Spatio-temporal assessment of inland surface water quality using remote sensing data in the wake of changing climate

Oupa E. Malahlela
Earth Observation Division, South African National Space Agency (SANSA), Pretoria, South Africa, 0184.
Email: omalahlela@sansa.org.za

Abstract. Remote sensing indices have been widely used for mapping water quality parameters such as the suspended material and chlorophyll-a concentration. This study aims at investigating the spatio-temporal performance of spectral indices for mapping suspended sediments in water using Landsat 8 data. The results show that the normalized suspended material index (NSMI) is the most effective index for multi-temporal mapping of suspended solids. The use of other spectral indices in the visible to shortwave infrared also offers reasonable estimation ($R^2 > 0.7; p < 0.05$) of water quality parameters. It is recommended that the performance of the tested indices be compared with other indices derived from high resolution data such as the Ziyuan-3 and Sentinel-2 satellites for operational monitoring of inland water bodies.

1. Introduction
Freshwater and estuarine ecosystems throughout the world are continuously impacted by changing climatic conditions. One of the ways in which such impacts are manifested is through changes in the quality of the surface water bodies [1]. In addition, climate change is interlinked with anthropogenic activities that often deteriorates the status of water quality within economically significant water bodies [2]. Flooding and droughts are often the main phenomena through which surface water bodies are impacted by climate change. Flooding modifies water quality through concentrations of dissolved and suspended materials, while in drought conditions, increases in water temperatures results in increased concentrations of dissolved materials, as a result of low oxygen concentrations [3,4]. Studying the concentrations of suspended materials and changes in water spatial extent has always been one of the crucial steps for surface effective water management, treatment and allocations.

As for studying the changes in surface water quality in major water bodies, recent scientific developments have shown that remote sensing is increasingly becoming the preferred method [5-7]. Much of the focus is thus placed on derivation of linear correlations between spectral data (e.g. spectral indices) and the water quality parameters (e.g. total suspended solid, chlorophyll-a and coloured dissolved organic matter) [5,8,9]. However, the temporal sensitivity and accuracy of spectral indices for mapping water quality parameters based on satellite data is not well understood. Understanding the sensitivity and predictive accuracy of spectral indices using multiple images through time will enhance the design of operational systems designed to assess water quality from remotely-sensed data at much larger scale (national, regional, continental and global). Therefore this study aims at assessing the quantification of inland surface water quality using multi-temporal remote sensing data in the wake of changing climate. The objectives are: (i) to extract surface water features;
(ii) to test the performance of various water quality (suspended sediments) spectral indices over time (seasonal and annual), and; (iii) to map the suspended material using the most robust linear model as identified in objective (ii).

2. Methods

2.1. Description of the study area
The main water bodies under consideration included the Spring Grove Dam, Midmar Dam, Nagle Dam, Albert Falls Dam and the Inanda Dam which are found in KwaZulu-Natal, South Africa. These dams are located in both the uMgeni River Catchment, covering an area of 66.10 km$^2$, within the geographic coordinates of 29.5547°S and 30.5522°E (figure 1). This river catchment and its reservoirs (dams) have elevated rates of turbidity, suspended solids levels, faecal contamination, eutrophication, concomitant pathogen risks and erosion sediment contamination [10].

![Figure 1. Map of the study area showing where the studied water bodies along with test points depicted by yellow colour where values for calculated indices were collected from Landsat 8 OLI imagery.](image)

2.2. Data pre-processing
Landsat 8 OLI data acquired between 2014 and 2018 were used for analysis. The images were downloaded from the United States Geological Surveys (USGS) earth explorer website free of charge (https://earthexplorer.usgs.gov/), comprising of 10 multispectral bands at 30 m spatial resolution. The details of the images used in this study are found in table 1. The Quick Atmospheric Correction (QUAC) correction module embedded in ENVI™ 5.1 software was used to minimize the effect of atmospheric interference on the spectral reflectance. The QUAC module was used because of its high robustness for water quality proxy retrieval, than the common fast line-of-sight atmospheric adjustment of spectral hypercubes (FLAASH) algorithm, particularly with the use of near infrared-red
band algorithms [11]. Following the atmospheric correction, the simple water index (SWI) was applied to extract the surface water body extents for selected time periods [12]. The regions of interest (ROIs), each comprising of minimum 16 pixels (approx. 14 400 m²) were selected within the studied dams to represent classes of low (1) to high (3) suspended sediment concentrations based on visual assessment of images. The pixels within ROIs of high concentrations of suspended sediments were considered as those with very similar spectral curve to nearby off-shore mixture of soil/vegetation spectral signature (figure 2).

Table 1. Characteristics of the image acquisition considered in the study.

| Year of acquisition | Time period 1 ($T_1$) | Time period 2 ($T_2$) | Data         |
|---------------------|-----------------------|-----------------------|--------------|
| 2014                | April                 | August                | Landsat 8 OLI |
| 2015                | April                 | August                | Landsat 8 OLI |
| 2016                | Oct                   | n/a (cloudy)          | Landsat 8 OLI |
| 2017                | Apr                   | September             | Landsat 8 OLI |
| 2018                | May                   | -n/a (cloudy)         | Landsat 8 OLI |

2.3. Data analysis and model development

In total, 36 ROI ($n = 36$) were defined for each image, belonging to the classes of low to high concentrations. From this dataset 50% ($n = 18$) was used for linear model building while the other 50% was used for validation. Four spectral indices were tested for the study, including the normalized suspended material index (NSMI), the water-sediment ratio index (WSRI), the normalized difference vegetation index (NDVI), and the enhanced green ratio index (EGRI) (table 2).

In this study, the empirical relationship between the visually suspended sediment material and the spectral indices was retrieved using a stepwise linear regression (SLR). In SLR, all variables (spectral indices) are added to the starting model one after another, and the variable is kept in the model if its addition led to a reduction in Akaike’s information criterion (AIC) [13,14]. The model development was done in R software using the glm2 (https://cran.r-project.org/web/packages/glm2/index.html) and MASS (https://cran.r-project.org/web/packages/MASS/index.html) packages. The resultant model was used for mapping the water quality condition within the waterbodies in the study area. The coefficient of determination ($R^2$), the root mean square error (RMSE) and the relative RMSE (rRMSE) were used as validation measures between the observed and predicted concentrations at each time period. The RMSE and rRMSE are defined by equations (1) and (2) respectively:

\[
RMSE = \sqrt{\frac{\sum (ss_{pred} - ss_{obs})^2}{n-1}} \tag{1}
\]

\[
rRMSE = \frac{\sqrt{\sum (ss_{pred} - ss_{obs})^2}}{\bar{x}} \times 100 \tag{2}
\]

where $ss_{pred}$ is the predicted suspended material, $ss_{obs}$ is the visually observed suspended material from Landsat 8 imagery, and $\bar{x}$ is the mean of the $ss_{obs}$.

3. Results and discussion

The results of the current study are shown in table 3. Generally all linear models yielded higher predictive accuracies ($R^2 > 0.7$) than prediction by chance ($R^2 = 0.5$) and all model coefficients were statistically significant ($P < 0.0001$). The relatively poor RMSE from the validation dataset was obtained with model obtained from August 2014 (0.49), while the lowest RMSE (0.23) was obtained with the dataset of September 2017 (table 4 and figure 3), indicating good agreement between
observed and predicted suspended sediments in the study area. The performance of these accuracy measures was compared seasonally and annually.

**Figure 2.** The average spectral signatures of both low and high suspended sediments used in the study.

**Table 2.** Formulation of the spectral indices for studying suspended materials derived from Landsat 8 OLI datasets.

| Index | Formula | References |
|-------|---------|------------|
| NSMI  | \[ \frac{(\rho_{\text{red}}) + (\rho_{\text{green}}) - (\rho_{\text{blue}})}{(\rho_{\text{red}}) + (\rho_{\text{green}}) + (\rho_{\text{blue}})} \] | [15] |
| NDVI  | \[ \frac{(\rho_{\text{NIR}} - \rho_{\text{red}})}{(\rho_{\text{NIR}} + \rho_{\text{red}})} \] | [16] |
| WSRI  | \[ 1 - \frac{(\rho_{\text{SWIR}} - \rho_{\text{blue}})}{(\rho_{\text{green}} + \rho_{\text{blue}})} \] | In this study |
| EGRI  | \[ \frac{\rho_{\text{Irr}}}{\rho_{\text{green}} + \rho_{\text{blue}}} \] | In this study |

The results show that the performance of spectral indices related to suspended sediment varies over time, and per index used. However, the NSMI was the most sensitive index to suspended sediments throughout the study period, except in August 2015. On the other hand, both the EGRI and WSRI have also showed slightly higher performance than the NDVI. This could essentially serve as indication that throughout the year the dams tend to be concentrated with suspended sediments, as opposed to algal concentrations which are common during summer months. Since the September 2017 image had low RMSE (0.86 mg/l), rRMSE (6.37%) and relatively high $R^2$, the spatial distribution map was thus produced from this model because of its robustness (high accuracy and low error rate). Figure 4 shows the spatial distribution of the turbidity across 5 water bodies in the KwaZulu-Natal area. The water bodies with high concentrations of suspended sediments are the Springgrove Dam and the Albert Falls Dam both of which are situated in the upper catchment area [17]. These water bodies tend to reduce the sediment load of water bodies downstream, thus reducing their own reservoir capacity in the area.
The models computed with three-band index tested in the study (WSRI) showed high sensitivity to spatial modeling of SS within the uMgeni River Catchment thus making it robust for assessing physical characteristics of water bodies under changing climate scenarios.

**Table 3.** Results of the predictive models across various times and spectral indices.

| Period  | Regression model                                                                 | $R^2$ | $r$RMSE (%) | $P$     |
|---------|----------------------------------------------------------------------------------|-------|--------------|---------|
| Apr 2014 | 6.383 – (6.31 * NSMI) + (2.42 * EGRI) – (1.80 * WSRI)                           | 0.71  | 11.42        | 0.0003  |
| Aug 2014 | 1.74 + (1.00 * NDVI) + (12.84 * NSMI) – (6.37 * EGRI) + (0.10 * WSRI)          | 0.87  | 13.66        | 0.0005  |
| Apr 2015 | 1.51 + (2.02 * NDVI) + (10.60 * NSMI) – (5.42 * EGRI) + (0.33 * WSRI)          | 0.82  | 8.46         | 0.0004  |
| Aug 2015 | 3.30 – (0.811 * NDVI) – (0.39 * WSRI)                                           | 0.73  | 9.22         | 0.0002  |
| Oct 2016 | 1.11 + (1.10 * NDVI) + (5.68 * NSMI) – (1.68 * EGRI)                           | 0.90  | 12.48        | 0.0001  |
| Apr 2017 | 2.34 + (2.0 * NDVI) + (4.91 * NSMI) – (4.05 * EGRI) + (0.14 * WSRI)            | 0.98  | 10.84        | 0.0000  |
| Sep 2017 | 3.82 – (3.24 * NSMI) – (0.77 * WSRI)                                            | 0.86  | 6.37         | 0.0001  |
| May 2018 | 1.71 + (1.24 * NDVI) + (6.62 * NSMI) – (3.15 * EGRI)                           | 0.85  | 8.25         | 0.0001  |

**Figure 3.** The graph showing the performance of linear models used for mapping suspended sediments in the subtropical region of South Africa.
Figure 4. The results of the more robust stepwise linear model produced from the September 2017 remote sensing indices. The Albert Falls Dam has shown the highest SS concentration predicted for 2017.

4. Conclusions and recommendations
This study aimed at assessing the sensitivity of spectral indices for mapping suspended sediments in a stepwise linear regression model over time. We conclude that the common NSMI is insensitive to changes in temporal variability of target spectral response for mapping concentrations of suspended material in inland water bodies. The indices derived from the visible-shortwave infrared region of electromagnetic spectrum showed greater accuracy for mapping water quality parameters than the two-band spectral indices derived from the visible-NIR region. It is thus recommended that the NSMI, EGRI, and WSRI be tested further for water bodies that are more murky/turbid before such indices can be used as part of operational system for water quality assessment at much larger scale. The testing of these indices should also be extended to other remote sensing instruments such as high resolution Ziyuan-3 and Sentinel-2.

Acknowledgement
I wish to thank South African National Space Agency (SANSA-EO) for support in this project. Special appreciation to United States Geological Survey, and the R Development Core Team for making both data and processing software freely available. I thank four anonymous reviewers for their contribution in improving the quality of this conference manuscript.

References
[1] Depla I, Jung AV, Baures M et al 2009 Impacts of climate change on surface water quality in relation to drinking water production Environment International 35(8) 1225-1233
[2] Han HJ, Allan D, Scavia D 2009 Influence of climate and human activities on the relationship between watershed nitrogen input and river export Environmental Science and Technology 43(6) 1916-1922
[3] Van Vliet MTH, Zwolsman JJJG 2008 Impact of summer droughts on the water quality of the Meuse River Journal of Hydrology 353(1-2) 1 – 17
[4] Giorgi F, Im ES, Coppola E, Diffenbaugh NS, Gao XJ, Mariotti L, Shi Y 2011 Higher hydroclimatic intensity with global warming Journal of Climatology 24 5309 – 5324
[5] Shi L, Mao Z, Wang Z 2018 Retrieval of total suspended matter concentrations from high resolution WorldView-2 imagery: a case study of inland rivers IOP Conference Series: Earth and Environmental Science 121 032036
[6] Petus C, Chust G, Gohin F, et al 2010 Estimating turbidity and total suspended matter in the Adour River plume (South Bay of Biscay) using MODIS 250-m imagery Continental Shelf Research 30(5) 379-392
[7] Sharaf El Din E., Zhang Y, Suliman A 2017 Mapping concentrations of surface water quality parameters using novel remote sensing and artificial intelligence framework International Journal of Remote Sensing 38(4) 1023-1042
[8] Keith DJ, Lunetta RS, Schaeffer BA 2016 Optical models for remote sensing of coloured dissolved organic matter absorption and salinity in New England, Middle Atlantic and Gulf Coast Estuaries USA Remote Sensing 8(283)
[9] Frouin R, Deschamps P, Gross-Colzy L, et al 2006 Retrieval of chlorophyll-a concentration via linear combination of ADEOS-II Global Imager data Journal of Oceanography 62(3) 331-337
[10] Umgeni Water Annual Report 2015/2016 Conserving our natural resources. pp 1-12
[11] Moses WJ, Gitelson AA, Perk RL et al 2012 Estimation of chlorophyll-a concentration in turbid productive waters using airborne hyperspectral data Water Research 46(4) 993-1004
[12] Malahlela OE 2016 Inland waterbody mapping: towards improving discrimination and extraction of inland surface water features International Journal of Remote Sensing 37(19) 4574 – 4589
[13] Huang C, Townsend JRG 2003 A stepwise regression tree for non-linear approximation: applications to estimating subpixel land cover International Journal of Remote Sensing 24(1) 75-90
[14] Zhu J, Huang Z, Sun H et al 2017 Mapping forest ecosystem biomass density for Xiangjiang River Basin by combining plot and remote sensing data and comparing spatial extrapolation methods Remote Sensing 9(3) 241
[15] Azad Hossain AKM, Chao X, Yafei J 2006 Development of remote sensing based index for estimating/mapping suspended sediment concentration in river and lake environment, The University of Mississippi
[16] Tucker CJ 1979 Red and photographic infrared linear combinations for monitoring vegetation Remote Sensing of Environment 8(2) 127-150
[17] Department of Water Affairs 2014 The uMkhomazi Water Project Phase 1 Report. Report no: WMA11/U10/00/3312/3/1/7