Geophysical Research Abstracts
Vol. 16, EGU2014-9653-3, 2014
EGU General Assembly 2014

Coastal water quality estimation from Geostationary Ocean Color Imager (GOCI) satellite data using machine learning approaches

Jungho Im (1), Sunghyun Ha (1), Yong Hoon Kim (2), Hokyung Ha (3), Jongkuk Choi (4), and Miae Kim (1)

(1) Ulsan National Institute of Science and Technology, Korea, Republic Of (ersgis@unist.ac.kr), (2) RPS ASA, USA (ykim@asascience.com), (3) Korea Polar Research Institute, Korea, Republic Of (ha@kopri.re.kr), (4) Korea Institute of Ocean Science and Technology, Korea, Republic Of (jkchoi@kiost.ac)

It is important to monitor coastal water quality using key parameters such as chlorophyll-a concentration and suspended sediment to better manage coastal areas as well as to better understand the nature of biophysical processes in coastal seawater. Remote sensing technology has been commonly used to monitor coastal water quality due to its ability of covering vast areas at high temporal resolution. While it is relatively straightforward to estimate water quality in open ocean (i.e. Case I water) using remote sensing, coastal water quality estimation is still challenging as many factors can influence water quality, including various materials coming from inland water systems and tidal circulation. There are continued efforts to accurately estimate water quality parameters in coastal seawater from remote sensing data in a timely manner.

In this study, two major water quality indicators, chlorophyll-a concentration and the amount of suspended sediment, were estimated using Geostationary Ocean Color Imager (GOCI) satellite data. GOCI, launched in June 2010, is the first geostationary ocean color observation satellite in the world. GOCI collects data hourly for 8 hours a day at 6 visible and 2 near-infrared bands at a 500 m resolution with 2,500 x 2,500 km square around Korean peninsula. Along with conventional statistical methods (i.e. various linear and non-linear regression), three machine learning approaches such as random forest, Cubist, and support vector regression were evaluated for coastal water quality estimation. In situ measurements (63 samples; including location, two water quality parameters, and the spectra of surface water using a hand-held spectroradiometer) collected during four days between 2011 and 2012 were used as reference data. Due to the small sample size, leave-one-out cross validation was used to assess the performance of the water quality estimation models. Atmospherically corrected radiance data and selected band-ratioed images were used as predictor variables. Results show that support vector regression outperformed the other two machine learning approaches as well as conventional statistical models, yielding calibration R2 of 0.9 and cross validation RMSE of 1.7 mg/m3 for chlorophyll-a concentration, and calibration R2 of 0.97 and cross validation RMSE of 11.4 g/m3 for suspended sediment. Relative importance of the predictor variables was examined and the spatiotemporal patterns of the water quality parameter distribution were analyzed along with tidal information.