ISPRS Journal of Photogrammetry and Remote Sensing 117 (2016) 227–236

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Spatio-temporal change detection from multidimensional arrays: Detecting deforestation from MODIS time series

Meng Lu [a,*], Edzer Pebesma [a], Alber Sanchez [a,b], Jan Verbesselt [c]

[a] Institute for Geoinformatics, Westfälische Wilhelms-Universitt Münster (WWU), Heisenbergstraße 2, 48149 Münster, Germany
[b] National Institute for Space Research (INPE), Av. dos Astronautas, 1758, 12227-007 São José dos Campos, Brazil
[c] Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, Droevendaalsesteeg 3, Wageningen 6708 PB, The Netherlands

ARTICLE INFO

Article history:
Received 8 June 2015
Received in revised form 19 January 2016
Accepted 9 March 2016
Available online 22 March 2016

Keywords:
BFAST
Time series analysis
Spatial correlation
Temporal correlation
Array database
Spatio-temporal change modeling

ABSTRACT

Growing availability of long-term satellite imagery enables change modeling with advanced spatio-temporal statistical methods. Multidimensional arrays naturally match the structure of spatio-temporal satellite data and can provide a clean modeling process for complex spatio-temporal analysis over large datasets. Our study case illustrates the detection of breakpoints in MODIS imagery time series for land cover change in the Brazilian Amazon using the BFAST (Breaks For Additive Season and Trend) change detection framework. BFAST includes an Empirical Fluctuation Process (EFP) to alarm the change and a change point time locating process. We extend the EFP to account for the spatial autocorrelation between spatial neighbors and assess the effects of spatial correlation when applying BFAST on satellite image time series. In addition, we evaluate how sensitive EFP is to the assumption that its time series residuals are temporally uncorrelated, by modeling it as an autoregressive process. We use arrays as a unified data structure for the modeling process, R to execute the analysis, and an array database management system to scale computation. Our results point to BFAST as a robust approach against mild temporal and spatial correlation, to the use of arrays to ease the modeling process of spatio-temporal change, and towards communicable and scalable analysis.

1. Introduction

Advanced earth observation satellite sensors provide remote sensing products that are rich in spatial, temporal, and spectral information. Open access policies of space agencies and the progress of remote sensing technologies make these products more accessible, which enables a wide range of novel applications, such as near real-time global change monitoring. This, however, calls for efficient handling and scalable processing of the massive amounts of available data. Major challenges include big data management, multidimensional data information extraction, and complex large-scale spatio-temporal change modeling procedures implementation and result visualization. These challenges call for novel data management and analytics tools and advanced spatio-temporal statistical algorithms.

Typical remote sensing satellite images are regularly discretised in space and time, and can naturally be represented as multidimensional arrays. The array data structure facilitates change modeling in many ways. Firstly, the array data structure allows a clean data processing procedure which simplifies data preparation, and avoid data structure conversions during the analysis. Wickham (2014) calls the unified data preparing process to "tidy data", and suggests restructuring all datasets into single, long tables. Since most earth observation data (i.e. earth information collected by remote sensing technologies) come as time series of multispectral images, and structuring such datasets into arrays is the more natural approach for data storage, analysis and visualization. In addition, the array data structure allows flexible application of spatio-temporal statistical algorithms (Zscheischler et al., 2013) and other information extraction methodologies (Mello et al., 2013), which was already exploited in the on-line analytical processing (OLAP) approach to analyze business data (Chaudhuri and Dayal, 1997; Viswanathan and Schneider, 2011). Finally, the array data structure facilitates parallelizing of the modeling process (Stonebraker et al., 2013). Array Data Management and Analytics Software (DMAS), which stores and operates on data as multidimensional arrays, can thus be used to scale the process and resolve the difficulties of large memory consumption and computational bottlenecks usually found in non-parallelized systems.

* Corresponding author.
E-mail address: meng.lu@uni-muenster.de (M. Lu).

http://dx.doi.org/10.1016/j.isprsjprs.2016.03.007
0924-2716/

Examples of array DMAS include SciDB (Cudre-Mauroux et al., 2009) and rasdaman (Baumann, 1994).

Remotely sensed image time series analysis (Verbesselt et al., 2010; Broich et al., 2011) has been drawing more attention in pixel-based change detection in recent years (Jianya et al., 2008; Banskota et al., 2014) due to the increased availability of long-term satellite image time series and improved computational power. Statistically, these methods can be classified as detecting change in mean (Kuan and Hornik, 1995), (e.g. by tests based on OLS (Ordinary Least Squares) residuals such as CUSUM (Cumulative Sum) test (Brown et al., 1975)), or change in regression parameters, (e.g. by tests that assess all regression coefficients such as supLM (supremum Lagrange Multiplier) test (Andrews, 1993; Zeileis and Hothorn, 2013)). Change detection with time series imagery solves many problems that are infeasible with bi-temporal analysis (Coppin et al., 2004; Jianya et al., 2008). There are several examples: (1) image time series analysis enables detection of unknown historical changes retrospectively, and monitoring of changes in near real-time (Verbesselt et al., 2012); (2) image time series analysis is able to classify land cover types that are of subtle differences in reflection. For example, one difficulty in analyzing tropical forest conservation from remotely sensed imagery *pairs* is to discriminate plantations from secondary forests (Lucas et al., 1993); (3) the regression model is flexible, and can integrate variables that will affect the process. For example, it is hard to distinguish between climate-induced forest drought and anthropogenic deforestation. Integration of climate variables, such as precipitation and temperature, can assist differentiating between these changes (Dutrieux et al., 2015); (4) in terms of reliability, satellite image time series analysis has the advantage of being more resistant to noise (Coppin et al., 2004).

One popular time series change detection tool that raised attention in image time series analysis is BFAST (Breaks For Additive Season and Trend) (Verbesselt et al., 2010, 2012). BFAST constitutes a change detection procedure on top of a comprehensive set of serial structural change detection tools. BFAST has been applied in various cases, such as detection of shifts in vegetation trends (Jong et al., 2012; Forkel et al., 2013). BFAST detects the structural change in trend and seasonality of a time series, which has many applications. For example, the seasonality between agriculture products (e.g. soybean) and rainforest are different, which enables the discrimination of different kinds of forest disturbances (e.g. changes from forest to agriculture vs. forest fire). BFAST treats observations as serially uncorrelated. Since it models pixel time series independently, possible spatial correlation around the area is ignored. Simple extensions to BFAST could model the residuals as an autoregressive (AR) process, and/or adopt a simultaneous autoregressive (SAR) model for the spatial residual process.

In this paper, we apply BFAST to our study region, and evaluate the effect of extending BFAST with temporal and spatial correlations. We want to do this in such a way that (1) it can be extended to global-scale data and (2) it is reproducible by other scientists within a reasonable effort. This means that we need to use a high-level data analysis language, such as R; that we need to use an open source Array Data Management and Analytics Software (DMAS) that allows parallel execution of the R code; and finally that we publish all the scripts to recreate the database and carry out the computational experiments on the data.

The study case concerns historical forest cover change detection with long-term MODIS image time series. We show how pixel-based time series analysis are extended to region-based joint spatio-temporal analysis, how the whole change modeling process and spatio-temporal information exploitation are simplified by multidimensional arrays, and how Array DMAS implement and scale the process. The study case is extensible and the methodologies are generic and can form the basis for further remote sensing data experiments.

The paper is organized as follows. Section 2 introduces and discusses multidimensional arrays. Section 3 describes how we model spatio-temporal change. Section 4 introduces the study case. Section 5 presents results, and Sections 6 and 7 finish with discussion and conclusions, respectively.

2. Multidimensional arrays

Most natural phenomena can be represented in multidimensional arrays once they are sampled and quantized in a computer system. The dimensionality of an array can be flexibly set for efficient information extraction and modeling. Examples of practical array abstraction include: 1-D ordered tables or time series (t); 2-D satellite images (x/y); 3-D satellite image time series (x/y/t); 4-D multi-spectral spatio-temporal data (band/x/y/t); subsurface hydrological data (x/y/z/t); and 5-D multi-sensor, multi-spectral spatio-temporal data (sensor/band/x/y/t).

2.1. Potential application of multidimensional arrays in remote sensing

As a multidimensional data structure, arrays have the potential to bring many advanced information extraction into practical use. For example, instead of using a single spectral layer (e.g. vegetation index), multi-spectral multi-temporal approaches (e.g. spectral-temporal surface (Mello et al., 2013)) use more information and thus are able to better represent the earth surface (Mello et al., 2013). This multi-spectral multi-temporal approach can be integrated with spatial information. Data can be organized as 4-D arrays with space, time and bands as four dimensions, and algorithms can be applied to them. Similar examples can be found in data fusion (Castanedo, 2013), where data from different sensors can be organized on two dimensions, and in spatio-temporal statistical modeling. In addition, the developed spatio-temporal statistical algorithms can be flexibly applied within array partitions that span the relevant array dimensions. This study especially demonstrates how array data can be used in spatio-temporal change modeling, and how an Array Data Management and Analytics Software System (DMAS) can be used for parallelization and scaling.

2.2. Tidy data with array data structure

The open source data analysis programming language R (R Core Team, 2015) provides rich data analysis tools. All entities R works on are objects. A special type of object is the array. For instance, the following code segment creates a $100 \times 100 \times 10 \times 5$ array, requests its dimensions, prints the length of the data vector (the product of the dimensions), and shows the length of a one-dimensional sub-array (vector) in the third dimension:

```
> a = array(NA, c(100,100,10,5))
> dim(a)
[1] 100 100 10 5
> length(a)
[1] 500000
> length(a[10,10,,1])
[1] 10
```

Such arrays are held in main memory, are dense, and hence do not scale to massive data or for sparse arrays. They allow to efficiently carry out functions over single dimensions (or sets of dimensions), such as is done in remote sensing time series analysis. Also, arrays keep no information on how dimensions or indexes relate to time, space, or other data properties, so they require additional book-keeping.