Ecological Informatics 43 (2018) 185–199

Contents lists available at ScienceDirect

Ecological Informatics

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

Extending geographically and temporally weighted regression to account for both spatiotemporal heterogeneity and seasonal variations in coastal seas

Zhenhong Du[a,b,*], Sensen Wu[a], Feng Zhang[a,b], Renyi Liu[a,b], Yan Zhou[c,d]

[a] School of Earth Sciences, Zhejiang University, 38 Zheda Road, Hangzhou 310027, China
[b] Zhejiang Provincial Key Laboratory of Geographic Information Science, 148 Tianmushan Road, Hangzhou 310028, China
[c] Zhejiang Fisheries Tech Extens Ctr, Hangzhou 310013, China
[d] Zhejiang Prov Acad Marine Sci, Hangzhou 310012, China

ARTICLE INFO

Keywords:
GcTWR
Spatiotemporal heterogeneity
Seasonal variations
Coastal seas
Chlorophyll-*a*

ABSTRACT

Space-time modelling has been successfully applied in numerous research projects and has been studied extensively in the field of geographical information science. However, the cyclical or seasonal variations in the temporal dimension of most spatiotemporal processes are rarely considered along with spatiotemporal nonstationarity. Seasonal variations are widespread and typical in marine environmental processes, and addressing both spatiotemporal heterogeneity and seasonal variations is particularly difficult in the turbid and optically complex coastal seas. By incorporating seasonal periodic effects into a geographically and temporally weighted regression (GTWR) model, we proposed a geographically and cycle-temporally weighted regression (GcTWR) model. To test its performance, modelling of chlorophyll-*a*, known as an important indicator of the coastal environment, is performed using the in situ data collected from 2012 to 2016 in the coastal sea of Zhejiang Province, China. GcTWR is compared with global ordinary least squares (OLS), geographically weighted regression (GWR), cycle-temporally weighted regression (cTWR), and GTWR models. In the results, the GcTWR model decreases absolute errors by 89.74%, 79.77%, 76.60% and 29.83% relative to the OLS, GWR, cTWR, and GTWR models, and presents a higher R^2 (0.9274) than the GWR (0.5911), cTWR (0.6465), and GTWR (0.8721) models. The estimation results further confirm that the seasonal influences in coastal areas are much more significant than the interannual effects, which accordingly demonstrates that extending the GTWR model to handle both spatiotemporal heterogeneity and seasonal variations are meaningful. In addition, a novel 3D visualization method is proposed to explore the spatiotemporal heterogeneity of the estimation results.

1. Introduction

Space and time are two fundamental dimensions pertaining to all geographic processes. Space-time analysis and modelling of geographic parameters has long been one of the main focuses of geographical information science (GIScience). Examples include investigating the spatiotemporal patterns of real estate prices (Fotheringham et al., 2015; Huang et al., 2010; Lu et al., 2014; Wu et al., 2014), environmental issues (Bai et al., 2016; Chu et al., 2015), land use (Wrenn and Sam, 2014), marine processes (Alam et al., 2016; Terry et al., 2013; Wang et al., 2015) etc. Although the temporal dimension has been incorporated into spatial analysis and modelling successfully in many research projects, cyclical or seasonal variations are rarely managed with spatiotemporal nonstationarity in most geographic processes.

Seasonal variations in marine environmental processes are widespread and typical, and require further exploration and researches (Dango, 2015; Khodse et al., 2007; Niu et al., 2015). Coastal seas are the richest marine regions in the world, and are the interfacial areas among the marine, terrestrial and aerial environments (Chen and Liu, 2015). Considering the key role of phytoplankton in biogeochemical cycles, phytoplankton biomass in terms of chlorophyll-*a* (Chl-*a*) is considered as the biological indicator of coastal environments and the most important element in coastal ecosystems (Paudel et al., 2016; Su and Weng, 1994). Therefore, investigating the spatiotemporal variations of Chl-*a* and understanding the interactions between Chl-*a* and other environmental factors are of great significance to recognize the ecological state of the coastal areas.

However, the variations of Chl-*a* are difficult to analyse and interpret due to the complicated interrelationships between phytoplankton and marine environmental factors (suspended matter, dissolved oxygen, nutrients, organic solute, etc.). Previous studies have established several global regression models to predict Chl-*a* concentrations. Celik (2006) used a multiple regression model to explore the relations between Chl-*a* and other

* Corresponding author at: School of Earth Sciences, Zhejiang University, 38 Zheda Road, Hangzhou 310027, China.
E-mail address: duzhenhong@zju.edu.cn (Z. Du).

https://doi.org/10.1016/j.ecoinf.2017.12.005
Received 16 July 2017; Received in revised form 9 December 2017; Accepted 11 December 2017
Available online 14 December 2017

water quality parameters (ammonium (NO_4), nitrate (NO_3) and phosphate (PO_4)). Partial least-squares regression (PLSR) was also employed to address the optimal count of factors that were suitable for estimating Chl-*a*, which was capable to account for 80% of the observed Chl-*a* variations (Ryan and Ali, 2016).

The hypothesis of space-time stationarity in global models is usually impractical since parameters tend to change across the research regions and over time. Therefore, some local models have been put forward to capture spatial variability in marine environmental processes (Freedman and Sen Roy, 2012; Keith et al., 2013). For instance, geographically weighted regression (GWR) model was formulated and fitted by Keith et al. (2013) to investigate the variations in the relations between macroalgae richness and environmental conditions over geographical regions. Furthermore, researches have been conducted to integrate temporal impacts into the GWR model to account for both spatial and temporal heterogeneities in recent years (Crespo, 2009; Fotheringham et al., 2015; Huang et al., 2010). For example, by extending the spatial distance to a spatiotemporal distance, a geographically and temporally weighted regression (GTWR) model was developed by Huang et al. (2010), which has been applied in various fields and has achieved significant performance. In 2015, Fotheringham et al. (2015) also proposed a new GTWR model to deal with local effects in both space and time, which was confirmed to be effective in the modelling of hedonic price.

However, few researches have been carried out on the quantitative effect of marine environmental processes in coastal seas on both spatial and temporal scales. Due to the complex optical and turbid characteristics of coastal regions, managing spatiotemporal nonstationarity in these areas is considerably difficult and challenging. Moreover, studies have shown that seasonal variations are quite significant in the coastal environment. For example, a long period of remote sensing data was used by Chen and Liu (2015) to demonstrate the temporal variations of Chl-*a* and suspended matter, which indicated significantly seasonal changes in the China's eastern coastal zones.

Therefore, to better understand the marine environmental processes in coastal seas, novel approaches should be put forward to deal with the spatiotemporal nonstationarity and seasonal variations simultaneously. By dividing temporal distance into seasonal periodic and interannual aperiodic parts, our study extends the GTWR model of Huang et al. (2010) and proposes a geographically and cycle-temporally weighted regression (GcTWR) model to capture cycle-temporal variations and spatiotemporal heterogeneity. In addition, an innovative 3D visualization method is proposed for presenting the spatiotemporal variations of the estimation results.

Our article is formed as follows. In Section 2, we describe the study area and data. The widely used GWR model is introduced in Section 3, followed up by the process of integrating GTWR with cycle-temporal variations to achieve the GcTWR model. The case study results and discussions of Chl-*a* modelling in the coastal sea of Zhejiang, China using GcTWR are presented in Sections 4 and 5. Furthermore, GcTWR is compared with global and other GWR-based models for performance examination. Finally, the study comes to an end with conclusions and summaries in Section 6.

2. Study area and data

2.1. Study area

The study area is situated in the Zhejiang coastal areas (ZCA) of the East China Sea (ECS), which is abundant in fishery resources and is a zone of frequent red tide events (Lou and Hu, 2014; Yang et al., 2013) (Fig. 1). It lies within a typical subtropical monsoon climate with four distinct seasons and is considerably affected by seasonal precipitation flows from the Yangtze River (Qiu et al., 2015), which results in strong seasonal variations in regional marine environmental processes. Every year, the Yangtze River transports about 240×10^6 t of sediment into the ECS, approximately 32% of which is stored in the ZCA and Fujian coastal region (Liu et al., 2006). In addition to the Yangtze River, several other rivers, including the Qiantang River, directly discharge large quantities of freshwater with high nutrients and sediments into the ZCA (Cong et al., 2014; He et al., 2013).

Tidal action in the ZCA is powerful and causes significant resuspension of sediment, especially in the Hangzhou Bay, one of the strongest tidal bays in the world. As a result, the ZCA has optically complex water and significant seasonal variations, which makes the spatiotemporal heterogeneities of its marine processes quite complicated.

2.2. Dataset

In situ data used in this paper were collected by the Marine Monitoring and Forecasting Center of Zhejiang Province (Hangzhou, China) through survey cruises and dip samples. The study data covered time period from 2012 to 2016 and were collected four times each year in winter (March), spring (May), summer (July–August), and autumn (October) with exception in 2012 and 2013. The detailed information of the dataset is shown in Table 1.

The spatial distributions of monitoring stations were different every year and the number has increased from 214 in 2012 to 309 in 2016. The stations of 2016 were displayed in Fig. 1 and a total dataset of 4820 observations was available (Table 1). The dataset provided full water quality parameters, e.g., suspended matter (SM), potential of hydrogen (PH), salinity (SAL), chemical oxygen demand (COD), dissolved oxygen (DO), ammonia nitrogen (NH_3), nitrate nitrogen (NO_3), nitrite nitrogen (NO_2), silicate (SiO_4), phosphate (PO_4), total phosphorus (TP), total carbon (TC), total nitrogen (TN), Chl-*a*, etc. Moreover, it also contained geographic coordinates and monitoring date information to enable our spatiotemporal analysis.

In the analysis process, Chl-*a* concentrations were used as the dependent variable while the explanatory variables included a total of 25 variables. Through correctional and multicollinearity analysis using SPSS 22.0 (Table 2), we found that DO, COD, TN and PO_4 were the most strongly correlated parameters with Chl-*a*, and had the lowest value of variance inflation factor (VIF). In addition, DO and COD are known as important indicators of phytoplankton respiration strength, while TN and PO_4 are fundamental nutrient substances for phytoplankton growth (Cole and Harmon, 1981; Steingrund and Gaard, 2005). Therefore, DO, COD, TN and PO_4 were chosen as the independent variable in our experiment.

3. Geographically and cycle-temporally weighted regression modelling

3.1. GWR model

The basis for the GWR methodology is that parameters in each point are estimated locally ground on distance-weighted subsampling at neighbouring locations (Brunsdon et al., 1998; Fotheringham et al., 2002). GWR model that takes the version of the Ordinary Least Squares (OLS) model is presented as follows:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \tag{1}$$

The index and coordinates of a spatial point are denoted as i and (u_i, v_i) in Eq. (1). Accordingly, y_i, x_{ik}, and ε_i represent the dependent variable, the kth independent variable and the error term for the ith point, respectively. $\beta_0(u_i, v_i)$ is the intercept term, and $\beta_k(u_i, v_i)$ stands for the coefficient of kth independent variable at location i, which are permitted to vary across space to capture spatial nonstationarity. The estimator using matrix representation can be expressed as:

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i)X)^{-1} X^T W(u_i, v_i)y \tag{2}$$

The $n \times n$ weights matrix associated with positions is represented by $W(u_i, v_i)$, with geographical weights in its leading diagonal and zeros in its off diagonal elements. A weighting function is established using the distance vector and a distance decay parameter such that neighbouring sample observations from the spatial data sample are allocated relatively more weight.