Spatial Statistics 13 (2015) 1–20

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Filtering remotely sensed chlorophyll concentrations in the Red Sea using a space–time covariance model and a Kalman filter

Denis Dreano [a], Bani Mallick [b], Ibrahim Hoteit [a,*]

[a] Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Saudi Arabia

[b] Department of Statistics, Texas A&M University, United States

A R T I C L E I N F O

Article history:
Received 2 October 2014
Accepted 7 April 2015
Available online 27 April 2015

Keywords:
Space–time statistics
Covariance models
Kalman filter
Geostatistics
Chlorophyll concentration

A B S T R A C T

A statistical model is proposed to filter satellite-derived chlorophyll concentration from the Red Sea, and to predict future chlorophyll concentrations. The seasonal trend is first estimated after filling missing chlorophyll data using an Empirical Orthogonal Function (EOF)-based algorithm (Data Interpolation EOF). The anomalies are then modeled as a stationary Gaussian process. A method proposed by Gneiting (2002) is used to construct positive-definite space–time covariance models for this process. After choosing an appropriate statistical model and identifying its parameters, Kriging is applied in the space–time domain to make a one step ahead prediction of the anomalies. The latter serves as the prediction model of a reduced-order Kalman filter, which is applied to assimilate and predict future chlorophyll concentrations. The proposed method decreases the root mean square (RMS) prediction error by about 11% compared with the seasonal average.

* Corresponding author.
E-mail address: ibrahim.hoteit@kaust.edu.sa (I. Hoteit).

http://dx.doi.org/10.1016/j.spasta.2015.04.002

1. Introduction

The Red Sea is an elongated basin situated between the Asian and the African shelves, connected to the Mediterranean Sea in the north through the Suez Canal, and to the Gulf of Aden in the south through the Strait of Bab el Mandeb Yao et al. (2014). It is one of the warmest and most saline seas in the world, with a rich ecosystem that has adapted to these extreme conditions (Heileman and Mistafa, 2009). This unique natural resource is however threatened by an abrupt increase of temperature since 1994 (Raitsos et al., 2011).

Phytoplankton are small, unicellular, photosynthetic algae. They are the primary producers for marine ecosystems, and at the base of the marine food chain. Phytoplankton concentration is therefore important for fisheries (Lo-Yat et al., 2011). By fixing atmospheric CO_2 and sinking to form sediment at the bottom of the sea, phytoplankton also acts as a biological pump. This phenomenon is crucial to understanding climate change (Mann and Lazier, 2006). In the Red Sea, phytoplankton is particularly important to the extensive coral reefs along its shores.

The Red Sea is generally deficient in major inorganic nutrients, and its productivity is relatively low (Acker et al., 2008; Triantafyllou et al., 2014). The high productivity observed in the southern Red Sea is attributed to the intrusion of nutrient-rich waters from the Gulf of Aden (Acker et al., 2008; Triantafyllou et al., 2014). Red Sea chlorophyll concentrations follow seasonal patterns, with a winter bloom following a weak summer productivity. Considerable interannual variability in chlorophyll concentrations has also been observed (Raitsos et al., 2012). However, the Red Sea ecosystem has not yet been fully explored, and very few in-situ measurements have been collected in its basin (Brewin et al., 2013), increasing the need for remotely sensed data. Satellite observations of chlorophyll concentrations have been shown to be reliable datasets to study the primary productivity of the oceans (McClain, 2009) and, up to this point, they constitute the basis of several studies in the Red Sea (Acker et al., 2008; Brewin et al., 2013; Raitsos et al., 2012). Here we are interested in forecasting chlorophyll concentrations in the Red Sea, and we therefore need a dynamical model to mimic its evolution.

There are two approaches to chlorophyll modeling: deterministic or data-driven (statistical). A broad range of deterministic models has been developed by the marine ecosystem research community, from the very simple NPZ model with only nutrients, phytoplankton and zooplankton as state variables, to much more complex models, such as the European Regional Seas Ecosystem Model (ERSEM) (Baretta et al., 1995). The latter distinguishes functional phytoplankton and zooplankton groups, and models the complete cycling of different nutrient groups and O_2 and CO_2, including the effect of higher trophic groups. An example of a 3D ERSEM coupled model has recently been implemented in the Red Sea (Triantafyllou et al., 2014). Simulating the ecosystem with such a model can be however very challenging, as it requires the coupling to an ocean circulation model, which provides the physical forcing. Configuring these models further requires considerable efforts and expertise (Petihakis et al., 2002), because of the high number of parameters (over 50 for ERSEM (Blackford and Radford, 1995)). This makes such models difficult to calibrate and validate, since there are usually not enough observations to constrain the parameters (Anderson, 2005).

An alternative approach is to follow a statistical framework to model the space–time evolution of chlorophyll concentrations. Space–time statistical methods have not been used yet in this field, which has so far relied on time-series observations. Artificial neural networks were applied to forecasting algal blooms in freshwater and marine systems (Lee et al., 2003; Recknagel, 1997), and generalized additive models have been used for finding explanatory variables for the chlorophyll concentrations in the Pagasitikos Gulf and the subartic North Atlantic (Raitsos et al., 2012, 2006).

Geostatistical spatio-temporal models are extensions of the spatial classical geostatistical methods (Kyriakidis and Journel, 1999). These methods consider space–time data as the realization of a Gaussian process, from which a mean and a covariance function can be estimated. Although, in most applications, such a stochastic modeling approach is not based on a dynamical framework, geostatistical methods may capture some patterns in the data and avoid the difficulties of developing dynamical models (Cressie and Wikle, 2011; Kyriakidis and Journel, 1999). These methods have been widely employed in meteorology to model the surface temperature over land and oceans (Handcock and Wallis, 1994; North et al., 2011), in an ecological context to study moth populations