Application of Machine-Learning Techniques Toward the Creation of a Consistent and Calibrated Global Chlorophyll Concentration Baseline Dataset Using Remotely Sensed Ocean Color Data

Ewa J. Kwiatkowska, *Member, IEEE*

***Abstract*—This paper introduces a machine-learning approach to satellite ocean color sensor cross calibration. The cross-calibration objective is to eliminate incompatibilities among sensor data from different missions and produce merged daily global ocean color coverage. The approach is designed and investigated using data from the Moderate Resolution Imaging Spectroradiometer (MODIS) onboard of the Terra satellite and Sea-viewing Wide Field-of-view Sensor (SeaWiFS). Data from these two sensors show apparent discrepancies originating from differences in sensor design, calibration, processing algorithms, and from the rate of change in the atmosphere and ocean within 1(1/2) h between sensor imaging of the same regions on the ground. The discrepancies have complex, noisy, and often contradictory time and space variabilities. Support vector machines are used to bring MODIS data to the SeaWiFS representation where SeaWiFS data are considered to exemplify a consistent ocean color baseline. Support vector machines are effective in learning and resolving convoluted data relationships between the two sensors given a variety of bio-optical, atmospheric, viewing geometry, and ancillary information. The method works accurately in low chlorophyll waters and shows a potential to eliminate sensor problems, such as scan angle dependencies and seasonal and spatial trends in data. The results illustrate that MODIS and SeaWiFS differences are noisy and highly variable, which makes it difficult to extrapolate the cross-calibration knowledge onto new time and space domains and to define representative global ocean color datasets for support vector machine training.**

***Index Terms*—Chlorophyll, data merger, fusion, machine learning, Moderate Resolution Imaging Spectroradiometer (MODIS), ocean color, regression, Sea-viewing Wide Field-of-view Sensor (SeaWiFS), support vector machines.**

I. INTRODUCTION

PHYTOPLANKTON are the principal source of organic matter in the oceans that sustain the marine food chain. They also act as a biological pump that sequesters carbon dioxide from the atmosphere to the deep ocean [1]. Some characteristics of the upper ocean, including phytoplankton concentrations, can be differentiated in terms of solar radiance scattered upward in the visible part of the electromagnetic spectrum. Phytoplankton concentrations are expressed in terms of the concentration of the main phytoplankton photosynthetic pigment, chlorophyll-a, which is often considered as an index of phytoplankton biomass [2]. Other factors that influence the backscatter signal are scattering by inorganic suspended material, scattering from water molecules, absorption by organic detrital material, and reflection off the sea bottom. Remote sensing satellites can detect these spectral-radiance signatures of the ocean surface. Phytoplankton-dedicated satellite sensors operate using a range of spectral bands within the visible and near-infrared spectrum, and the derived data are called ocean color. Satellite data need to undergo challenging sensor calibration [3], [4], atmospheric, and other corrections [5], and normalizations for satellite viewing and sun angles before water-leaving radiances can be obtained. Normalized water-leaving radiances (nLw) are then converted to chlorophyll-a concentrations using empirical algorithms [6].

There have been many satellite ocean color sensors on-orbit, and many are planned for the near future. The Coastal Zone Color Scanner (CZCS) was an initial proof-of-concept mission. The sensor was launched by the National Aeronautics and Space Administration (NASA) in 1978 and operated for eight years, providing global coverage in ocean color data. The subsequent ocean color missions are shown in Table I.

NASA's Sensor Intercomparison and Merger for Biological and Interdisciplinary Oceanic Studies (SIMBIOS) Project at Goddard Space Flight Center (GSFC) has been involved in collection, processing, archival, and documentation of calibration and bio-optical datasets obtained from various *in situ* and on-orbit instruments [7]–[9]. The SIMBIOS Project has been pursuing sensor data comparisons [10]; calibrations [11]–[13]; development, validation, and implementation of algorithms [7]; and full dataset processing [14].

The ultimate objective of the SIMBIOS Project is to integrate information from multiple satellite ocean color sensors for global information enhancement. The goal is to define a consistent series of multiinstrument, multiplatform, and multiyear observations based on accurate and uniform calibration and validation over the lifetime of the measurement [15]. The effort is to bring data from different sensors to a uniform, consistently calibrated, and well-validated ocean color baseline.

The most obvious benefit of the data merger is improvement in spatial and temporal ocean color coverage. Ocean color

Manuscript received January 31, 2003; revised June 4, 2003. This work was supported by the National Aeronautics and Space Administration's Sensor Intercomparison and Merger for Biological and Interdisciplinary Oceanic Studies (SIMBIOS) Project.

The author is with the Science Applications International Corporation, NASA Goddard Space Flight Center, Greenbelt, MD 20771 USA (e-mail: ewa@simbios.gsfc.nasa.gov).

Digital Object Identifier 10.1109/TGRS.2003.818016

TABLE I
OCEAN COLOR SENSORS

Global Coverage Missions	Missions with Principally Limited Coverage Applications
Ocean Color and Temperature Scanner (OCTS), Japan, 1996-1997, 10:41am Equator crossing local time, descending orbit	Modular Optoelectronic Scanner (MOS), Germany/India, 1996-
Polarization and Dimensionality of the Earth's Reflectances (POLDER I), France, 1996-1997, 10:41am Equator crossing local time, descending orbit	Ocean Color Monitor (OCM), India, 1999-
SeaWiFS, USA, 1997-, 12:20pm Equator crossing local time, descending orbit	Ocean Color Imager (OCI), Taiwan, 1999-2002
MODIS on Terra, USA, 1999-, 10:30am Equator crossing local time, descending orbit	Ocean Scanning Multispectral Imager (OSMI), South Korea, 1999-
Multi-angle Imaging SpectroRadiometer (MISR), USA, 1999-, 10:30am Equator crossing local time, descending orbit	Chinese MODIS (CMODIS), China, 2002-
MODIS on Aqua, USA, 2002-, 1:30pm Equator crossing local time, ascending orbit	Chinese Ocean Color and Temperature Scanner (COCTS), China, 2002-
Medium Resolution Imaging Spectrometer (MERIS), France, 2002-, 10:00am Equator crossing local time, ascending orbit	Coastal Zone Imager (CZI), China, 2002-
Global Imager (GLI), Japan, 2002-, 10:30am Equator crossing local time, descending orbit	
POLDER II, France, 2002-, 10:30am Equator crossing local time, descending orbit	
Visible/Infrared Imager Radiometer Suite (VIIRS), USA, to be launched in 2006, 10:30am Equator crossing local time, descending orbit	

sensors are placed on polar orbiting satellites. Single-sensor daily global coverage is severely limited by gaps between the consecutive satellite overflight coverages (i.e., swaths). Global missions revisit each worldwide location every second day, while the revisits of the limited coverage missions are restricted to given imaging schedules and target locations. Gaps are also caused by clouds, sun glint, and other phenomena that disable the extraction of ocean color [16]. Based on collection 4 Moderate Resolution Imaging Spectroradiometer (MODIS)–Terra, collection 3 MODIS–Aqua, and reprocessing 4 Sea-viewing Wide Field-of-view Sensor (SeaWiFS) data, daily global 9-km resolution ocean coverages from either MODIS–Terra or MODIS–Aqua are increased by approximately 40% from adding SeaWiFS observations.[1] MODIS–Terra daily global ocean coverage at 9-km resolution is improved by 55% from adding MODIS–Aqua data. Combined daily ocean representations from MODIS on Terra and Aqua are additionally raised by 15% by incorporating SeaWiFS data. All these statistics only

[1]See MODIS Science Team Meeting Presentation by E. J. Kwiatkowska, "Daily global oceans coverage improvements based on MODIS–Terra, MODIS–Aqua, and SeaWiFS matchups." http://seawifs.gsfc.nasa.gov/staff/ewa/SeaMODISTerra/coverage_improvment.html.

TABLE II
MODIS AND SEAWIFS COMPARABLE STANDARD PRODUCTS

Standard Products	MODIS	SeaWiFS
Normalized water-leaving radiances (nLw) at visible wavelengths	nLw_412 nLw_443 nLw_488 nLw_531 nLw_551 nLw_667 nLw_678	nLw_412 nLw_443 nLw_490 nLw_510 nLw_555 nLw_670
Chlorophyll-a concentration	chlor_a_2 chlor_MODIS chlor_a_3	chlor_a
Aerosol optical thickness (AOT)	Tau_865	tau_865
Diffuse attenuation coefficient (K)	K_490	K_490
Aerosol radiance ratio between the two near-infrared bands 765nm and 865nm (ε)	Eps_78	eps_78

apply to high-quality chlorophyll-a concentration data from all three sensors. Overall, merged good quality daily data from MODIS on Terra and Aqua platforms and SeaWiFS provide about 38% of the global ocean and inland-water coverage. The other critical benefit of data merger is an increase in statistical confidence in extracted bio-optical parameters. Furthermore, fusion of data from multiple sensors can enable definition of a variety of ocean color products, including long-term time series and climatological datasets. Sensor-varying properties, such as spectral, spatial, temporal, and ground coverage characteristics, can be utilized by regional and local merger algorithms that can enable broadening the scope of ocean color applications [17].

VI. CONCLUSION

The objective of creating global ocean color datasets from multiple satellite sensors is important in the era of many concurrent ocean-observing missions on orbit. This goal is, however, hampered by incompatibilities in data products between the missions.

The investigations presented in this paper show that MODIS and SeaWiFS ocean color datasets reveal significant discrepancies which are dependent on sensor calibrations and operational characteristics. These discrepancies inhibit the creation of consistent daily global merged datasets from both sensors. This paper proposes and investigates the application of machine-learning techniques to cross-calibrate MODIS and SeaWiFS data. Support vector machines are trained to emulate SeaWiFS baseline chlorophyll from MODIS data. The ultimate objective is to produce joint MODIS and SeaWiFS daily global coverages which have the accuracy and the spatial and temporal consistency of SeaWiFS datasets.

Machine-learning regression is a promising tool for the data merger. Support vector machines are able to accurately learn complex relationships between MODIS and SeaWiFS data and effectively eliminate sensor discrepancies, such as seasonal trends, scan angle dependencies, and spatial variation.