Available online at www.sciencedirect.com

ScienceDirect

Procedia Environmental Sciences 3 (2011) 105–110

Procedia Environmental Sciences

1st Conference on Spatial Statistics 2011- Mapping Global Change

Modeling spatial-temporal change of Poyang Lake marshland based on an uncertainty theory - random sets

Xi Zhao[ab]*, Alfred Stein[a], Xiaoling Chen[bc], Lian Feng[b]

[a]*Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede,7500AE, The Netherlands*
[b] *State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, ,430079, China*

[c]*Key Laboratory of Poyang Lake Ecological Environment and Resource Development, Jiangxi Normal University, Nanchang 330027, China*

Abstract*

Uncertainty modeling now engages the attention of researchers in spatial temporal change analysis in remote sensing. Some studies proposed to use random sets for modeling the spatial uncertainty of image objects with uncertain boundaries, but none have considered the parameter determination problem for large datasets. In this paper we refined the random set models for monitoring monthly changes in wetland vegetation areas from series of images. Twelve cloud-free HJ-1A/1B images from April 2009 to March 2010 were used for monitoring spatial-temporal changes of Poyang Lake wetlands. We applied random sets to represent spatial uncertainty of wetland vegetation that were extracted from normalized difference vegetation index (NDVI) maps. Time series of random sets reflect the seasonal differences of location and extents of the wetlands, whereas degree of uncertainties indicated by *SD* and *CV* indices reflect the gradual change of the wetland vegetation in space. Results show that the uncertain extents of wetland vegetation change through the year, achieving the largest range and uncertainty degree in autumn. This coincides with the highly heterogeneous vegetation status in autumn, since the wetland recovers gradually after flooding and young vegetation emerges at gradually changing densities, thus providing forage in different ecological zones for different types of migratory birds. We conclude that the random set model enriches spatial-temporal modeling of phenomena which are uncertain in space and dynamic in time.

Selection and/or peer-review under responsibility of Spatial Statistics 2011
Keywords: Uncertainty; random set; spatial-temporal change; Poyang Lake

1. Introduction

Poyang Lake is the largest fresh water lake in China and its seasonal flooding creates a large extent of wetland which provides important forages and habitats for migratory birds. A better understanding of the dynamic of wetlands is important for the wetland ecosystem monitoring and habitat assessment [1]. Due to the gradual transition of elevation from vegetated area to open water area as well as seasonal water level dynamics, the wetland in Poyang Lake has gradual and unclear boundaries in space and is dynamic in time [2]. Such spatial uncertainty varying with time, therefore, needs to be considered when precisely extracting wetlands from remote sensing images and representing them in spatial temporal modeling. Hui et al. [3] studied the spatial temporal dynamics of water coverage and inundation of Poyang lake wetlands by analyzing a temporal sequence of remote sensing images. The temporal

* Corresponding author. Tel.: +86-15972171719
E-mail address: xzhao@itc.nl, zhaoxichina@gmail.com.

1878-0296
doi:10.1016/j.proenv.2011.02.019

sequence of water extent was determined from images by means of crisp classification methods, but spatial uncertainty has not been taken into account.

A random set method was proposed by Zhao et al. [4] for modeling the boundary extensional uncertainty of spatial objects extracted from images. The methods were based on the assumption that interpreting boundaries e.g. by digitizing, thresholding or segmentation, with more uncertainties/vagueness will result in bigger differences and more randomness in different interpreting outputs, so the overall set of outputs can be considered as a random set whose variability can reflect the degree of uncertainty. The previous study focused on modeling the extensional uncertainty of objects at certain time point. The aim of this research is to examine the application of random sets to monitor the multi-temporal changes of the Poyang Lake wetland with uncertain spatial extents.

2. Study area and image data

Poyang Lake (28° 22′ - 29° 45′ N, 115° 47′ -116° 45′ E) is the largest freshwater lake in China, located in the middle reach of Yangtze River. It exchanges water with Yangtze River through the 40 km long water channel in the northern part, and weakens flood peaks by storing discharges from five tributaries: River Gan, Xiu, Xin, Fu and Rao. The biggest magnitude of fluctuation between high and low water level of Poyang Lake is 16.68 m: the highest water level is 22.58m and the corresponding water surface is approximately 4070 km^2, whereas the lowest is 5.9m and the water surface area is about 30 km^2 [5]. Interactions between the lake, rivers and surrounding land make an aquatic-terrestrial ecosystem in the southern part of the Poyang Lake.

HJ data, acquired by HJ-1A/1B, were obtained from the China Centre for Resource Satellite Data and Applications (CRESDA). Twelve HJ images were chosen based on the availability, geographic coverage, and cloud cover. Nine of them were on 2009: April 14, May 15, June 19, July 17, August 19, Sep 3, Oct 6, Oct 24, Nov 24, and three were in 2010: Jan 13, Feb 21, March 19. After the atmospheric correction and image-to-image geo-correction, all the HJ images were then spatially subset to the Poyang Lake region. Due to limited optical bands that HJ image has, we chose the most commonly used vegetation index: Normalized Difference Vegetation Index (NDVI) to map vegetation and differentiate water from them. These NDVI maps are the inputs of the further processing.

3. Methods

3.1. Refining random set model in spatial-temporal analysis

A grey-scale NDVI image inside a window W can be treated as a function f: $W \rightarrow$[-1, 1]. Each pixel $x \in W$ has a grey value $f(x)$ which lies between -1 and 1. The NDVI image can also be interpreted as a collection of sets $F = \{x \in W: f(x) \geq R\}$, $0 \leq R \leq 1$. F becomes a random set when R is a random variable on [-1, 1]. The distribution of a random set is uniquely determined by f and the random variable R [6]. Two types of distribution were used in our previous research [7]: the uniform distribution and the normal distribution. The uniform distribution can be chosen when no emphasis needs to put on special values, whereas choice for the normal distribution is based on the assumption that pixel values close to object boundary has more chance to be selected by the image interpreters when labeling the boundary pixels than the values far away [7]. We adopt the terms proposed by Friel and Molchanov [6], and call such random set models uniformly weighted and normally weighted respectively. Demining the parameters of the models, however, is more challenging in spatial-temporal analysis. In [4], the mean and standard deviation of the normal distribution are user-defined parameters and was determined by pre-experiments in that specific case. It is difficult, however, to determine the parameters by users or based on collecting training data for each image, mainly because of the large volume of image sets in spatial-temporal analysis.

Another possibility is to consider the histogram of the whole or part of the image as the distribution of the variable. In this case, we call it a histogram weighted random set model. Instead of treating the whole image as a random set, we consider the area of interest (denote as *AOI*) in an image as a random set. Different realizations of a random set are generated by thresholding or segmenting *AOI* by using a random variable *PAR* as the parameter in the thresholding or segmenting algorithm. Suppose pixel values in the area of interest fall in the range [a, b], $-1 \leq a \leq b \leq 1$, *PAR* then follow the distribution of the histogram of the image pixels in [a, b]. To determine a and b, all the pixels on an image were devided into three classes: pixels with high NDVI for grassland, pixels with low NDVI for open water area, and pixels falling in-between for uncertain boundary areas, e.g. marshland connecting grassland to open water. Consequently, the lower and upper limits of the marshland determine the range [a, b]. The natural breaks Jenks method is based on Fisher-Jenks algorithm and the general idea is to identity breakpoints between classes by minimizing the sum of the variance with each of the classes [8].

After obtaining several realizations of a random set, we estimate statistical moments of the random set such as the mean set and standard deviation to quantify the uncertainty of *AOI*. We chose the covering function to characterize the distribution of the random set Γ. The covering function $\Pr_\Gamma(x)$ at pixel x is defined as $P(\{x\} \cap \Gamma \neq 0) = P(x \in \Gamma)$. It can be estimated as $\hat{\Pr}_\Gamma(x) = (1/n)\sum_{i=1}^{n} I_{O_i}(x), x \in R^2$, where O_i are realizations of Γ and $I_{O_i}(x)$ is the indicator function of $O_i(x)$. Pixels with $\Pr_\Gamma(x) \geq p$ construct a p-level set of Γ. The 0-level set, the 0.5-level set and the 1-level set are called the support set (Γ_s), median set and core set (Γ_c), respectively. The set-theoretic variance of Γ equals $\Gamma_{var}(x) = \mathrm{E}(I_{Oi}(x) - \Pr_\Gamma(x))^2$. Two indicators, *SD* and *CV*, derived from $\Gamma_{var}(x)$ quantify the extensional uncertainty of *AOI*. The *SD* is the sum of Γ_{var} over Γ_s as $SD = \int_{\Gamma_s} \Gamma_{var}(x)dx$, whereas the *CV* is the coefficient of variation as $CV = \int_{\Gamma_s} \sqrt{\Gamma_{var}(x)}dx / \int_{\Gamma_s} \Pr_\Gamma(x)dx$. A higher value of the indicator corresponds to a larger extensional uncertainty, which means the boundaries of *AOI* have more gradual transitions to the background [4].

We used the natural breaks Jenks method in ArcGIS to partition the image histogram into the three classes and estimate the uncertain intervals. The random sets generation and statistic analysis, as well as producing average difference map were programmed and implemented in ENVI/IDL software.

3.2. ODF mean set for hardening the uncertain boundaries

We considered the uncertain *AOI* as random sets in the modeling. Since crisp results are required to compare with the uncertainty modeling results, the mean set of a random set can be adopted. In this study, we chose the mean set based on oriented distance function (ODF), because ODF also consider the distance of each pixel to Γ [9]. All pixels in the crisp classification are classified both according to their grey value as well as to their positions with respect to *AOI*.

Let $O_1, \ldots, O_n$ be the observed realizations of Γ. We denote the ODF of O_i at pixel x as $b_{O_i}(x)$: $b_{O_i}(x) = d_{O_i}(x) - d_{O_i^c}(x), x \in R^2$, where $d_{O_i}(x)$ and $d_{O_i^c}(x)$ is the distance function from x to O_i and the complement of O_i, respectively. We use the Euclidean distance function that $d_{O_i}(x)$ equals the distance from x to the nearest point of O_i. Let $\hat{b}_O(x) = (1/n)\sum_{i=1}^{n} b_{O_i}(x)$ be the sample mean ODF at pixel x. The estimated mean set is consequently given by the zero-level set of the empirical ODF: $\hat{\Gamma}_{ODF} = \{x : \hat{b}_O(x) \leq 0\}$.

3.3. Detecting and analyzing changes in grasslands

The interval $[a^{(t)}, b^{(t)}]$ indicates the range of pixel values classified as marshland on the image at moment t. The observed grassland have vague extensions in space and the status of grassland changes during the seasons resulting in changing of the reflectance or NDVI values of the pixels. Therefore, $[a^{(t)}, b^{(t)}]$ is not necessary to be the same for a time series of images and be precise for each image. Next, histogram distributed random set models were built for each image: $\Gamma^{(1)}, \Gamma^{(2)}, \ldots, \Gamma^{(t)}$. Here we considered using time t only as the attribute. The covering function $\hat{\Pr}_\Gamma(x,t)$, set-variance $\Gamma_{var}(x,t)$, oriented distance function $\hat{\Gamma}_{ODF}^{(t)}$ and indicators $SD^{(t)}$ and $CV^{(t)}$ were defined for $\Gamma^{(t)}$.

The area of $\Gamma^{(t)}$ was summarized by plotting areas of 100 p-level sets, including $\Gamma_s^{(t)}$, $\Gamma_c^{(t)}$ and $\hat{\Gamma}_{ODF}^{(t)}$ in a boxplot. By comparing the boxplots of $\Gamma^{(1)}, \Gamma^{(2)}, \ldots, \Gamma^{(t)}$, we can explore how the amount of grassland areas, water areas and uncertain boundary areas change through time. $SD^{(1)}, SD^{(2)}, \ldots, SD^{(t)}$ and $CV^{(1)}, CV^{(2)}, \ldots, CV^{(t)}$ are plotted with time order, to see how the extensional uncertainties of grasslands change with time in different months and seasons.

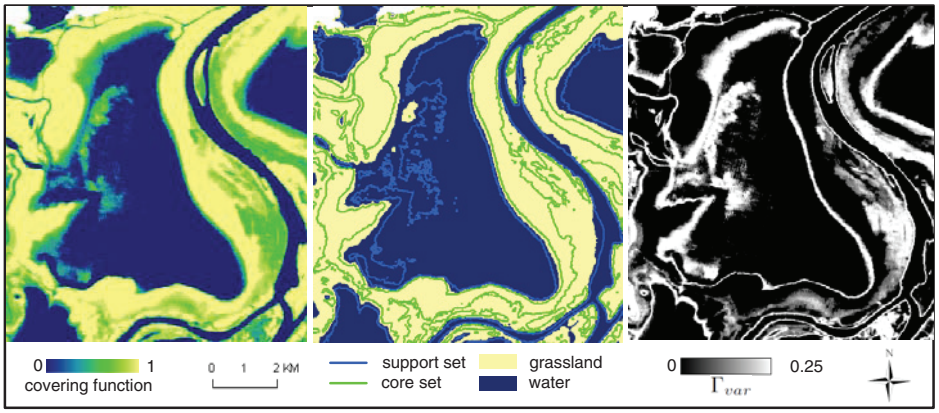

Fig. 3 The histogram distributed random set model for image I3. Part of the study area around Banghu lake is shown, with the covering function, contours of support set and core set (left), the crisp classification by the ODF mean set (middle), and the set-theoretic variance (right)

Fig. 2 Monthly change of wetland vegetation in Poyang Lake represented by random sets

4. Results

Fig. 1 shows an example in a subset of the study area. The covering function of a pixel indicates the probability of that pixel belonging to the random set for grassland. In this real case, we put 0.95-level set and 0.05 level set as Γc and

Γs respectively. Pixels located in Γc have more than 95% probability to be classified as grassland and those located within Γs have more than 5% probability. Pixels outside Γs are considered surely belonging to water class. The middle part of Fig. 3 show the crisp classification in which all pixels are assigned to be water or vegetation, by the ODF mean set method. The crisp boundary between water and vegetation are smooth, and locate between the contours of Γc and Γs.

All maps in Fig. 2 show a spatial-temporal pattern of wetland grassland and open water from spring to winter. Wetland vegetation dynamics are highly related to seasonal extent of the water body in this environment. Time series of covering functions of random sets show that there is either uni-model or bi-modal growing seasons discernable, depending on the elevation of the vegetation grow. The largest flooding extent appears on August, so the vegetated areas shown in yellow on August are mature on high elevations and remain emerge above the water during the summer flooding (June, July, August). They are off peak in other seasons, e.g. becoming senescent in winter and shown in cyan on I10. Vegetation on low elevations, usually near water or on the shore of lakes has another growing peak in autumn beside spring. In autumn, large land areas emerge following the gradually decreased water level, and vegetation such as *Carex* flooded beneath now turn green and become mature, whereas young *Carex* are flush on the newly exposed lake banks. This gradual emerging of land on different elevation makes vegetation growing status (including density, height and greenness) be different and thus forms diverse vegetation zones for different kinds of migratory birds foraging in their favorite zones. By looking at the areas of modest covering functions, we find the vegetation statuses are much more diverse in autumn and winter (I10 to I12) than those in summer (I7 to I8), and slightly more than those in spring (I4 to I5).

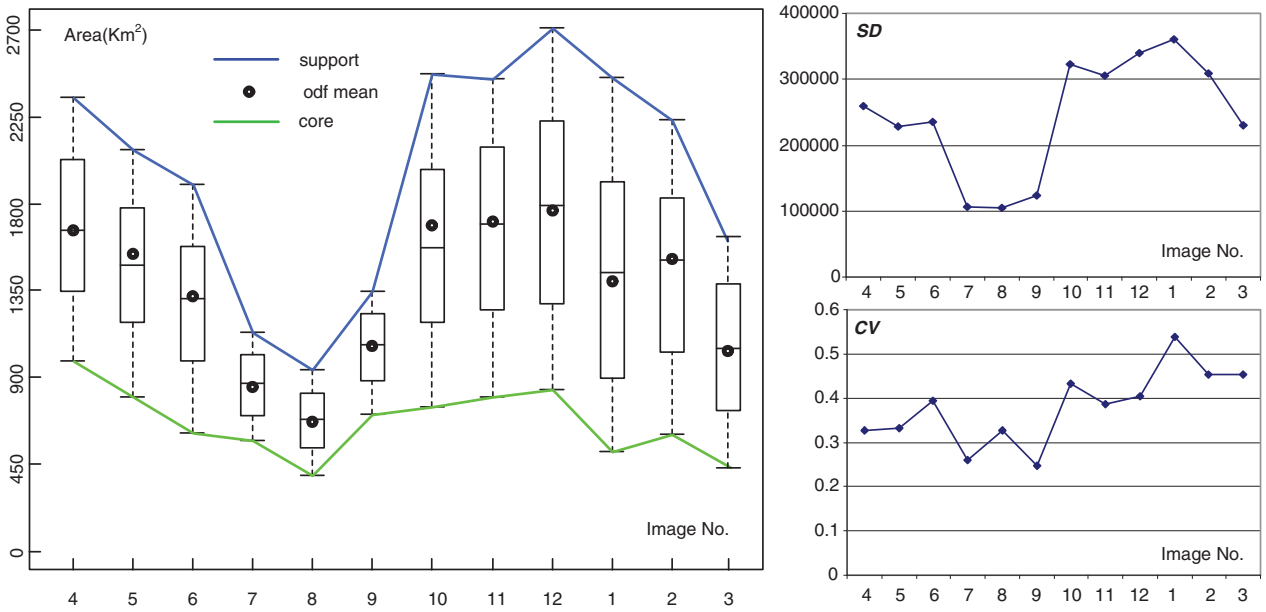

Fig. 3 Trend of the area change of support set, ODF mean set and core set in random sets, which presenting the dynamic of the vegetation and wetland with uncertain boundries (left); trend of the uncertainty indicators *SD* (top right) and *CV* (bottom right)

From Fig. 3 we can see the areas of grassland have different ranges in different months. In the summer months, areas of grassland have smallest ranges, but have large ranges in autumn. This coincides with our observations from Fig. 2. The ODF mean sets have similar areas with the median sets. But it does not necessarily mean that they have similar spatial distribution. The green line for Γ_c show how the areas of high green biomass changes through time. For example, the grassland area with high biomass in I4 is more than that in I10. This coincide with the fact that all kinds of wetland vegetations turn mature and have high density in April, but only few kinds of vegetation grow in autumn with high biomass. The area of Γ_s in I4, however, is smaller than that in I10. This is because young sedges with very low density are flush on the shore gradually in October after the flooding gone, these areas are counted in the support set area. After I12, the range of uncertain area should decrease, due to the senescence of vegetation. The abnormal high range in I1 and I2 correspond to the heavy image noises.

On the right part of Fig. 3, the curves of indicators *SD* and *CV* have similar trends: a sudden decrease from I6 to I7 and stay at relatively low level and then jump to a high level in I10 and remain high until I2. The differences between these two curves are due to the absolute uncertainties and uncertainties per unit area. For example, from I3 to I4, the

SD increases because the total amount of uncertain area increase, whereas *CV* decreases because the increased amount of uncertainties is less than that of the increased total vegetation area, including the certain part and uncertain part. High values of both indicators from I10 to I2 show that the wetland in autumn has vegetations with gradually changing status: greenness, density and height. The valley of the curves in I7, I8 and I9 indicate grasslands are homogeneous, because only vegetation on high elevations emerges above the flood and they become mature in summer. These coincide with our observation in Fig. 2. Therefore, *SD* and *CV* serve to quantify and summarize the status of different vegetation zones as numerical indicators.

5. Discussion and conclusions

The uncertain spatial extents of the Poyang Lake wetland are changing through years. Several statistic moments of random sets and two uncertainty indices derived from them were used to reflect the vagueness, diversity and heterogeneity in the grassland boundary zones. Series of covering functions and trend lines of uncertainty indices show the features of seasonal dynamic of wetlands and coincide with our local knowledge. We conclude that the random set model is suitable and applicable for spatial temporal modeling of phenomena which are uncertain in space and dynamic in time.

The essential part of building random set model is to generate realizations which can fully characterize the distribution of random set. In this study, the classified grassland on one image was modeled as a random set, and a realization of it was a binary set which partitions the grassland from the non-grassland class by a threshold. Due to the spatial uncertainty of the grassland on an image, the threshold should vary in a reasonable interval and follow a certain underlying distribution e.g. normal or uniform. To adjust the random set model in large image dataset context, we chose histogram distributed model and use natural breaks Jenks method to determine the threshold interval. More advanced method can be applied to find narrow the interval and thus result in more precise random set.

Acknowledgements

This Research was funded by NSFC projects (Grant Nos.41021061, 41071261，40910191). The authors would like to thank CRESDA for providing HJ images and support from Planet Action.

References

1. Lyon, J.G., *Wetland landscape characterization: techniques and applications for GIS, mapping, remote sensing and image analysis*. 2001: Ann Arbor Press.
2. Wu, Y. and W. Ji, *Study on Jiangxi Poyang Lake national nature reserve*. 2002, Beijing: Forest Publishing House.
3. Hui, F., et al., *Modelling spatialtemporal change of Poyang Lake using multitemporal Landsat imagery.* International Journal of Remote Sensing, 2008. **29**(20): p. 5767-5784.
4. Zhao, X., et al., *Quantification of Extensional Uncertainty of Segmented Image Objects by Random Sets.* IEEE Transactions on Geoscience and Remote Sensing, 2011. **In Press**.
5. Hu, X. and X. Xiong, *water level character and wetland ecological conservation of Poyang Lake.* Jiangxi Forestry Science and Technology (In chinese), 2002. **4**: p. 1-6.
6. Friel, N. and I. Molchanov, *A new thresholding technique based on random sets.* Pattern Recognition, 1999. **32**: p. 1507-1517.
7. Zhao, X., A. Stein, and X. Chen, *Application of random sets to model uncertainties of natural entities extracted from remote sensing images.* Stochastic Environmental Research and Risk Assessment, 2010. **24**: p. 713-723.
8. Slocum, T.A., *Data Classification*, in *Thematic Cartography and Visualisation*. 1999, Prentice Hall: New Jersey. p. 61-83.
9. Jankowski, H. and L.I. Stanberry, *Expectations of random sets and their boundaries using oriented distance functions.* Journal of Mathematical Imaging and Vision, 2010. **36**: p. 291-303.