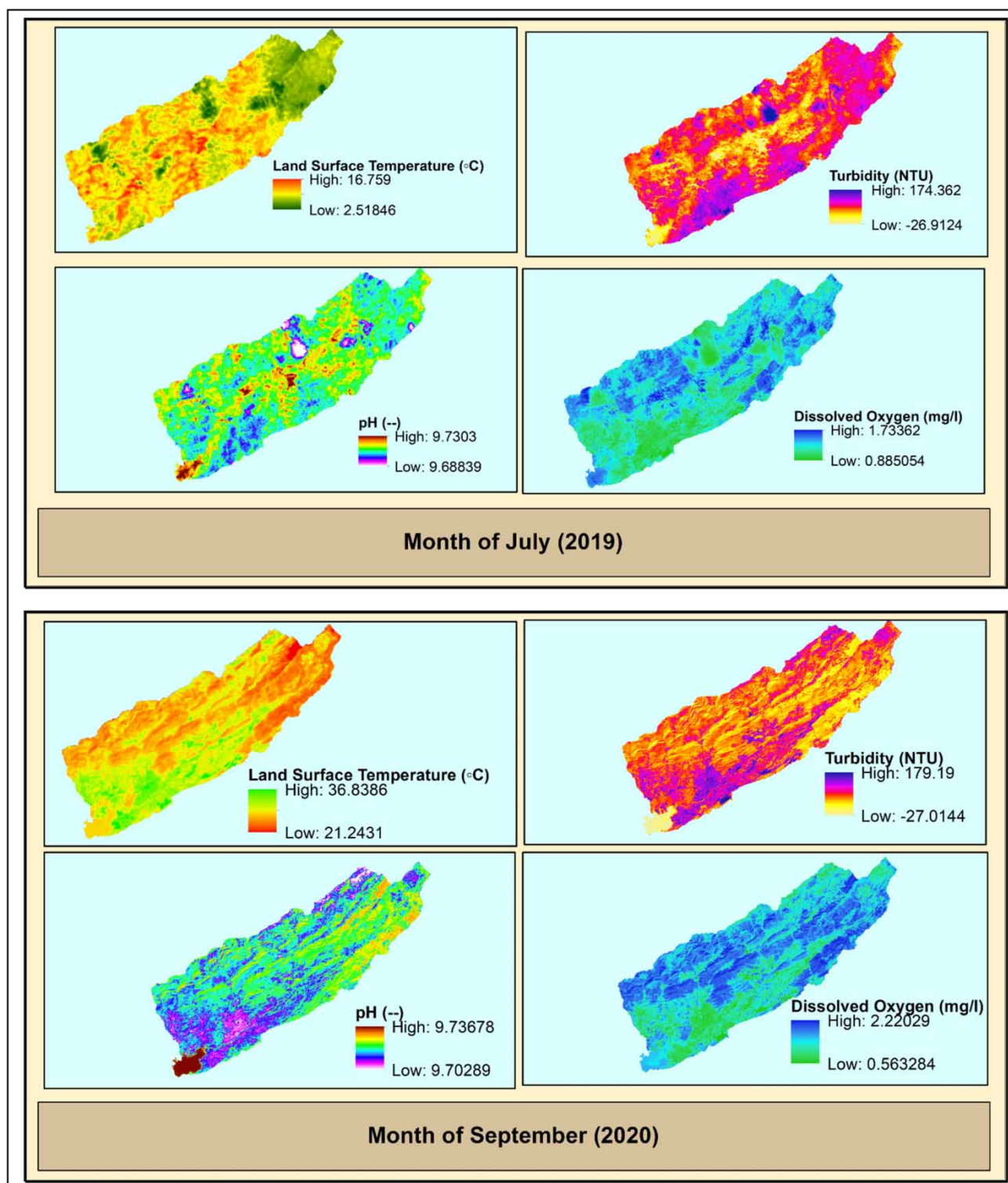
Figure 6 | Physico-chemical parameters extracted from LANDSAT 8 images for the months July, 2019 and September, 2020.

parameters in the Rawal watershed for the monsoon seasons in June to September mostly had the same trend, throughout the four-year time period.

For the year 2017, the turbidity parameter has positive correlations with topographical parameters i.e., slope, aspect and soil type. The pH observed for the Rawal watershed varied from 5 to 9 and was calculated with the equations mentioned
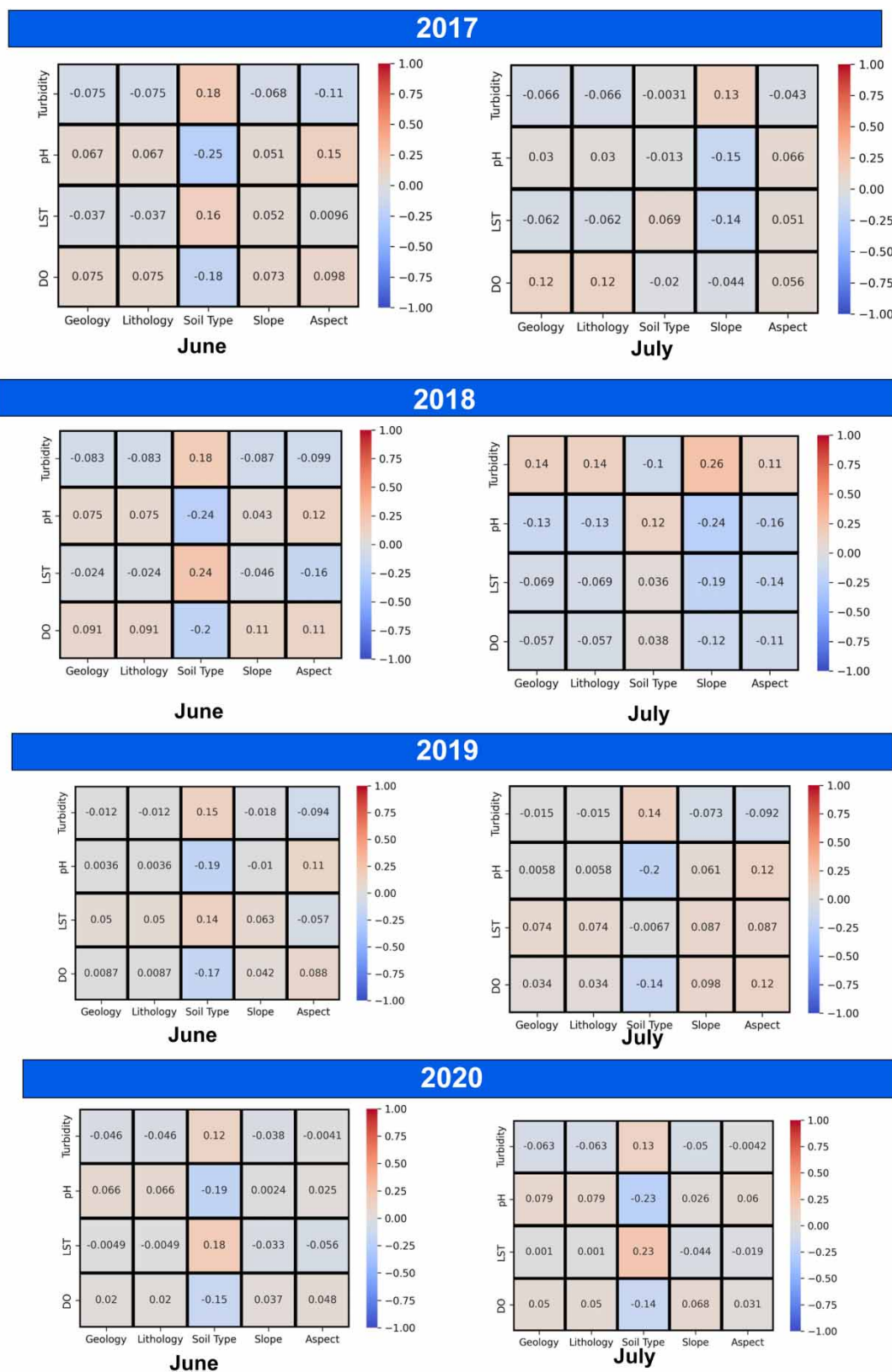
Figure 7 | Correlations among water quality and topographical, geological, lithological and soil type parameters observed for the months of June and July (2017–2020).
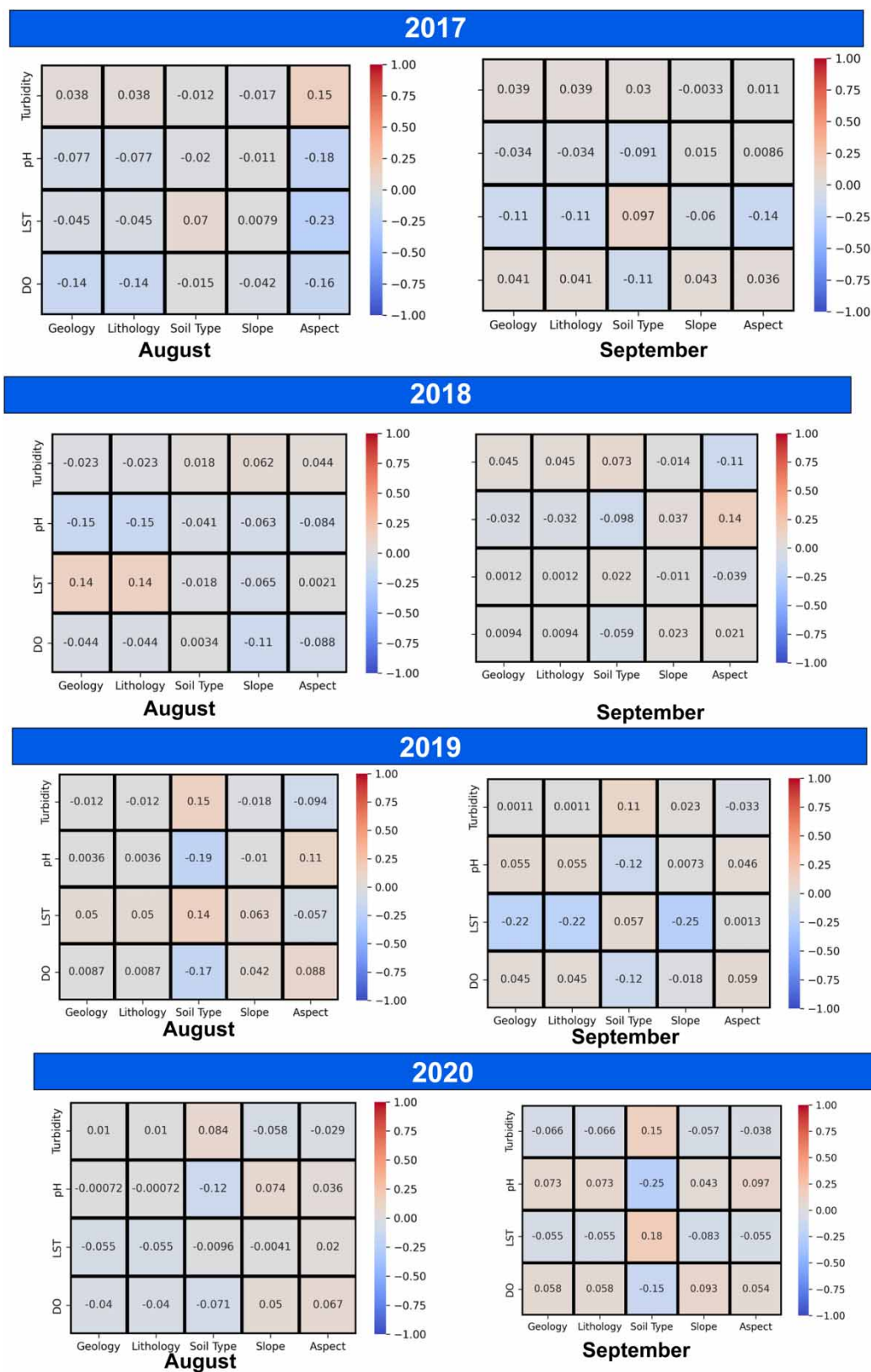
Figure 8 | Correlations among water quality and topographical, geological, lithological and soil type parameters observed for the months August and September (2017–2020).

in Table 2. The pH parameter had weak correlations with the topographical, geological, lithological and soil type parameters. The LST showed weak correlations with all parameters except for the soil type parameter (ranging 0.07–0.2). The DO parameter has positive correlations with all parameters including geology (ranging 0.04–0.12), lithology (ranging 0.04–0.12), slope (ranging 0.04–0.07) and aspect (ranging 0.04–0.1) except for the soil type parameter. The best correlations were observed for the month June, 2017. For the months July, August and September, 2017, it showed a weak to fair relationship.

For the year 2018, the correlations of turbidity with slope (ranging −0.01 to 0.26), geology and lithology (ranging −0.02 to 0.1) showed a weak to fair relationship. The relationship with soil type was observed to be fair (ranging −0.1 to 0.2). A fair correlation was observed for DO with slope and aspect (ranging −0.1 to 0.1). With geology and lithology, correlation with a range of −0.06 to 0.1 was observed. However, a weak relationship was observed with soil type (ranging −0.02 to 0.04). The LST values observed lie in the range 10 °C–40 °C. The LST parameter showed a better relationship with the soil type parameter compared to other parameters. The pH parameter showed a fair relationship with aspect parameter compared to the geological, lithological and soil type parameters. The best correlations were observed for the months June and July, 2018.

The values were observed for turbidity range from 13 to 300 NTU. For the year 2019, the turbidity parameter showed a fair relationship with soil type (range 0.11–0.15), but no relationship with the topographical, geological and lithological parameters. The pH parameter showed a fair relationship with aspect parameter. The relationship between LST and soil type was observed to be fair (ranging 0.05–0.2). DO did not show any relationship with any parameters and a weak relationship with soil type was observed. The best correlations were observed for the months June and August, 2019.

Dissolved oxygen varied from the range of 0.9–2.2 mg/l. For the year 2020, the relationship of turbidity and LST parameters with the soil type parameter showed a fair relationship (ranging 0.12–0.23). These parameters did not show any relationship with the topographical, geological and lithological parameters. DO and pH parameters showed weak correlations overall. The best correlations were observed for the months July and September, 2020.

In the four-year time period, turbidity and LST parameters showed a fair relationship with the soil type parameter, but a weak relationship with the geological and lithological parameters. This is relevant as the water turbidity is caused by clay and other microscopic organisms that affect the water clarity. The DO parameter only showed positive results for the year 2017 and did not show any good results in the later years i.e., 2018–2020. The pH parameter gave a fair relationship with the aspect parameter for 2018 and 2019 compared to the other parameters. The results revealed that there is in fact a relationship found between the physico-chemical and other environmental factors. Thus, pH and DO did not show a very strong relationship with the observed points extracted from satellite data. The LST and turbidity parameters show the presence of its effect on all topographical, geological, lithological and soil type parameters. The best results were observed for the month of July in all four years. However, these results are based on only 100 points extracted for a four-year time period, i.e. 2017–2020 which has 16 satellite images for the observed study area. The increase in collection of data would make it easier to explore the actual relationship and impact of the topographical, geological, lithological and soil type parameters on physico-chemical features of water.

CONCLUSION

In this study, physico-chemical features were calculated from Landsat 8 images and topographical, geological, lithological and soil type features were extracted with SRTM DEM data retrieved for the study area of Rawal watershed for the years 2017–2020. Four water quality parameters including dissolved oxygen, pH, turbidity and temperature, and five topographical, geological, lithological and soil type parameters consisting of slope, aspect, geology, lithology and soil type were taken as input to observe the impact of the geographical or environmental factors on the quality of water. For this, Pearson's correlation was applied on the features to observe any relationships amongst the parameters. The results observed with the correlations of topographical, geological, lithological and soil type with physico-chemical parameters in the monsoon seasons of June to September for the years 2017–2020 show the presence of a similar trend where the turbidity and temperature parameters had good positive correlations with soil type parameter, which is the proof of a relationship as turbidity is caused by the presence of elements such as clay. However, pH and dissolved oxygen physico-chemical parameters do not show any strong bond with topographical, geological, lithological and soil type parameters. Nevertheless, these results are based on the images retrieved for the monsoon seasons over four years i.e., 2017–2020 and the best results were only seen for the month of July. The selection of the Landsat images is primarily based on the availability of data for the chosen watershed area. As future work, more satellite imagery extracted from different satellite sensors can be analysed to present a better

picture of the relationship between these features. The application of machine learning techniques on such features will also be performed to predict the water quality of the study area.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

Abdullah, H. S., Mahdi, M. S. & Ibrahim, H. M. 2017 Water quality assessment models for Dokan Lake using Landsat 8 OLI satellite images. *Journal of Zankoy Sulaimani, Pure and Applied Sciences* **19–3** (4), 25–44.

Ali, M., Qamar, A. M. & Ali, B. 2013 Data analysis, discharge classifications, and predictions of hydrological parameters for the management of Rawal dam in Pakistan. In: *12th International Conference on Machine Learning and Applications*. Vol. 1. IEEE, pp. 382–385.

Avdan, U. & Jovanovska, G. 2016 Algorithm for automated mapping of land surface temperature using LANDSAT 8 satellite data. *Journal of Sensors* **2016**, 1–8.

Bajjali, W. 2018 Watershed delineation. In: *ArcGIS for Environmental and Water Issues*. Springer Textbooks in Earth Sciences, Geography and Environment. Springer, Cham. pp. 235–245.

Beeson, P. C., Sadeghi, A. M., Lang, M. W., Tomer, M. D. & Daughtry, C. S. 2014 Sediment delivery estimates in water quality models altered by resolution and source of topographic data. *Journal of Environmental Quality* **43** (1), 26–36.

Bhatti, N., Siyal, A. & Qureshi, A. 2018 Groundwater quality assessment using water quality index: a case study of Nagarparkar, Sindh, Pakistan. *Sindh University Research Journal-SURJ (Science Series)* **50** (2), 227–234.

Bonansea, M., Bazán, R., Ledesma, C., Rodriguez, C. & Pinotti, L. 2015 Monitoring of regional lake water clarity using Landsat imagery. *Hydrology Research* **46** (5), 661–670.

Chen, L., Tan, C.-H., Kao, S.-J. & Wang, T.-S. 2008 Improvement of remote monitoring on water quality in a subtropical reservoir by incorporating grammatical evolution with parallel genetic algorithms into satellite imagery. *Water Research* **42** (1–2), 296–306.

Dogliotti, A. I., Ruddick, K., Nechad, B., Doxaran, D. & Knaeps, E. 2015 A single algorithm to retrieve turbidity from remotely-sensed data in all coastal and estuarine waters. *Remote Sensing of Environment* **156**, 157–168.

Fichot, C. G., Downing, B. D., Bergamaschi, B. A., Windham-Myers, L., Marvin-DiPasquale, M., Thompson, D. R. & Gierach, M. M. 2016 High-resolution remote sensing of water quality in the San Francisco Bay–delta estuary. *Environmental Science & Technology* **50** (2), 573–583.

Hamzaoui-Azaza, F., Ketata, M., Bouhlila, R., Gueddari, M. & Riberio, L. 2011 Hydrogeochemical characteristics and assessment of drinking water quality in Zeuss–Koutine aquifer, Southeastern Tunisia. *Environmental Monitoring and Assessment* **174** (1–4), 283–298.

Hsu, H. H., Chen, L., Kou, C. H., Yeh, H. C. & Wang, T. S. 2009 Applying multi-temporal satellite imageries to estimate Chlorophyll-a concentration in Feitsui reservoir using ANNs. In: *International Joint Conference on Artificial Intelligence*. IEEE, pp. 345–348.

Issaka, S. & Ashraf, M. A. 2017 Impact of soil erosion and degradation on water quality: a review. *Geology, Ecology, and Landscapes* **1** (1), 1–11.

Jian-hua, L. I. U. 2011 Discussion on the method of making slope classification map with DEM in ArcGIS. *Geomatics & Spatial Information Technology* **1**.

Kapalanga, T. S. 2015 *Assessment and Development of Remote Sensing Based Algorithms for Water Quality Monitoring in Olushandja Dam, north-central Namibia*. PhD Thesis, MSc Thesis, University of Zimbabwe.

Khalil, M. T., Saad, A., Ahmed, M., El Kafrawy, S. B. & Emam, W. W. 2016 Integrated field study, remote sensing and GIS approach for assessing and monitoring some chemical water quality parameters in Bardawil lagoon, Egypt. *International Journal of Innovative Research in Science Engineering and Technology* **5** (8), 10–15680.

Khan, H. Q. 2010 Water quality index for municipal water supply of Attock city, Punjab, Pakistan. In: Gökçekus, H., Türker, U. and LaMoreaux, J. W. (eds) *Survival and Sustainability*. Springer, Berlin, Heidelberg, pp. 1255–1262.

Khattab, M. F. & Merkel, B. J. 2014 Application of Landsat 5 and Landsat 7 images data for water quality mapping in Mosul dam lake, northern Iraq. *Arabian Journal of Geosciences* **7** (9), 3557–3573.

Lim, J. & Choi, M. 2015 Assessment of water quality based on Landsat 8 operational land imager associated with human activities in Korea. *Environmental Monitoring and Assessment* **187** (6), 1–17.

Liu, H., Xu, M. & Beck, R. 2018 An ensemble approach to retrieving water quality parameters from multispectral satellite imagery. In: *IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, pp. 9284–9287.

Magesh, N. S., Jitheshlal, K. V., Chandrasekar, N. & Jini, K. V. 2013 Geographical information system-based morphometric analysis of Bharathapuzha river basin, Kerala, India. *Applied Water Science* **3** (2), 467–477.

Markogianni, V., Kalivas, D., Petropoulos, G. & Dimitriou, E. 2017 Analysis on the feasibility of Landsat 8 imagery for water quality parameters assessment in an oligotrophic Mediterranean lake. *International Journal of Geological and Environmental Engineering* **11** (9), 906–914.

Nazeer, S., Hashmi, M. Z. & Malik, R. N. 2014 Heavy metals distribution, risk assessment and water quality characterization by water quality index of the river Soan, Pakistan. *Ecological Indicators* **43**, 262–270.

Oyedotun, T. D. T. 2019 Land use change and classification in Chaohu Lake catchment from multi-temporal remotely sensed images. *Geology, Ecology, and Landscapes* **3** (1), 37–45.

Oyedotun, T. D. T. 2020 Quantitative assessment of the drainage morphometric characteristics of Chaohu lake basin from SRTM DEM data: a GIS-based approach. *Geology, Ecology, and Landscapes*. DOI: 10.1080/24749508.2020.1812147.

Pazand, K. & Hezarkhani, A. 2012 Investigation of hydrochemical characteristics of groundwater in the Bukan basin, northwest of Iran. *Applied Water Science* **2** (4), 309–315.

Qadir, A., Malik, R. N. & Husain, S. Z. 2008 Spatio-temporal variations in water quality of nullah Aik-tributary of the River Chenab, Pakistan. *Environmental Monitoring and Assessment* **140** (1–3), 43–59.

Sharaf El Din, E. 2020 A novel approach for surface water quality modelling based on Landsat-8 tasselled cap transformation. *International Journal of Remote Sensing* **41** (18), 7186–7201.

Theologou, I., Patelaki, M. & Karantzalos, K. 2015 Can single empirical algorithms accurately predict inland shallow water quality status from high resolution, multi-sensor, multi-temporal satellite data? *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* **40** (7), 1511.

Wen, X. P. & Yang, X. F. 2010 Monitoring of water quality using remote sensing techniques. In: Honghua Tan (ed.) Editor: Honghua Tan City: Baech, Switzerland *Applied Mechanics and Materials*. Vol. 29. Trans Tech Publications Ltd, Baech, Switzerland, pp. 2360–2364.