Estuarine, Coastal and Shelf Science 177 (2016) 1–7

Contents lists available at ScienceDirect

Estuarine, Coastal and Shelf Science

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

Prediction of cadmium enrichment in reclaimed coastal soils by classification and regression tree

Feng Ru [a], Aijing Yin [b], Jiaxin Jin [c], Xiuying Zhang [c], Xiaohui Yang [a], Ming Zhang [d], Chao Gao [a, *]

[a] School of Geographic and Oceanographic Sciences, Nanjing University, Nanjing, 210093, China
[b] Institute of Agricultural Resources and Environment, Jiangsu Academy of Agricultural Sciences, Nanjing, 210014, China
[c] International Institute for Earth System Science, Nanjing University, Nanjing, 210093, China
[d] Nanjing Center, China Geological Survey, Nanjing, 210000, China

ARTICLE INFO

Article history:
Received 17 September 2015
Received in revised form
15 January 2016
Accepted 27 April 2016
Available online 7 May 2016

Keywords:
Cadmium enrichment
CART
Reclamation
Coastal land

ABSTRACT

Reclamation of coastal land is one of the most common ways to obtain land resources in China. However, it has long been acknowledged that the artificial interference with coastal land has disadvantageous effects, such as heavy metal contamination. This study aimed to develop a prediction model for cadmium enrichment levels and assess the importance of affecting factors in typical reclaimed land in Eastern China (DFCL: Dafeng Coastal Land). Two hundred and twenty seven surficial soil/sediment samples were collected and analyzed to identify the enrichment levels of cadmium and the possible affecting factors in soils and sediments. The classification and regression tree (CART) model was applied in this study to predict cadmium enrichment levels. The prediction results showed that cadmium enrichment levels assessed by the CART model had an accuracy of 78.0%. The CART model could extract more information on factors affecting the environmental behavior of cadmium than correlation analysis. The integration of correlation analysis and the CART model showed that fertilizer application and organic carbon accumulation were the most important factors affecting soil/sediment cadmium enrichment levels, followed by particle size effects (Al_2O_3, TFe_2O_3 and SiO_2), contents of Cl and S, surrounding construction areas and reclamation history.

1. Introduction

Reclamation of coastal land is currently the main approach to gaining arable land (Airoldi and Beck, 2007; Doody, 2004). However, it has long been acknowledged that anthropogenic interference with coastal land has a variety of detrimental effects (Flemming and Nayandwi, 1994), such as littoral sea eutrophication, ecosystem degradation and heavy metal contamination (Chen and Jiao, 2008; Dejonge et al., 1993; Hodoki and Murakami, 2006). Heavy metal contamination is one of the most dangerous types of pollution because of its diverse sources, innate toxicity, persistence and non-degradability (Meng et al., 2008; Wang et al., 2013). As one of the most hazardous trace elements, the toxicity of cadmium received intensive attention since the outbreak of a severe disease in Japan in the early 20th century (Waisberg et al., 2003). The concentration of cadmium in soil is highly differential and predominantly depends on the inputs of human activities, including industrial production, agricultural practice and waste discharge (Luo et al., 2009; Nicholson et al., 2003).

China has one of the longest coastlines in the world, and these coastal areas often have high population density and intense human activities. The rapid development of Chinese coastal areas is accompanied by a serious environmental problem with regard to cadmium enrichment and pollution (Meng et al., 2008; Pan and Wang, 2012), which are partially attributed to coastal reclamation for cultivation and aquaculture. The practices, such as excessive and inappropriate application of fertilizer and fish feed, as well as wastewater irrigation, lead to the deteriorating situation of cadmium accumulation and contamination in these regions (He et al., 2005; Liu et al., 2011; Onsanit et al., 2010; Zhang et al., 2012).

In general, cadmium threatens the ecological environment and human health via the food chain and ground water (Wang et al., 2002), and intake through the soil-food system is one of the

* Corresponding author.
E-mail address: chgao@nju.edu.cn (C. Gao).

http://dx.doi.org/10.1016/j.ecss.2016.04.018
0272-7714/

major pathways for human exposure to environmental cadmium in agricultural areas (Wei and Yang, 2010). Thus, prediction of cadmium enrichment following reclamation is of vital importance for the sustainable management of coastal land and security of food production. Classification and regression tree (CART) analysis, a suitable and practical prediction model, has emerged and been widely applied in remote sensing image processing to classify land use and make thematic maps (de Colstoun et al., 2003; Pal and Mather, 2003) as well as being used in soil science in recent years for modeling the role of soil properties in influencing heavy metal concentrations (Vega et al., 2009, 2010) and simulating heavy metal concentrations in urban areas (Zhang et al., 2008a, 2008b). This method has outstanding features, including freedom from assumptions of normality, the ability to cope with variables with internal relationships and higher-order interactions, and analysis results that are easy to understand and interpret (Bruland and DeMent, 2009).

The DFCL in eastern China is a typical reclamation region for agriculture and aquaculture faced with the risk of cadmium enrichment. Taking account of the interaction of natural and anthropogenic influence with cadmium concentrations, we assumed that the CART model is appropriate for prediction of cadmium enrichment in this region. The primary objectives of this study were to (1) reveal possible environmental factors that impact cadmium concentrations in reclaimed soils and sediments and (2) validate the effectiveness of predicting cadmium enrichment levels using CART.

2. Materials and methodology

2.1. Study area

The study area (Dafeng Coastal Land, DFCL) is located between latitudes 30°00′ and 33°30′N and longitudes 120°30′ and 121°00′E in the eastern part of Dafeng County, Jiangsu Province, adjacent to the Yellow Sea (Fig. 1). The study area, covering an area of approximately 1900 km^2, was divided into five zones in this study based on the time of embankment. The history of land reclamation was obtained by the interpretation of satellite images and by reviewing the local historical literature (Fig. 1). Zone A was a current natural tidal flat, normally 2–3 km in width. Zones B, C, D, and E (in order from tidal flat to inland side) were reclaimed lands with conversion periods of approximately 10, 30, 60, and 90 years, respectively. The main land use was aquaculture in Zone B. This zone was dominated by fishing ponds of various sizes. In Zones C, D, and E, reclaimed lands were intensively used for farming. Major crops in these zones included rice (*Oryza sativa*), wheat (*Triticum aestivum*), barley (*Hordeum vulgare*) and cotton (*Gossypium* spp).

2.2. Sampling and chemical analysis

Surficial soil and sediment samples were collected from 227 geo-referenced locations in the DFCL in 2012. At each sampling site, five randomly distributed samples were taken at depths of 0–20 cm and then combined to obtain a more representative compound sample. After collection, samples were air-dried and sieved to 2 mm before chemical analysis.

Cadmium (Cd) concentrations in sediment and soil samples were measured by ICP-MS after grounding and digestion. Elemental concentrations of Si, Al, Fe, K, C, N, P, S and Cl were measured using X-ray fluorescence spectrometry (XFS) after 5 g of each dried sample was ground to <200 mesh in an agate mortar and compacted into a disc. Organic matter (OM) content was determined by wet combustion method using the mixture of potassium dichromate and concentrated H_2SO_4, followed by titration with a standard solution of ammonium ferrous sulphate (Lu, 1999). Soil/sediment pH was determined with a glass electrode in 1:2.5 soil-to-water ratios. Certified reference material of Chinese national geostandard sample (GSS28) was used to validate the analysis. Recoveries for Cd ranged from 95 to 119%, whereas they were within 92–108% of the certified values for other items. One in every ten samples was randomly selected for repeated measure to ensure precision and the differences were all within 10%.

2.3. Classification and regression tree (CART)

Considering the joint influence of natural and anthropogenic factors on Cd concentrations in soils and sediments of the DFCL, it is necessary to adopt a suitable and practicable method to predict Cd enrichment in soils and sediments. The CART model is a suitable choice because it could not only predict Cd enrichment levels but also reflect the further relationships between Cd and other parameters (Bruland and DeMent, 2009; Zhang et al., 2008a) which could not be reflected by statistical analysis. The rules for deducing levels of Cd enrichment could be attained from the sample data through the CART model, in which environmental factors were selected as independent variables and levels of Cd enrichment were dependent variables. The precondition of CART was the selection of independent variables, and the statistical analysis, fieldwork and remote sensing interpretation were used for this purpose. The data used for CART analysis were processed with the software S-PLUS 8.0 for Windows 7, and the data preparation for CART analysis was conducted using the following scheme.

2.3.1. Dependent variable

In CART, Cd enrichment levels need to be classified because the CART model can only obtain ranked data rather than exact values. In this study, the enrichment level index (I_n) was used to assess the Cd level in the soils/sediments of the DFCL. If $I_n \leq 1.00$, it showed that there was no Cd accumulation; otherwise, it meant that there was Cd enrichment in the soil or sediment. The enrichment level index was calculated as follows:

$$I_n = C_n/B_n$$

where C_n is the measured concentration of Cd in the soil or sediment (n) in units of $\mu g\ kg^{-1}$ and $I_n = 90.0\ \mu g\ kg^{-1}$ is the geochemical background value (Chi and Yan, 2007). Then, I_n was divided into six levels (Table 1).

2.3.2. Independent variables

In this study, the independent variables consisted of three portions based on the assumption of possible affecting factors based on different approaches to data acquisition.

On the basis of the Pearson's correlation analysis, the soil/sediment properties and element concentrations that were significantly correlated with Cd concentrations were selected for further analysis by the CART model. These possible factors affecting Cd concentrations in soils and sediments of the DFCL could be validated and selected by the CART model to construct the prediction rules of Cd enrichment levels. Moreover, compared with the direct measurement of Cd concentrations, these element concentrations and soil/sediment properties were conveniently measured.

Then based on the reclamation history, the DFCL was divided into five sampling zones: I (corresponding to Zone A), II (B), III (C), IV (D) and V (E). It has been demonstrated that Cd shows an important increase in soils with a history of increasing agricultural activity (Andreu and Gimeno-Garcia, 1999; Mann et al., 2002). We inferred that different reclamation histories may influence Cd concentrations in soils and sediments of the DFCL.