Data Reduction for Great Variability Images in a Temporal Analysis

Mára Regina Labuto Fragoso da Silva
Espírito Santo Federal University
marar@npd.ufes.br

Abstract

This paper introduces a new methodology for data reduction concerned on a time sequence group of images having an important problem: the images composing the group are very different ones – either because of noise or because of an event with dynamic behavior (even both causes). In this case, PCA does not perform very well and the data must be prepared before applying this method. In order to improve PCA the proposed method tries to correlate the data before PCA is applied, using two techniques: noise removal by pixel-to-pixel pure pixel mean and grouping by a user-defined similarity index. A case study using sea surface temperature images acquired by satellite remote sensing is described with good results.

1. Introduction

Data reduction for great variability images in a temporal analysis is a very complex problem. This paper works with a very good example. With the arrival of Satellite Oceanography, there are new tools for collecting information by means of images acquired from special sensors onboard satellites. However, an important problem resulted: the need for dealing with a huge number of available observations. So, several methods for data reduction are used to extract information from a series of time subsequent ocean parameters images, called temporal analysis. The proposed methods in the literature to reduce image dataset concerning on sea surface temperature (SST) images temporal analysis are the following: Principal Components Analysis (PCA): [1]; Mean: [2]; Comparison with in situ data: [3]; Visual analysis and Color enhancement: [4]; Artificial neural network (NLPCA): [5]. But the two most used methods are PCA and NLPCA. The PCA finds the eigenmodes of the data covariance matrix and, with this result, one is able to reduce dimensionality and analyze the main patterns of variability present in the first modes of the dataset. However, PCA solves the eigenmode problem using a linear approach and sea surface temperature variability has a nonlinear nature. Nevertheless, using non-linear approach as NLPCA, has an important limitation: saturation phenomena presented in Neural Networks prevents the use of this approach to handle large datasets and large matrices (as images). For REVIZEE Programme[1] this author had to work with 212 SST images (24 hours temporal resolution) and each image dimension was 1112 X 1556 pixels (1,2 km^2 spatial resolution). So, a new method had to be implemented.

2. Methodology for Data Reduction

The SST time subsequent images can change dramatically from one day to another. This fact happens because: (a) the sea surface temperature images have large cloudy regions, detected by small or negative temperatures, which give no information; (b) this parameter behavior has a great variability due to sea currents and meteorological events. These two problems turn the time sequence images in a series of de-correlated ones. In order to correlate the data set preparing the data for better results using the PCA method, the methodology proposed realizes two processing techniques, each one to solve one described problem: 1. Cloud removal technique (realized with all time subsequent images): based on a pixel-to-pixel mean including only the pure pixels (not cloud) at each operation over a group composed by a variable number of time subsequent images, until we have almost 100% of information in all the determined area of study. This method gives a better transition (smooth) between subsequent images and tries to keep the best information acquired without noise (cloud, in this context); 2. Grouping technique (realized with all time subsequent mean images generated by cloud removal technique described in the previous item): based on similarities or distances (natural grouping) according to a defined similarity index, not according to a fixed

[1] www.secirm.mar.mil.br/psrm/revizee/revizee.htm. Access: April, 24, 2005.

human scale as weekly, monthly, yearly, 10 days, etc. as found in literature. Sometimes an event occurring on March, 31st belongs to the group of April, 1st, and it does not belong to the last group of March, because Nature does not follow human decisions or rules.

3. The Similarity Index

As the study of the SST parameter in this work had the purpose of determining the ocean live resources[1], the selection of the index took into consideration the significant threshold for SST change for migration or extinction of species. After consulting the specialists, the selected threshold was ±1°C. The change detection algorithm analyzed the time subsequent images (first processed with cloud removal technique) and compared an image and its antecessor, generating a similarity index (number of pixels with difference less than ±1°C between the two images, except cloud information). The algorithm proceeded to all images and calculated the similarity mean index. Afterwards, all the time subsequent images were grouped if their index was less than or equal to the mean index and generate a new group if it was bigger. So, the interface between groups indicated some event and the date of this event could be even defined. For other applications a new similarity index must be generated.

4. Experimental Results: Cabo Frio Case

Images from Cabo Frio City, Rio de Janeiro, Brazil, August, 1999 were chosen for testing because there was a known upwelling event. Images comprised latitude: 20°29'55,32'' to 23°59'33,40'' S; longitude: 39°30'10,91'' to 42°30'40,68'' W, total area 58240 pixels → 70470,4 Km^2. There were only 12 images available for this month. Data reduction using the methodology described (cloud removal + similarity index) resulted in 3 groups. Fig.1 shows PCA obtained with the 3 groups generated by the proposed methodology. A visual analysis of the daily images showed a great change at August, 31st, 1999, with temperatures ranging from 16 to 21°C. This event is showed in the first mode of PCA using the proposed method, but it was not detected by the PCA generated without the method. It could be observed a small area of cold water from 17-18°C in the coast between 26-28, August, enhanced by the first mode of PCA using the proposed method, but it was less detected (subtle) by the PCA generated without the method.

Concerning about data reduction, the first mode of PCA using the proposed method had 71% of information (98.4% in two modes), but the first mode of PCA generated by weekly mean had only 43% of information (68.7% in two modes).

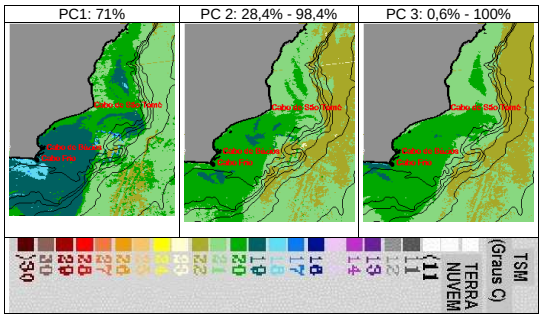

Figure 1. Cloud removal+similarity index PCA.

5. Conclusions and Future Work

There are several tools for sea surface temperature temporal analysis. The most used technique – PCA – is not well suited for nonlinear data behavior as SST. So, the data must be prepared before using PCA in order to reduce the nonlinearity and correlate the data. The proposed method tries to adequate the data so as to take advantage of PCA potentiality of data reduction. This method was tested with several types of SST images [6]: small to great images (260 X 224 pixels to 1112 X 1556 pixels) and small to medium temporal resolution (GOES: acquisition each 0.5 hour and NOAA: acquisition each 24 hours) and worked well in all cases. As was expected, the smaller the area, the better the similarity index. Improvements/adjustments on the threshold choice, the rules for grouping (mean index or a percentage of the index) and the dimension of the image concerning about a specific event are some of the suggestions for a future work.

6. References

[1] Parada, M., Cantón, M., 1998, "Sea surface temperature variability in Alborán sea from satellite data", International Journal of Remote Sensing, v.19, n.13, pp. 2439-2450.

[2] Huang, W, Fang, M., 1998, "Monthly composites of sea surface temperature of the South China Sea", International Journal of Remote Sensing, v.19, n.13, pp. 2435-2438.

[3] Malm, J. and Lennart, J., 1994, "Water Surface Temperature Characteristics and Thermal Bar Evolution during Spring in Lake Ladoga", Remote Sensing of Environment, v.48, pp. 332-338.

[4] Narayana, A., et al.,1995, "Geometric correction and radiometric normalisation of NOAA AVHRR data for fisheries applications", International Journal of Remote Sensing, v.16, n.4, pp.765-771.

[5] S. Botelho, et al.,2005 "C-NLPCA: Extracting Nonlinear Principal Components of Image Datasets". XII Simpósio Brasileiro de Sensoriamento Remoto, pp.3495-3502.

[6] M. R. L. F. Silva, "Aplicações de TSM Orbital em Engenharia Oceânica", DSc. Thesis.Universidade Federal do Rio de Janeiro, Brazil, 2002, 139 p.