Journal of Hydroinformatics

Journal of Hydroinformatics Vol 25 No 2, 432 doi: 10.2166/hydro.2023.137

Quality monitoring of inland water bodies using Google Earth Engine

P. Y. Sherjah [a,b,*], N. Sajikumar[a,c] and P. T. Nowshaja[a]

[a] Department of Civil Engineering, Government Engineering College, Thrissur, Kerala, India
[b] A. P. J. Abdul Kalam Technological University, Thiruvananthapuram, Kerala, India
[c] WRPM Consultants, Thrissur, Kerala 80010, India
*Corresponding author. E-mail: sherjah@gmail.com

PYS, 0000-0002-1006-8530

ABSTRACT

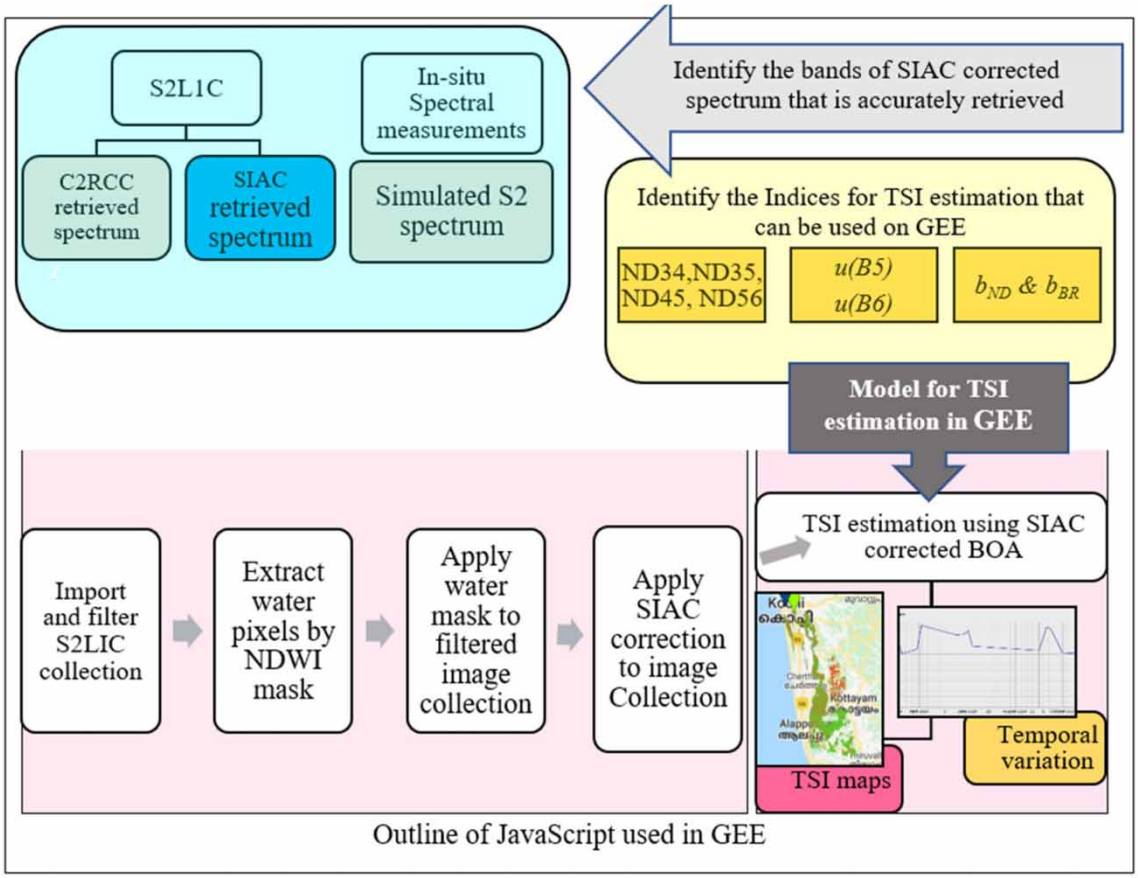
Regular quality monitoring of inland water bodies is vital for identifying the areas with deteriorating water quality. Satellite remote sensing has been used for obtaining long-term trends that require the processing of many images. The computational load of processing a large number of satellite imageries can be eased by utilizing the cloud computing capabilities of Google Earth Engine (GEE). The present study explores the possibility of using the GEE platform for mapping the Trophic State Index (TSI) of an inland water body. The bottom of atmosphere (BOA) reflectance retrieved by the SIAC algorithm (used in the GEE platform) is assessed for its accuracy. The algorithm could retrieve only BOA reflectance at bands B3 and B4 of Sentinel 2L1C (S2) with reasonable accuracy. The study has identified the Normalized Difference of B3 and B4 bands of S2 (i.e., ND34) as the tool for mapping TSI of a water body using GEE. TSI from six imageries of three lakes was estimated with a mean error $<17\%$. The capability of GEE as a rapid water quality monitoring tool is demonstrated by displaying the temporal and spatial variations of water quality across Vembanad Lake for the period 2016–2021.

Key words: GEE, Sentinel 2, TSI estimation, water quality monitoring

HIGHLIGHTS

- Model for TSI estimation feasible for GEE identified.
- Normalized difference of B3 and B4 bands of Sentinel 2 feasible for GEE.
- Long-term temporal variations and the synoptic view of water quality obtained in less time and effort.

GRAPHICAL ABSTRACT



INTRODUCTION

The increase in population and man's negligence in incorporating the principle of sustainability in technological developments has led to uncontrolled urbanization, which is taking its toll on the environment (Owa 2013; Dörnhöfer & Oppelt 2016). Regular monitoring of inland water bodies is necessary for the identification of the areas of deteriorating water quality so that corrective measures can be adopted before the impairment becomes irreversible (Blondeau-Patissier *et al.* 2014; Barrett & Frazier 2016). There are various parameters based on which quality of a water body is assessed, namely, dissolved oxygen, salinity, pH, temperature, turbidity, Total Nitrogen (TN), Total Phosphorus (TP) and chlorophyll-*a* (Chl-*a*). Water quality indices are simple indices used for classifying the quality of a water body based on these parameters (Tyagi *et al.* 2013). Trophic State Index (TSI), (Carlson 1977) is a widely used water quality index for classifying water quality status (Papoutsa *et al.* 2014; Shi *et al.* 2019; Zhu & Mao 2021). It is computed from log-transformed values of Secchi depth (SD), TN, TP and/or Chl-*a*.

Remote sensing has been of interest to researchers and environmentalists owing to its ability to get a synoptic view and almost real-time source of information. Water quality assessment using remote sensing is carried out in terms of optically active constituents, namely, Chl-*a*, suspended sediments and coloured dissolved organic matter (CDOM). There are several algorithms developed for quantifying water quality from remotely sensed data, but most of them were developed by using *in situ* spectral measurements and used the spectral bands available in ocean monitoring satellites like MODIS (e.g., 412, 620, 622, 730, 700 nm) (Dörnhöfer & Oppelt 2016). These bands are not available in the land monitoring satellites like Sentinel 2 and Landsat 8 and the coarse spatial resolution of ocean monitoring satellites is insufficient to capture the variations in inland water bodies. A recent study has identified the potential of Normalized Difference (ND) indices of green and red bands (B3, B4, B5 and B6) of Sentinel 2, single scattering albedo (u) and backscattering coefficient (b) for estimating TSI(SD) (i.e. TSI computed from SD) (Sherjah *et al.* 2022a, 2022b). Bottom of atmosphere (BOA) reflectance retrieved by the C2RCC (Case 2

Regional Coast Colour) atmospheric correction processor (ACP) was used for deriving these indices. The good correlation between SD and turbidity (Effler 1989) can be utilized for quantifying TSI(SD) from turbidity readings as SD measurements are prone to error (difficulty of taking measurements in water bodies with undercurrents and the readings can vary with the visibility of the observer). Moreover, the causes of deteriorating water quality (like suspended matter and algae blooms) affect the clarity of the water (Moore 1980). Turbidity readings have also been endorsed as a proxy indicator for cyanobacteria and water quality in general (Beck *et al.* 2019).

Cloud computing platforms ease the computational load of downloading, processing and analysing large datasets. Google Earth Engine (GEE) is a cloud-based computing platform for geospatial analysis using an online repository of satellite imagery. It utilizes Google's computational capabilities for monitoring societal issues like deforestation, climate change, shoreline changes, water management, etc. This platform saves the time and effort of downloading and pre-processing satellite images. The capabilities of GEE have been utilized for algae bloom mapping (Cardall *et al.* 2021; Lobo *et al.* 2021), forest area monitoring (Yang *et al.* 2019), land cover monitoring and flood mapping (Amani *et al.* 2020). However, their utilization for water quality monitoring studies is limited. The major limitation in utilizing the capabilities of GEE for water quality monitoring studies is the atmospheric inference that amounts to 80–90% of the total received signal sensed from the water bodies. The accuracy of any algorithm developed for water bodies is greatly influenced by the retrieval accuracy of the ACP (Sherjah *et al.* 2022b). The ACP available on the GEE platform is SIAC (Sensor Invariant Atmospheric Correction). The SIAC-corrected surface reflectance from land pixels of Sentinel 2 was validated (Yin *et al.* 2019), but studies on the accuracy of the algorithm in retrieving the surface reflectance of water pixels are limited. Hence, the atmospheric correction accuracy of the SIAC algorithm on water bodies needs to be assessed. The prospect of using GEE for obtaining a quick analysis of the spatial and temporal variations of TSI needs to be explored.

The objective of the present study is to explore the feasibility of using the models developed for TSI estimation on the GEE platform. The accuracy of the BOA reflectance spectrum retrieved from Sentinel 2 L1C imagery by the SIAC algorithm is to be assessed so that the bands that can be used on the GEE platform can be identified.

MATERIALS AND METHODS

Study area

The area chosen for the study is Vembanad Lake located in Kerala, India (Figure 1(a)). The lake is 96 km long, the largest in the state of Kerala. The wetland system is the largest in India (2,033.02 km^2) and is one of the three protected wetlands by the Ramsar convention, 2002. A 1,252 m long mud regulator (Thanneermukkom saltwater barrier) is built across the lake to prevent saltwater intrusion into the Kuttanad paddy fields. Vembanad has ecological importance by harbouring thousands of waterfowl. The major sources of revenue in the area are farming, lime shell collection, fishing, coir retting and tourism.

The mining of shells from the lake bed in an uncontrolled manner is posing a threat to the ecosystem. The past few decades have witnessed heavy disposal of domestic and industrial sewage into the lake making it highly polluted. The estuary and contributing rivers have been polluted with anthropogenic loadings (Viswanathan 2019). The untreated sewage released into the lake has reduced the dissolved oxygen content in the water body. Many studies were conducted on the quality status of the lake, all of which reveal the disastrous situation of this ecological hotspot. The substratum condition which is ideal for the growth of shrimp is impacted by the pollution deposits in the lake (Sruthy & Ramasamy 2017). The sediments of the lake were found to be polluted by toxic heavy metals like Zinc, Cadmium, Nickel and Cobalt (Priju & Narayana 2007). The need for regularly monitoring the trophic status of the lake is the reason for choosing the lake for this study. *In situ* measurements were taken from Kochi Kayal located on the north of the lake.

The models for TSI estimation that are being tested in GEE were developed from the atmospherically corrected (by C2RCC) Sentinel 2 LIC imagery and turbidity data of Milford Lake, KS, USA (Sherjah *et al.* 2022a, 2022b). Milford Lake is a man-made reservoir with a surface area of 64 km^2 and a catchment area of 6,200 km^2 (Figure 1(b)). Agricultural land (51%) and grassland (36%) are the major land use classes of the catchment. The excess fertilizer application from these areas and the effluents from livestock production units were the major causes of the water quality deterioration of the lake (Leiker *et al.* 2019).

The model/models referred to in the previous paragraph that can be applied to GEE were validated with six sets of data (imagery and water quality) of Milford Lake, Conroe Lake (Figure 1(c)) and Vembanad Lake. Conroe Lake is located in Montgomery County, Texas, United States and has a surface area of 85 km^2. The lake runs through classic East Texas Piney

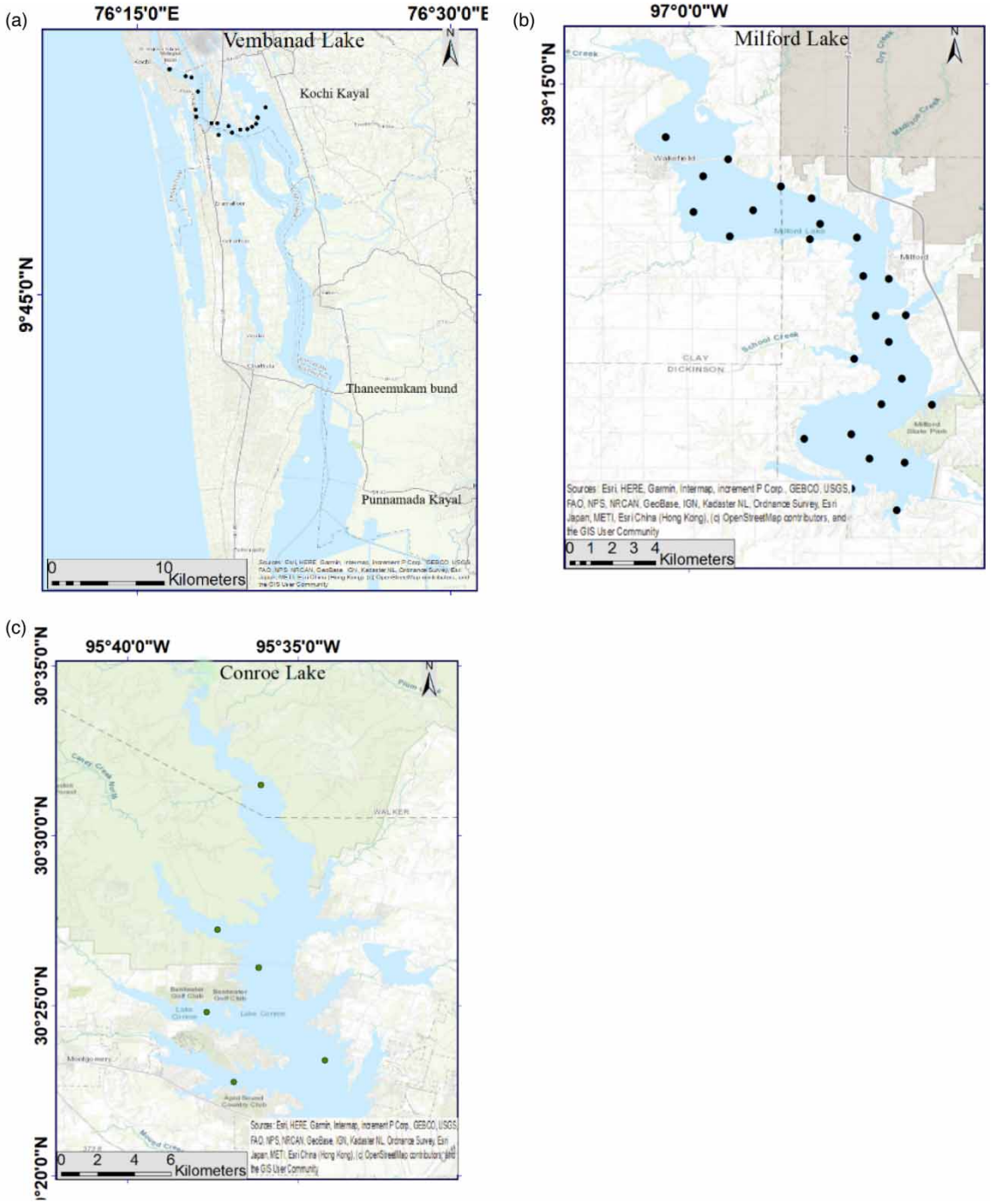


Figure 1 | Location maps of the study area.

Woods forests. Water quality is generally good, with an average depth of 6.2 m, and a maximum depth of 23 m. The details of these lakes with the dates of sample collection, satellite sensing date, No. of sample points and coordinates are given in Table 1.

Satellite data

Sentinel 2 Level 1C (S2) imagery was used for the study. The imagery has 13 radiometrically and geometrically corrected Top of Atmosphere (TOA) reflectance bands (4 in the visible region, 3 in the red edge, 3 in the near-infrared (NIR) region and 3 in

Table 1 | Details of the study area with no. of sample points, the date of sample collection and date of satellite sensing

Study area	Surface area of lake (km^2)	No. of sample points	Date of sample collection	Sensing date of imagery
Vembanad Lake, Kerala, India	2,033	22	26 Jan 2021	28 Jan 2021
Milford Lake, KS, USA	64	26	10 Jul 2018	09 Jul 2018
		26	09 Aug 2018	03 Aug 2018
		13	16 Oct 2018	22 Oct 2018
		13	17 Oct 2018	
Conroe lake, TX, USA	84	7	04 Feb 2020	06 Feb 2020
		6	26 Sep 2021	21 Sep 2021

short wave Infrared (SWIR). The spatial resolution of the bands in the visible region, B2–B4 and the NIR band B8 is 10 m. B5, B6 and B7 bands have 20 m and B1, B9 and B10 have 60 m spatial resolutions. These images are available with a high temporal resolution of 5 days (2–3 days for mid-latitudes). They are freely downloadable at Copernicus Open Access Hub (https://scihub.copernicus.eu. last accessed on 31/12/2021).

The pre-processing steps of the imagery were carried out by using SeNtinel Application Platform (SNAP) software (version 8.0.2) developed by the European Space Agency (ESA). There are several open-source toolboxes and bundled packages for processing MERIS, Sentinel and Landsat images in SNAP. The resampling of the image to 10 m and the masking of the area of interest from the downloaded images were carried out. The raster computations were carried out using the 'Band Maths' tool. The images were atmospherically corrected by the C2RCC processor plug-in available in the SNAP toolbox. The software ArcGIS 10.1 is used for the presentation of maps.

Google Earth Engine

GEE is a cloud computing platform, launched in 2010, for assisting in geospatial data analysis. It has a massive repository of publicly available datasets of various satellites and aerial imaging systems, weather and climate forecasts, land cover maps and socio-economic datasets. It has a parallel image processing platform which distributes the computation tasks across high-performance petabyte CPUs. This enables the processing of hundreds of images within a few seconds. It also simplifies the entire procedure of analysis as the satellite images need not be downloaded for their processing or analysis. Users can utilize the data libraries and computing facilities of GEE via Earth Engine programs in JavaScript or Python scripts.

Water quality data

Water quality data of the lakes in the US were downloaded from the USGS National Water Information System (https://nwis.waterdata.usgs.gov; last accessed on 31/12/2021). The models used in the study were developed from the S2 imagery and the *in situ* measurements of Milford Lake (Sherjah *et al.* 2022a, 2022b).

The turbidity measurements and the spectral measurements from Kochi Kayal, Kerala, were taken on 26th January 2021. Water samples from 22 points of Kochi Kayal (Figure 1(a)) were collected and stored in ice bags as per standard protocol. Turbidity of the water samples was measured using Eutech ECTN-100IR. The above-surface spectral reflectance measurements from the sample locations were taken using an SVC HR 512i spectroradiometer. The instrument has a spectral range from 350 to 1,050 nm and a spectral resolution of less than 1.5 nm. The spectrum is obtained in radiance, as well as reflectance units. The reflectance spectrum was used in the study for assessing the accuracy of retrieval of the ACP, SIAC. Spectral measurements were taken between 10.30 a.m. and 12.00 p.m. and then from 2.00 to 3.30 pm on the day of measurement. The measurements were repeated three to four times at each point and the averaged spectrum at each point was then convoluted to simulated S2 bands by the following equation.

$$\mathrm{R_{rs}(B}_i) = \frac{\int_{\lambda_{\min}}^{\lambda_{\max}} SRF(\lambda) \cdot R_{rs}(\lambda) d\lambda}{\int_{\lambda_{\min}}^{\lambda_{\max}} SRF(\lambda) d\lambda}, \quad (1)$$

where R_{rs} (B_{i}) is the simulated band reflectance of $\mathrm{B_i^{th}}$ band of S2, $\lambda_{\min}$ and $\lambda_{\max}$ are the starting and ending wavelengths of the

band B_i, R_{rs} is the *in situ* measured reflectance at the wavelength λ, SRF is the spectral response function of S2 (https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/document-library/-/asset_publisher/Wk0TKajiISaR/content/sentinel-2a-spectral-responses) (last accessed on 15 April 2021).

The atmospherically corrected BOA spectrum obtained from SIAC was compared with the simulated spectrum and that from C2RCC to assess its retrieval efficiency.

METHODS

Atmospheric correction processor

The models used in this study were developed from S2 data whose atmospheric correction was carried out by C2RCC. Case 2 Regional Coast Colour processor (C2RCC) is an ACP developed for water bodies by Doerffer & Schiller (2007). The BOA reflectance from the satellite imagery is derived by using Artificial Neural Networks (ANN) based on a large database of around 5 million cases of simulated water-leaving radiances and related TOA radiances (generated from *in situ* measurements obtained from the NOMAD database, Coastcolour database and simulated results from the radiative transfer model, HYDROLIGHT). The processor extracts the water pixels from the imagery and retrieves the water-leaving reflectance in 6 bands (B1–B6).

The atmospheric correction tool available on GEE is SIAC. This algorithm is based on the inference that changes in the reflectance from land surfaces are insignificant when the images are a few days apart. Thus, any change observed in the reflectance in these images is caused by the change in the atmospheric composition. SIAC computes the atmospheric inference (i.e., aerosol optical thickness (AOT) and total columnar water vapour (TCWV)) based on Bayesian decision theory using the MODIS MCD43 BRDF descriptor product and the Copernicus Atmosphere Monitoring Service (CAMS) operational forecasts of AOT and TCWV. MODIS MCD43 BRDF (Bi-directional Reflectance Distribution Function) descriptor product provides the surface directional reflectance (SDR) at the coarse resolution of MODIS. The simulated band measurements from MODIS bands are used to predict the reflectance at the other sensors. A Gaussian point spread function (PSF) is inferred based on the correlation between the TOA reflectance and BOA reflectance. A model based on MODIS PSF is built to deal with the scale differences between MODIS and Sentinel 2/Landsat 8. Linear spectral mappings are used to convert the reflectance of MODIS bands to the spectral configuration of the target sensor (i.e., Sentinel or Landsat 8) (Yin *et al.* 2019). BOA reflectance is derived from TOA reflectance using the atmospheric parameters obtained from the 6S model. Table 2 gives the bands of MODIS and Sentinel 2 that share similar spectral coverage. The spatial resolution of MODIS for red (B1) and NIR (B2) bands is 250 m and that for all other bands is 500 m. There are no bands in MODIS that have similar spectral coverage as bands B1, B5 and B6 of Sentinel 2.

Indices for trophic status

TSI (Carlson 1977; Kratzer & Brezonik 1982) was used in this study for quantifying water quality. TSI is considered as a reliable tool for quantifying trophic status by the ecological community and the environmental monitoring centres like United States Environmental Protection Agency (United States Environmental Protection Agency 2007), CETESB (Environmental Company of Sao Paulo State) (Martins *et al.* 2020; Lobo *et al.* 2021) and National Environmental Monitoring Center

Table 2 | Bands and the spectral coverage of MODIS and Sentinel 2

MODIS		Sentinel 2	
Band No.	**Wavelength (nm)**	**Band No.**	**Wavelength (nm)**
3	459–479	2	457–523
4	545–565	3	542–578
1	620–670	4	650–680
2	841–876	8	784–900
2	841–876	8A	855–875
6	1,628–1,652	11	1,565–1,655
7	2,105–2,155	12	2,100–2,280

(China) (Martins *et al.* 2020; Zhu & Mao 2021). 'Water Atlas', the program created by the University of South Florida, uses TSI to provide information about water quality of more than 7,700 water bodies to the stakeholders in water resources management. These indices can be computed from log-transformed values of SD, TN, TP and Chl-*a*. The objective of the present study is to assess the possibility of using the models developed for TSI estimation (Sherjah *et al.* 2022b, 2022a) in the GEE platform. It was observed that TSI(SD) had the best correlation to *in situ* measurements in the study and hence, TSI is computed in this study from SD values by using Equation (2). Though the index can take infinite values, the scale is limited to 100, with SD of 64 m as 0, and 0.06 m as 100. TSI value <60 is considered a 'good' quality lake that meets all lake use criteria (swimmable, fishable and supports healthy habitat) (University of South Florida 2018).

Turbidity readings were used to derive SD based on the well-established relation between turbidity and SD (Equation (3)) (Myre & Shaw 2006).

$$\mathrm{TSI} = 10\left(6 - \frac{\ln(\mathrm{SD})}{\ln(2)}\right), \tag{2}$$

$$\mathrm{SD} = 244.13 * \mathrm{turbidity}^{-0.662}, \tag{3}$$

where SD is in cm and turbidity in FNU.

The trophic conditions of water bodies are classified into four classes as mentioned in Table 3.

Methodology adopted for the study

The objective of the paper is to analyse the feasibility of using the models developed for TSI estimation on the GEE platform. The methodology adopted is briefed in Figure 2 and is represented by steps as follows.

1. Identify the bands that are accurately retrieved by the SIAC atmospheric correction algorithm by comparing the SIAC-generated spectrum with that of C2RCC and the simulated spectrum.
2. Based on the accuracy of retrieved bands, identify the indices and the model that are feasible for GEE when the SIAC algorithm is used for atmospheric correction.
3. Develop the JavaScript for processing the imagery and computing TSI using the identified model on GEE.

Models for TSI estimation

The models that are tested in this study were derived from the atmospherically corrected S2 imagery and turbidity data of Milford Lake (Figure 3).

Table 3 | Classification of TSI (Carlson 1977; Kratzer & Brezonik 1982)

TSI interval	Classification	SD (m)
0–10	Oligotrophic	64–32
11–20		31.9–16
21–30		15.9–8
31–40		7.9–4
41–50	Mesotrophic	3.9–2
51–60	Eutrophic	1.9–1
61–70	Hypereutrophic	0.9–0.5
71–80		0.49–0.25
81–90		0.24–0.12
91–100		0.11–0.062
100 <		0.062 >

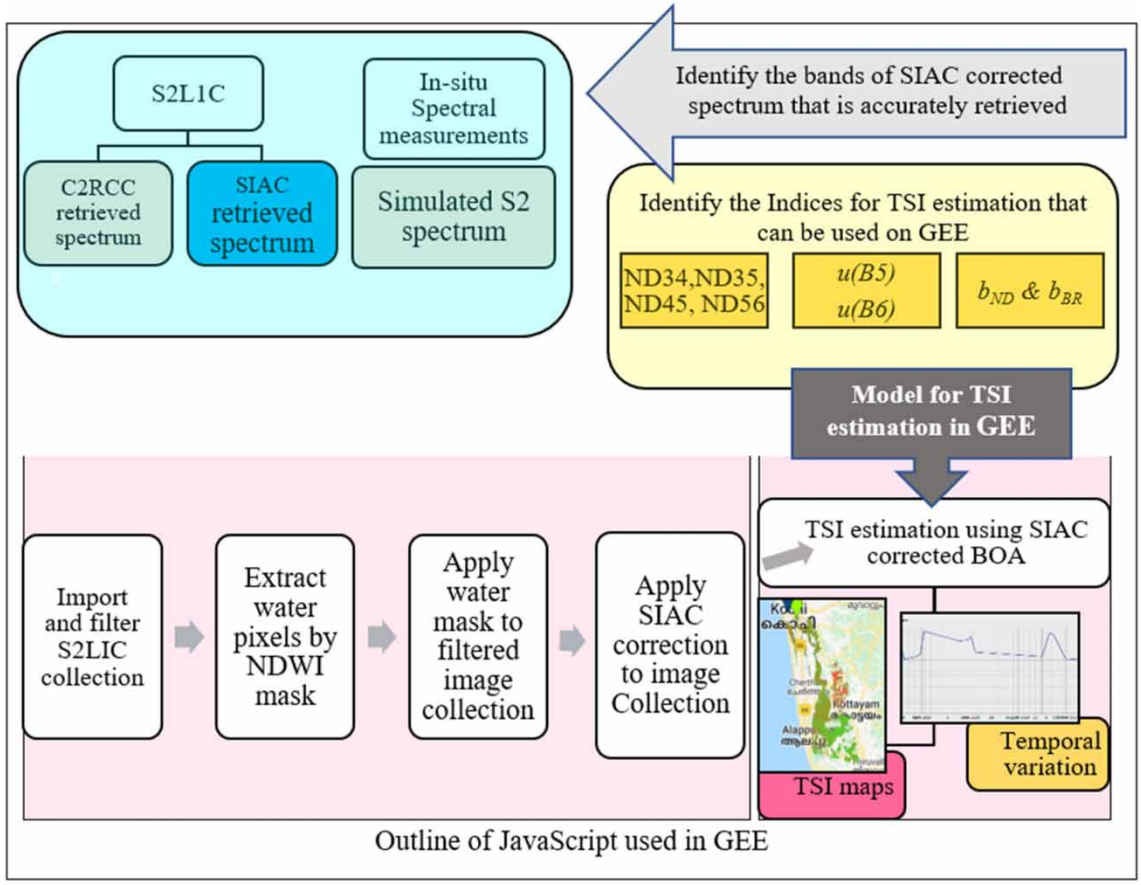


Figure 2 | Methodology in brief.

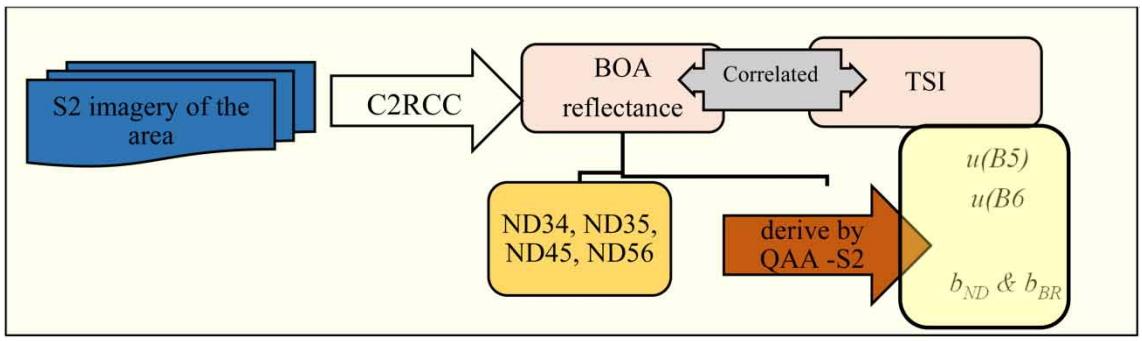


Figure 3 | Methodology adopted for deriving the models used in the study.

ND indices are represented in this paper as NDxy and computed using the following equation.

$$\mathrm{ND}xy = \frac{R_{\mathrm{rs}}(x) - R_{\mathrm{rs}}(y)}{R_{\mathrm{rs}}(x) + R_{\mathrm{rs}}(y)} \tag{4}$$

where $R_{\mathrm{rs}}(x)$ and R_{rs} (y) are the BOA reflectance in the bands x and y, respectively.

QAA-S2 (Sherjah *et al.* 2022b) is the modified version of the Quasi Analytical Algorithm (QAA v6) (Lee *et al.* 2002) parametrized for S2 bands. The QAA-S2 algorithm derives the Inherent Optical Properties (IOP), namely, single scattering albedo

(u) and absorption coefficient (a) and backscattering coefficient (b). The steps of QAA-S2 for deriving IOP from the above-surface reflectance (R_{rs}) are briefed below (Equations (5)–(10)). u(B5), u(B6) and ND of b were identified as good estimators of TSI (Sherjah *et al.* 2022b) and hence are considered in this study.

$$r_{rs}(\lambda) = \frac{R_{rs}(\lambda)}{0.52 + 1.7R_{rs}(\lambda)} \tag{5}$$

where $R_{rs}(\lambda)$ is the above-surface reflectance at any wavelength λ and $r_{rs}(\lambda)$ is the below-surface reflectance at any wavelength λ

$$u(\lambda) = \frac{-g_0 \pm \sqrt{g_0^2 - 4g_1 * r_{rs}(\lambda)}}{2g_1} \tag{6}$$

where $u(\lambda)$ is the single scattering albedo, $g_o = 0.0895$ and $g_1 = 0.125$ (Lee *et al.* 2002)

$$u(\lambda) = \frac{b(\lambda)}{a(\lambda) + b(\lambda)} \tag{7}$$

where $a(\lambda) = a_w(\lambda) + a_p(\lambda)$, $b(\lambda) = b_w(\lambda) + b_p(\lambda)$, $a(\lambda)$ = total absorption coefficient at any wavelength λ, $b(\lambda)$ = total backscattering coefficient at any wavelength λ, and subscript 'w' for water, 'p' for particulate.

$$a(\lambda_0) = a(670) = a_w(670) + 0.39\frac{R_{rs}(670)}{R_{rs}(443) + R_{rs}(490)} \tag{8}$$

where $R_{rs}(670) = R_{rs}(B4)$, $R_{rs}(443) = R_{rs}(B1)$, $R_{rs}(490) = R_{rs}(B2)$

$$b(\lambda) = b_w(\lambda) + b_p(\lambda o)\left(\frac{\lambda o}{\lambda}\right)^{\eta} \tag{9}$$

$$\eta = 2(1 - 1.2 * \exp\left(-0.9\frac{r_{rs}(\lambda_1)}{r_{rs}(\lambda_2)}\right) \tag{10}$$

where $\lambda_1 = 665$, i.e., $B4$; $\lambda_2 = 740$, i.e., $B6$.

The indices capable of estimating TSI from R_{rs} are as follows:

a. The ND indices of BOA reflectance between bands B3, B4, B5 and B6, namely, ND34, ND35, ND45 and ND56 (Sherjah *et al.* 2022a).
b. Single scattering albedo derived by QAA-S2 (Sherjah *et al.* 2022b).
c. ND indices of backscattering coefficients derived by QAA-S2 (Sherjah *et al.* 2022b).

Javascript for TSI estimation

The processing steps for deriving TSI maps from satellite imagery are briefed below. The JavaScript is shown in the Appendix included as supplementary data.

1. The S2LIC collection is imported from the repository and the dataset is filtered based on dates, cloud cover percentage & region of interest. The filtered dataset is named '*S2_col*'.
2. The water area is extracted by using an NDWI mask. NDWI is computed by Equation (11) for a whole year and the temporal variation charts for selected points on the lake closer to the banks are drawn to identify the period with mean water area, i.e., the months where the water body is neither inundated nor dried up. The NDWI > 0 is the water area, and the NDWI mask '*S2_mask*' is computed for a cloud-free day in this period. For the Vembanad area, the month of May is used

for computing the NDWI mask.

$$NDWI = \frac{R_{rs}(B3) - R_{rs}(B8)}{R_{rs}(B3) + R_{rs}(B8)} \tag{11}$$

3. The S2L1C filtered collection is clipped by this NDWI mask to obtain only the water area. The collection is further named '*S2_col_mask*'.
4. The SIAC correction module is imported and applied to each image in collection 'S2_col_mask' to get the output collection 'S2_boa', which has 12 bands of the bottom of atmosphere reflectance.
5. The obtained BOA reflectance is used for computing TSI.
6. The date of the imagery whose TSI map is to be visualized is chosen and a date stamp function is used to identify the imagery.
7. The time series of TSI is computed for the points of interest and period of interest.
8. The TSI maps and the corrected imagery are exported to drive for future reference and further analysis, if necessary.

Accuracy assessment

The Pearson correlation coefficient (r) was used to assess the degree of correlation and is computed by the following equation

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{12}$$

where x and y are variables, whose correlation is studied.

The accuracy of estimation using the model was assessed by Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE)

$$\text{MAE} = \frac{\sum_{i=1}^{N}|y_{m_i} - y_{p_i}|}{N} \tag{13}$$

$$\text{MAPE} = \frac{\sum_{i=1}^{N}\left|\frac{y_{m_i} - y_{p_i}}{y_{m_i}}\right|}{N} * 100, \tag{14}$$

where y_m and y_p are measured and estimated variables.

RESULTS AND DISCUSSIONS

Accuracy assessment of the spectrum retrieved by the SIAC algorithm

One of the objectives of the study is to assess the accuracy of BOA reflectance (R_{rs}) spectra retrieved from the water bodies by the SIAC algorithm. The R_{rs} values for the bands B1–B6 of S2 (at a point) are termed (here) as the 'spectrum' at that point. C2RCC has been established to be the suitable ACP for water bodies in the previous studies (Pereira-Sandoval *et al.* 2019; Warren *et al.* 2019; Sherjah *et al.* 2022a) based on its better performance when compared to other ACPs. However, the C2RCC processor was not available in the GEE platform. Only the SIAC algorithm is available on the GEE platform and hence, the spectra retrieved by SIAC are compared with the C2RCC retrieved spectra to identify the bands at which R_{rs} agree with each other. The SIAC corrected spectra (of the sample points) retrieved from the imagery of Kochi Kayal (28th January 2021) and the C2RCC corrected spectra (of the same sample points) of the same imagery are shown in Figure 4(a) and 4(b), respectively. The simulated spectrum for each sample point (Figure 4(c)) is generated from the *in situ* spectral measurements for the bands of S2 by using Equation (1). The simulated spectrum represents the spectrum that can be obtained by an ACP, if it retrieves R_{rs} all the bands with 100% accuracy. The real atmospheric composition and its

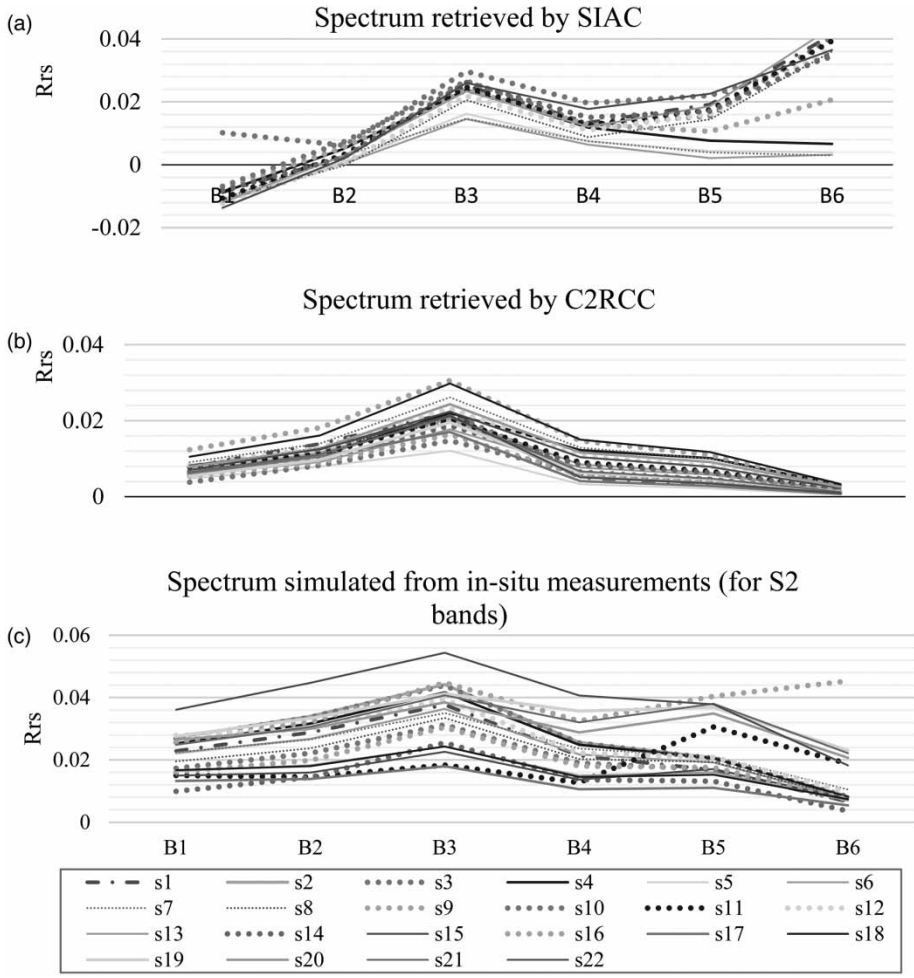


Figure 4 | BOA spectrum of S2 imagery of Kochi obtained from SIAC processor, C2RCC processor and the simulated spectrum from *in situ* spectral measurements.

contribution to the TOA reflectance at each band is too complex to be accurately represented by a model. The deviations of the spectra retrieved by the ACP (from the simulated spectra) help to assess the efficiency of the ACP.

The accuracy of R_{rs} retrieved by an ACP for a band depends on the algorithm it uses. The ACPs developed for land may not perform well on water pixels (Pereira-Sandoval *et al.* 2019; Sherjah *et al.* 2022b). SIAC computes the atmospheric interference based on the inference that the changes in the TOA reflectance from land are insignificant when the images are a few days apart and any change observed in the TOA reflectance in these images is caused by the change in the atmospheric composition. The atmospheric contributions (scattering and absorption effect of aerosols) are subtracted from the TOA reflectance to retrieve R_{rs} from the target. It is evident from Figure 4(a) and 4(b) that only R_{rs} at B3 and B4 retrieved by SIAC match with that of C2RCC. R_{rs} at B1 and B2 are underestimated (sometimes even negative), and at B5 and B6 are overestimated for most points. The assumption (of SIAC) that any change observed in the TOA reflectance is solely the contribution of the atmosphere cannot be applied to the atmospheric correction of water bodies. However, this assumption may not be valid for water bodies as their compositions can vary within the temporal resolution of the imagery. Also, the SIAC algorithm is based on the MODIS BRDF product that has SRF at the bands of MODIS. The reflectance at these bands was then convoluted to that at the bands of S2 by spectral mapping. There are no bands in MODIS corresponding to the spectral ranges of the bands B1, B5 and B6 of S2 (Table 2). These may be the reasons for the inaccuracy in the retrieved spectra by SIAC. Even though R_{rs} for B3 and B4 retrieved by SIAC are less than that of the simulated spectra, those values are comparable with the values retrieved from the C2RCC processor.

Hence, from the comparison, it can be understood that only the models /indices with B3 and B4 bands can be considered for estimating the TSI on the GEE platform.

SIAC corrected R_{rs} in the models

The models developed for TSI estimation were based on the C2RCC corrected datasets of S2 imagery and the TSI measurements of Milford Lake for the three days shown in Table 1. Hence it is necessary that the SIAC corrected R_{rs} match with that from the C2RCC processor to give accurate TSI estimation using these models. It is evident from the previous section that only R_{rs} of B3 and B4 retrieved by SIAC come close to that obtained from the C2RCC processor. Hence the ND with R_{rs} at B3 and B4 (ND34) has a good probability of giving sound results with the SIAC dataset. Regarding the indices developed from IOP, the QAA-S2 algorithm requires the R_{rs} at B1, B2 and B6 for computing $a(\lambda o)$ (Equation (8)), and R_{rs} at B4 and B6 for computing η (Equation (10)). The inaccuracy in the retrieval of B1, B2 and B6 bands will affect the retrieval of IOPs using the QAA-S2 algorithm. Hence, out of the indices for TSI estimation, only ND34 has the potential to be used for the estimation of TSI in the GEE platform.

TSI can be estimated using ND34 by the following equation:

$$\text{TSI} = 77.15 - 38.22 * \text{ND34} \quad \text{(C2RCC model)} \tag{15}$$

when TSI was estimated from six images by using this equation, MAE was <10 and MAPE $< 17\%$ (Figure 5).

A good correlation ($r > 0.8$) was also observed between R_{rs} at B3 and B4 of SIAC-corrected S2 imagery of Milford lake and the TSI values. An alternate model (Equation (16)) was also developed with the SIAC corrected BOA dataset of Milford lake and the TSI values.

$$\text{TSI} = 73.96 - 89.10 * \text{ND34} \quad \text{(SIAC model)} \tag{16}$$

This equation was also used to estimate TSI from the other three images (MAE shown in Figure 5). However, the model based on the C2RCC dataset performed better than the SIAC dataset and hence, the former may be approved as the model to be used for TSI estimation on the GEE platform.

Temporal and spatial variations of TSI

Regular monitoring of the spatial and temporal variations in the quality status of a water body is essential to identify any deterioration in its condition before the situation becomes grave and irreversible. From the previous sections, the accuracy of the estimation of TSI using R_{rs} retrieved by SIAC was found to be acceptable. The spatial variation in TSI can be visualized by applying the model (Equation (15)) to atmospherically corrected S2 image of the lake. Numerous images have to be processed for obtaining the temporal variations of TSI of the lake. Normally, downloading and pre-processing of images requires considerable time and effort using the conventional method. GEE eases the computational load of downloading and pre-processing a large number of satellite images. The identified model can be applied to GEE via the JavaScript (shown in

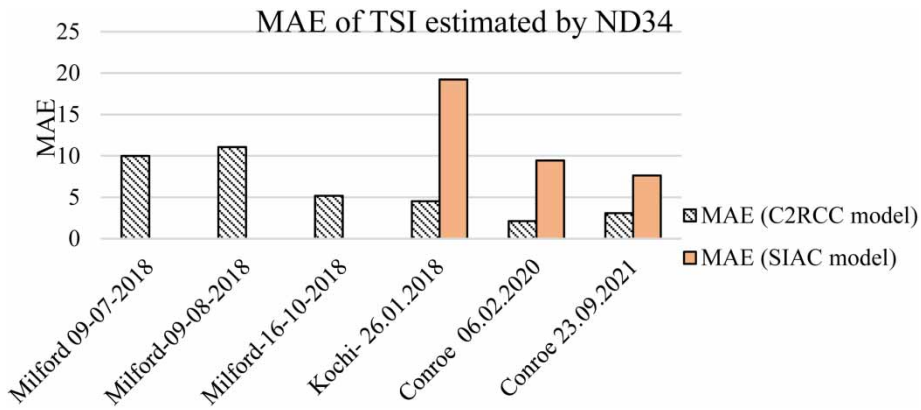


Figure 5 | MAE of TSI estimated from the SIAC dataset using C2RCC model and SIAC model.

Appendix) for obtaining TSI maps of a water body from the S2 images. This section depicts the usefulness of GEE for obtaining temporal and spatial TSI variations in the water body. TSI maps are prepared by applying the model to all the water pixels in all the images for the prescribed duration. These maps give a synoptic view of TSI variations in the water body. For simplicity, the S2 image corrected by the C2RCC processor is named '$R_{rs\ C2RCC}$' and that corrected by the SIAC processor as '$R_{rs\ SIAC}$'.

Initially, the TSI maps of Vembanad Lake (28th January 2021) prepared from the S2 imagery (Figure 6) by using $R_{rs\ C2RCC}$ (Figure 7) and by using $R_{rs\ SIAC}$ (Figure 8) are compared. TSI estimated from this image (Vembanad Lake on 28th January 2021) was validated with *in situ* measurements from Kochi Kayal. MAPE of 7% and MAE of 4.5 was obtained for Kochi Kayal when TSI was estimated by using both $R_{rs\ SIAC}$ (Figure 5) and $R_{rs\ C2RCC}$ (Sherjah *et al.* 2022a). The colour variation in both the TSI maps (Figures 7 and 8) gives a notion that these two estimations are somewhat different and this is due to the narrow range of legend that is used for the representation. The TSI map derived from $R_{rs\ C2RCC}$ (Figure 7) depicts the condition of the lake on the border of hypereutrophic condition (49–60) with only the Kochi Kayal on the north and the area next to Thaneermukkam bund (south) showing higher TSI values of 67. The TSI map derived from $R_{rs\ SIAC}$ (Figure 8) displays the lake with higher TSI values (63–73) and a few areas giving values above 73. A scatter plot of the estimated TSI from these images against *in situ* measurements from Kochi Kayal (Figure 9) shows that the TSI values calculated from $R_{rs\ C2RCC}$ are observed to be slightly underestimated. In order to further verify the accuracy of these predictions, the TSI levels of Kochi Kayal as per *in situ* measurements are overlaid on the TSI map prepared from $R_{rs\ SIAC.}$ *In situ* values of TSI lie between 61 and 76 and these values agree with the TSI estimated by using $R_{rs\ SIAC}$ (Figure 10). Thus, the accuracy of the model with the SIAC corrected dataset is sufficient for estimating the TSI of water bodies by using the GEE platform.

In order to evaluate the computational efficiency of GEE, the temporal variations in TSI across the years 2016–2021 for 11 arbitrary points spread uniformly over Vembanad Lake (p1–p11 shown in Figure 7) are estimated using the conventional method of processing the S2 imagery with SNAP software (atmospheric correction by C2RCC) and by using the GEE platform (atmospheric correction by SIAC). The average time for downloading and processing an image to obtain a TSI map from S2 imagery using the conventional method is around 1 h and the computational load for processing is also high. The TSI

Figure 6 | S2 Imagery of Vembanad Lake (28.01.2022) (subset).

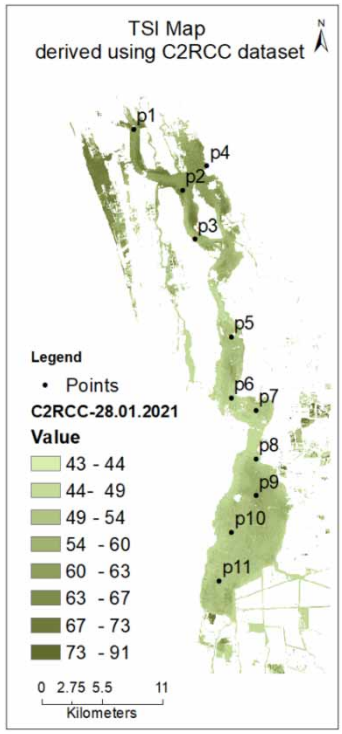


Figure 7 | TSI maps derived using C2RCC corrected R_{rs}.

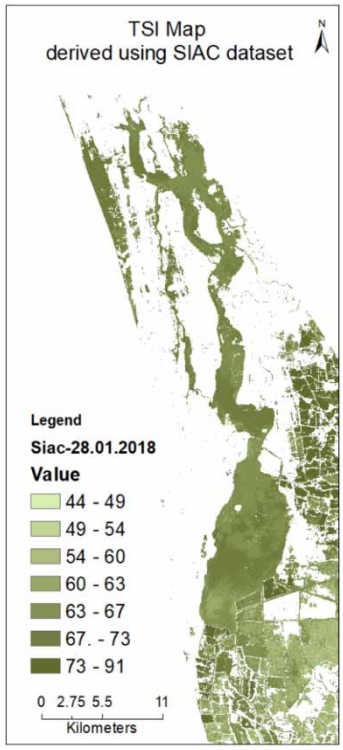


Figure 8 | TSI maps obtained from GEE using SIAC corrected R_{rs}.

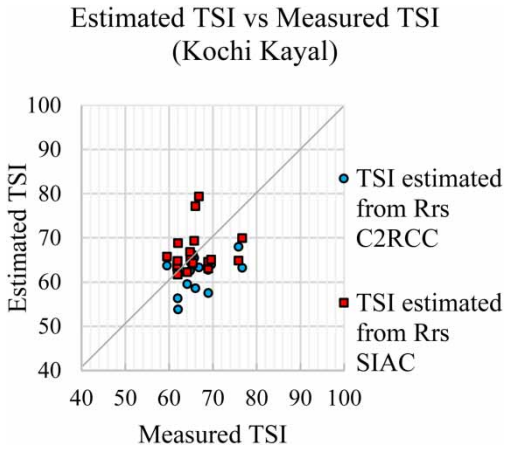


Figure 9 | TSI estimated with C2RCC and SIAC corrected R_{rs} values vs measured values (Kochi Kayal).

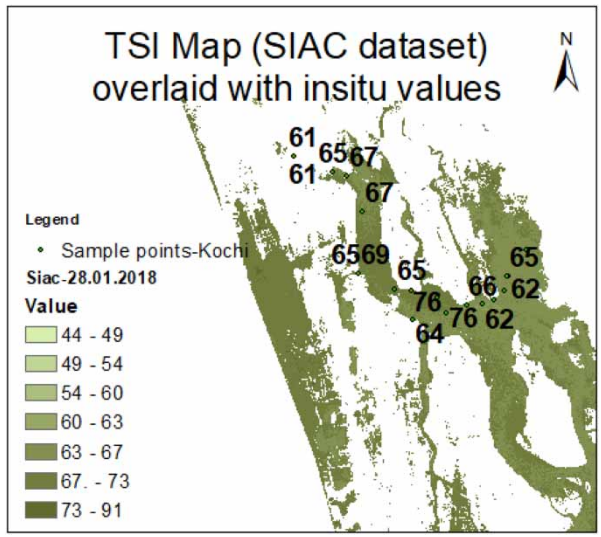


Figure 10 | TSI map prepared using SIAC-corrected S2 image of Kochi overlaid with *in situ* measurements of TSI.

maps were prepared by using eight $R_{rs\ C2RCC}$ images of Vembanad Lake and the temporal variation charts in TSI are displayed in Figure 11. TSI was found to be between 50 and 60 for most of the points. TSI maps were prepared using the same eight images corrected by SIAC, and the temporal variation chart is displayed in Figure 12. The TSI values are around 70 for all the points when estimated with $R_{rs\ SIAC}$.TSI estimated from $R_{rs\ C2RCC}$ was lower than that estimated from $R_{rs\ SIAC}$. This happens because of the difference in ACP. Hence the efficiency of the model is governed by the ACP. There must be efforts to incorporate C2RCC into the GEE platform for full utilization of the developed models. Nevertheless, the methodology of using GEE for water quality monitoring is illustrated in this study.

The computational efficiency of GEE is established by computing the temporal variation of TSI on Vembanad Lake for the period 2016–2021 by processing 164 images (Figure 13). The time for processing 164 images and obtaining the temporal variation charts for those 11 points (mentioned in earlier paragraph) was less than an hour. If this processing has to be carried out by the conventional method, it would have taken at least 164 h, counting 1 h for downloading and processing of an imagery. The temporal variations in TSI across various points on Vembanad Lake for the period 2016–2021 derived from $R_{rs\ SIAC}$ are shown in Figure 13.

The fluctuations in TSI are minor, and the trophic status of the water body is hypereutrophic. TSI at point p1 is observed to go above 75 for the last months of 2020 and the year 2021. The point p1 is located on the Kochi Kayal whose banks are

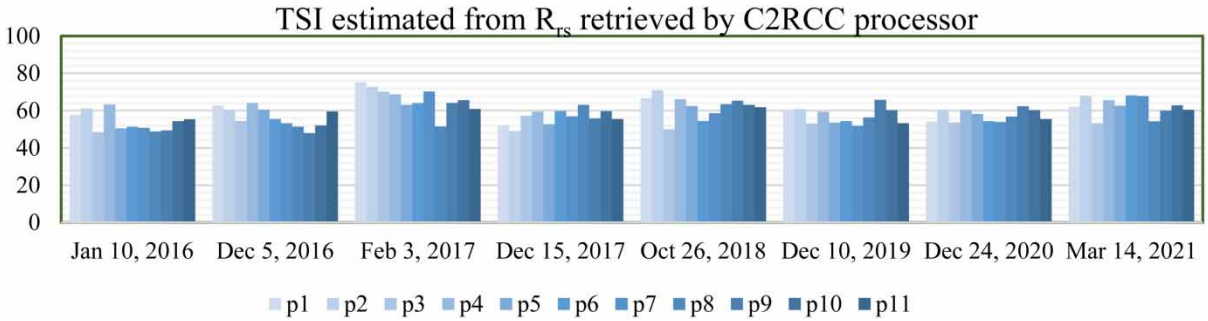


Figure 11 | TSI variation at points on Vembanad Lake (2016–2021) derived using C2RCC corrected dataset (8 images).

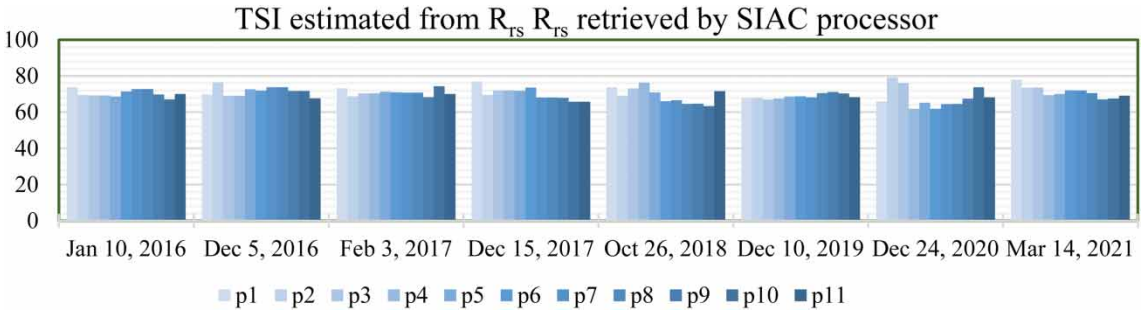


Figure 12 | TSI variation at points on Vembanad Lake (2016–2021) derived using SIAC (8 images).

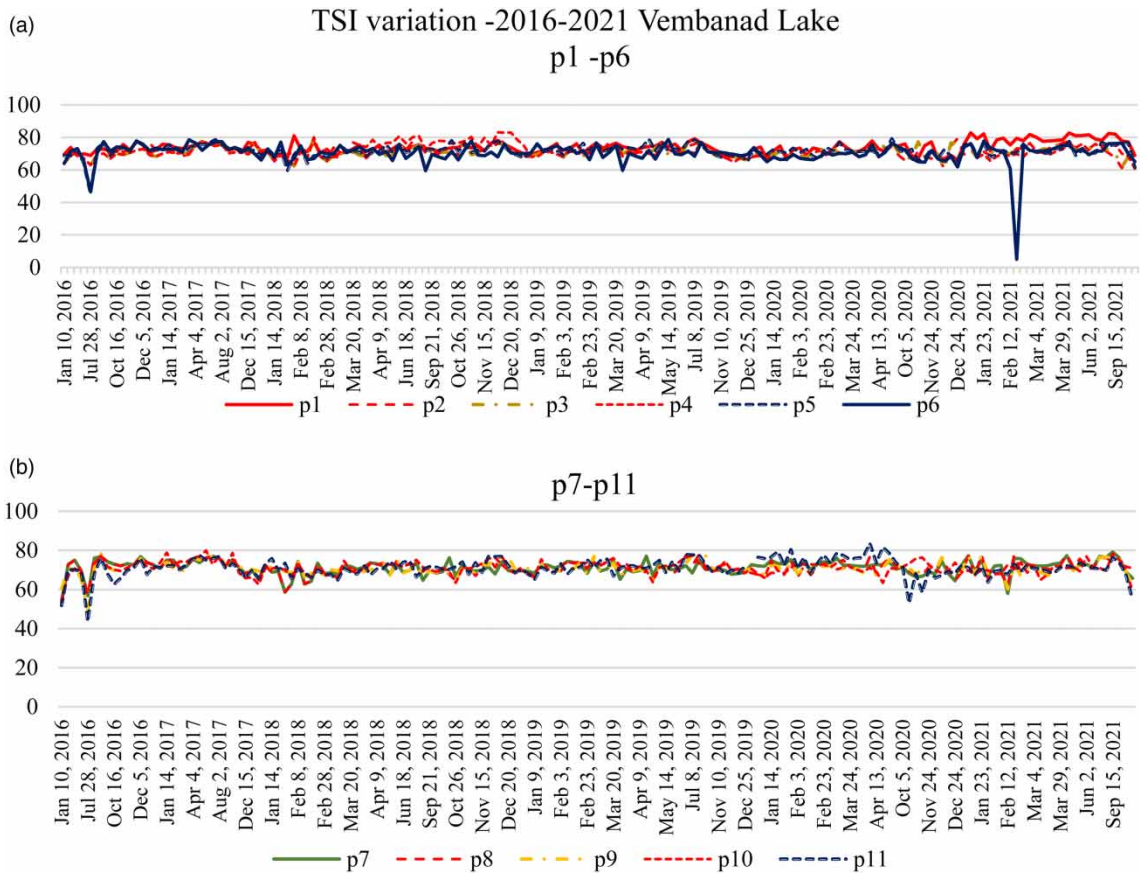


Figure 13 | Temporal variation of TSI in Vembanad Lake obtained from GEE.

populated with industrial units and tourist rest homes. The increase in TSI value for the year 2021 may be due to the rigorous attempts of the society to recover the losses incurred after many months of lockdown (Covid-19). No significant variations are observed that could be linked to the seasonal changes. TSI estimated from most of the points on 'July 28, 2016' imagery dropped below 50; the reason for this was the presence of clouds on this S2 imagery. Hence, it is suggested that images chosen for processing in GEE should be viewed prior to using them for TSI estimation.

The potential of GEE to be used as a quick TSI monitoring tool for inland water bodies is demonstrated in this study. The evaluation of the quality status of inland water bodies by observing the long-term trends and the spatial variation can be obtained with less time and effort by utilizing the GEE platform and the identified index. TSI is an index commonly used by the ecological community and the environmental monitoring centres like United States Environmental Protection Agency (United States Environmental Protection Agency 2007), CETESB (Environmental Company of Sao Paulo State) (Martins *et al.* 2020; Lobo *et al.* 2021) and National Environmental Monitoring Center (China) (Martins *et al.* 2020; Zhu & Mao 2021). Thus, GEE can assist these and other government and non-government authorities in regularly monitoring the quality status of the water bodies by minimizing the time, effort and expense required for the monitoring process. Continuous near real-time quality monitoring of the variation in the water quality is also necessary for deciding and reviewing the government policies and also observing the impact of the industrial and anthropogenic developments on the health of the water bodies.

CONCLUSION

The present study has identified a model for TSI estimation that can be applied to the GEE platform. The SIAC algorithm for atmospheric correction (used in GEE) of water bodies is evaluated for its performance. The study demonstrates the capability of the cloud computing platform GEE for water quality monitoring by processing a large database of Sentinel 2 MSI imagery using less time and effort than required for the conventional method.

The conclusions of the study are briefed below.

- The assessment of the BOA reflectance retrieved by the SIAC processor shows that only reflectances at B3 and B4 have acceptable accuracy to be used for the indices of TSI estimation.
- Of the three sets of indices developed for TSI estimation, only the ND34 index is feasible for GEE. TSI estimation capability of GEE using this index is established by obtaining a reasonable level of average percentage error while validating the index with six images of three lakes.
- The temporal and spatial variation of TSI of Vembanad Lake for the period 2016–2021 is displayed by processing 164 images in a small spell of time on GEE by using the identified ND34 index model.

The study depicts the usefulness of the Google Earth cloud computing platform for obtaining long-term temporal variations and the synoptic view of water quality of inland water bodies quickly so that the critical areas that require attention can be identified at the earliest. The continuous spatial and temporal monitoring of a water body is of utmost importance for maintaining the health of the ecosystem and it is made easier by the developed model on the GEE. This tool may prove beneficial for the government authorities for regular monitoring of the quality of water bodies so that deviations in the trend can be noticed without much delay. Early detection of any deterioration in the quality is necessary for adopting measures to curb further decline.

Limitations of the study and future scope for research

This study is aimed at demonstrating the utility of GEE for water quality monitoring using TSI. The accuracy of estimation is greatly influenced by the accuracy of the retrieval of BOA reflectance. It is the need of the hour to develop better atmospheric correction algorithms for water bodies that can be utilized on the GEE platform. Incorporation of the C2RCC algorithm into the GEE platform can improve the prospects of GEE for water quality monitoring studies. The potential of GEE is immense in this area of research. GEE can be used for monitoring the water quality parameters that are detectable by remote sensing (namely, CDOM, Chl-*a*, suspended sediment and turbidity) and that have established algorithms for their retrieval from satellite reflectance.

DATA AVAILABILITY STATEMENT

All relevant data are available from Copernicus Open Access Hub (https://scihub.copernicus.eu).

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Amani, M., Ghorbanian, A., Ahmadi, S. A., Kakooei, M., Moghimi, A., Mirmazloumi, S. M., Moghaddam, S. H. A., Mahdavi, S., Ghahremanloo, M., Parsian, S., Wu, Q. & Brisco, B. 2020 Google Earth Engine cloud computing platform for remote sensing big data applications: a comprehensive review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **13**, 5326–5350.

Barrett, D. C. & Frazier, A. E. 2016 Automated method for monitoring water quality using landsat imagery. *Water (Switzerland)* **8** (6), 1–14.

Beck, R., Xu, M., Zhan, S., Johansen, R., Liu, H., Tong, S., Yang, B., Shu, S., Wu, Q., Wang, S., Berling, K., Murray, A., Emery, E., Reif, M., Harwood, J., Young, J., Nietch, C., Macke, D., Martin, M., Stillings, G., Stumpf, R., Su, H., Ye, Z. & Huang, Y. 2019 Comparison of satellite reflectance algorithms for estimating turbidity and cyanobacterial concentrations in productive freshwaters using hyperspectral aircraft imagery and dense coincident surface observations. *Journal of Great Lakes Research*, International Association of Great Lakes Research **45** (3), 413–433.

Blondeau-Patissier, D., Gower, J. F. R., Dekker, A. G., Phinn, S. R. & Brando, V. E. 2014 A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans. *Progress in Oceanography*, Elsevier Ltd **123**, 123–144.

Cardall, A., Tanner, K. B. & Williams, G. P. 2021 Google earth engine tools for long-Term spatiotemporal monitoring of Chlorophyll-a concentrations. *Open Water Journal* **7** (1), 4.

Carlson, R. E. 1977 A trophic state index for lakes. *Limnology and Oceanography* **22** (2), 361–369.

Doerffer, R. & Schiller, H. 2007 The MERIS case 2 water algorithm. *International Journal of Remote Sensing* **28** (3–4), 517–535.

Dörnhöfer, K. & Oppelt, N. 2016 'Remote sensing for lake research and monitoring – recent advances. *Ecological Indicators*, Elsevier Ltd **64**, 105–122.

Effler, S. W. 1989 Sechhi disk transparency and turbidity. *Journal of Environmental Engineering* **114** (6), 1436–1447.

Kratzer, C. R. & Brezonik, P. L. 1982 A carlson-type trophic state index for nitrogen in Florida Lakes. *Water Resources Bulletin, America Water Resources Association* **17** (4), 713–715.

Lee, Z., Carder, K. L. & Arnone, R. A. 2002 Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. *Applied Optics* **41** (27), 5755–5772.

Leiker, B. M., Abel, J. R., Graham, J. L., Foster, G. M., King, L. R., Stiles, T. C. & Buley, R. P. 2019 *Spatial and Temporal Variability of Nutrients and Algae in the Republican River and Milford Lake, Kansas, June Through November 2017 and May Through November 2018. U S Geological Survey*.

Lobo, F. d. L., Nagel, G. W., Maciel, D. A., de Carvalho, L. A. S., Martins, V. S., Barbosa, C. C. F. & de Moraes Novo, E. M. L. 2021 Algaemap: algae bloom monitoring application for inland waters in Latin America. *Remote Sensing* **13** (15), 2874.

Martins, I. A., Fein, D., Pompêo, M. L. M. & Bitencourt, M. D. 2020 Determination of the trophic state index (Tsi) using remote sensing, bathymetric survey and empirical data in a tropical reservoir. *Limnetica* **39** (2), 539–553.

Moore, G. K. 1980 Satellite remote sensing of water turbidity. *Hydrological Sciences-Bulletin-des Sciences Hydrologiques* **25** (4), 407–421.

Myre, E. & Shaw, R. 2006 The Turbidity Tube : Simple and Accurate Measurement of Turbidity in the Field. Michigan Technological University, (April), 1–15.

Owa, F. D. 2013 Water pollution: sources, effects, control and management. *Mediterranean Journal of Social Sciences* **4** (8), 65–68.

Papoutsa, C., Akylas, E. & Hadjimitsis, D. 2014 Trophic state index derivation through the remote sensing of case-2 water bodies in the Mediterranean region. *Central European Journal of Geosciences* **6** (1), 67–78.

Pereira-Sandoval, M., Ruescas, A., Urrego, P., Ruiz-Verdú, A., Delegido, J., Tenjo, C., Soria-Perpinyà, X., Vicente, E., Soria, J. & Moreno, J. 2019 Evaluation of atmospheric correction algorithms over Spanish inland waters for sentinel-2 multi spectral imagery data. *Remote Sensing*, Multidisciplinary Digital Publishing Institute (MDPI) **11** (12), 1469.

Priju, C. P. & Narayana, A. C. 2007 Heavy and trace metals in Vembanad Lake sediments. *International Journal of Environmental Research* **1** (4), 280–289.

Sherjah, P. Y., Sajikumar, N. & Nowshaja, P. T. 2022a Trophic status mapping of inland water bodies using normalized difference indices derived from Sentinel 2 MSI imagery. *Journal of Hydrologic Engineering* **28** (1), 04022035.

Sherjah, P. Y., Sajikumar, N. & Nowshaja, P. T. 2022b Semi-analytical model for TSI estimation of inland water bodies from Sentinel 2 imagery. *Journal of Hydroinfomatics* **24** (2), 444–463.

Shi, K., Zhang, Y., Song, K., Liu, M., Zhou, Y., Zhang, Y., Li, Y., Zhu, G. & Qin, B. 2019 A semi-analytical approach for remote sensing of trophic state in inland waters: bio-optical mechanism and application. *Remote Sensing of Environment*, Elsevier **232** (September 2018), 111349.

Sruthy, S. & Ramasamy, E. V. 2017 Microplastic pollution in Vembanad Lake, Kerala, India: the first report of microplastics in lake and estuarine sediments in India. *Environmental Pollution* **222**, 315–322.

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.

United States Environmental Protection Agency 2007 'Carlson's Trophic State Index.' *Aquatic Biodiversity*.

University of South Florida 2018 'Trophic State Index (TSI).' *Lake County Water Atlas*. Available from: http://www.lake.wateratlas.usf.edu/shared/learnmore.asp?toolsection=lm_tsi.

Viswanathan, M. 2019 'Toxic contamination of Vembanad lake in Kochi poses major health hazard.' *New Indian Express*, Kochi.

Warren, M. A., Simis, S. G. H., Martinez-Vicente, V., Poser, K., Bresciani, M., Alikas, K., Spyrakos, E., Giardino, C. & Ansper, A. 2019 Assessment of atmospheric correction algorithms for the Sentinel-2A MultiSpectral Imager over coastal and inland waters. *Remote Sensing of Environment* Elsevier **225** (September 2018), 267–289.

Yang, Z., Li, W., Chen, Q., Wu, S., Liu, S. & Gong, J. 2019 A scalable cyberinfrastructure and cloud computing platform for forest aboveground biomass estimation based on the Google Earth Engine. *International Journal of Digital Earth* Taylor & Francis **12** (9), 995–1012.

Yin, F., Lewis, P. E., Gomez-Dans, J. L. & Wu, Q. 2019 *A Sensor Invariant Atmospheric Correction : Sentinel-2/Msi and Landsat 8/OLI*. 1–42.

Zhu, S. & Mao, J. 2021 A machine learning approach for estimating the trophic state of urban waters based on remote sensing and environmental factors. *Remote Sensing* **13** (13), 2498.

First received 3 September 2022; accepted in revised form 23 December 2022. Available online 12 January 2023