PRINCIPAL COMPONENT ANALYSIS OF C-SAR IMAGES FOR FLOOD MAPPING - SANTA FE PROVINCE, ARGENTINA*

ANÁLISE DE PRINCIPAIS COMPONENTES DE IMAGENS C-SAR PARA MAPEAMENTO DE INUNDAÇÃO - PROVÍNCIA DE SANTA FE, ARGENTINA

Jones Zamboni Graosque
Universidade Federal do Rio Grande do Sul
Programa de Pós-Graduação em Geografia (Brasil)
ORCID 0000-0002-8451-904X    jones.eu@gmail.com

Laurindo Antonio Guasselli
Universidade Federal do Rio Grande do Sul
Programa de Pós-Graduação em Geografia (Brasil)
ORCID 0000-0001-8300-846X    laurindo.guasselli@ufrgs.br

ABSTRACT

Flood events are phenomena associated with heavy rainfall. In Argentina, floods have high economic and social costs, including loss of human life. In this paper, principal component analysis (PCA) is used to map flood-prone areas along the Paraná river in Santa Fe, Argentina. The Sentinel-1B (S1B) images, sensor C-SAR with VH polarisation Interferometric type (IW) Ground Range Detected (GRD) with spatial resolution of 10 m, from 2016, were referenced and the PCA method was used to extract the four first principal components. The flood-affected images make it possible to accurately define the flooded area. In targets with dense vegetation, however, there is no pixel backscatter pattern. PC2 better highlighted the threshold of pixel intensity, with an accuracy of 70%, and 93% of the mapped area was shown to be flood-prone. Procedures to map floods remotely are pivotal because they can quickly obtain precise data on flood areas that may not be accessible for fieldwork or that have not yet been mapped in great detail.

Keywords: Sentinel-1, time series analysis, principal components, Argentina.

RESUMO

Inundações são associadas a chuvas intensas. Na Argentina é o evento natural que causa mais perdas econômicas, sociais e de vidas humanas. O objetivo desse trabalho é mapear a área de inundação do rio Paraná, em Santa Fe, por Análise de Componentes Principais (ACP). As imagens Sentinel-1B, sensor C-SAR, polarização VH do tipo Interferométrico (IW) Ground Range Detected (GRD), pixel de 10 m, ano 2016, foram referenciadas, extraíndo as quatro primeiras ACP. As imagens sob efeito de inundação permitiram delimitar com precisão a área inundada. No entanto, em áreas com densidade de vegetação não há um padrão de retroespalhamento dos pixéis. A PC2 destacou melhor o limiar de intensidade dos pixéis de inundação, com uma precisão de 70%, sendo que 93% da área mapeada é suscetível à inundação. A cartografia de risco de inundação obtida a partir de sensoriamento remoto revela-se essencial, pois possibilita a obtenção de resultados rápidos e precisos das áreas de inundação, em áreas cujo trabalho de campo não seja possível ou não se encontrem disponíveis mapas detalhados das áreas atingidas.

Palavras-chave: Sentinel-1, análise temporal, principais componentes, Argentina.


Este artigo é parte integrante da Revista Territorium, n.º 27 (II), 2020, © Riscos, ISSN: 0872-8941.
Introduction

Among natural disasters, floods may be the most common ones causing more destruction than any other natural hazard worldwide (Sanyal & Lu, 2004; Tingsanchali, 2012; Rahman & Thakur, 2018).

In Argentina, floods are the main natural cause for economic and social losses as well as loss of human life. In the Litoral region alone – comprising the provinces Formosa, Chaco, Missiones, Corrientes, Santa Fe, and Entre Ríos - 165 people perished and almost 6,000 were injured due to fluvial flooding events between 1970 and 2004. Nearly one million people had to be evacuated, and approximately 18,000 houses were destroyed. An estimated 1.7 million head of cattle were lost, and 28 million hectares of farm land were affected (Celis, 2006). This region has already sustained substantial losses in agriculture and cattle husbandry inflicting significant damage on these sectors (Noticias Agrícolas, 2018; Correio do Brasil, 2018; Agrolink, 2018).

Flood maps of the city of Santa Fe were developed based on fieldwork and optical imaging of flooding events in 2003 and 2007 (Vallejos et al., 2014). This data was paramount to devise emergency plans and protection structures for the city of Santa Fe (Fe, 2018).

However, flood maps based on optical imaging may not render the correct dimensions of the flooded area due to cloudy skies or vegetation. Some studies suggest that Synthetic Aperture Radar (SAR) sensors provide better results to detect floods than optical sensors (Hess & Melack, 1994; Townsends & Walsh, 1998; Tralli et al., 2005; Clement et al., 2018; Markert et al., 2018).

The Global Flood Monitoring System (GFMS) provides updated maps on floods and precipitation online (Wu et al., 2014). In addition, tools such as Earth Data and GFMS provide valuable and low-cost hydrological data which may contribute to flood maps and policy development (Yan et al., 2015). GFMS is an efficient system that provides maps shortly after the event. The spatial resolution, however, amounts to 1 km compromising the level of detail of the flood area (Wu et al., 2014). In general, the flood maps of the region of the city of Santa Fe mainly indicate how far the water proceeds in urban areas; they address potentially flooded areas in smaller communities or rural areas only secondly.

Using radar images to map floods can offset the restrictions of optical imaging. One of the most commonly applied techniques is thresholding. It is a conventional method to map non-forested areas that maximises the contrast between water and land facilitating the representation of floods in histograms (Brivio et al., 2002; Malnes et al., 2002; Solbø & Solheim, 2004; Rahman & Thakur, 2018).

When applied to a single image, the thresholding method may not size a flooding event precisely; however, when applied to a time series of images, it becomes increasingly accurate since even flood water covered by vegetation is detected. Considering that a greater amount of images is required to conduct the analysis and that the procedure is more complex, the maps can be more precise, especially in vertical targets such as wooded areas (Townsend, 2001 and 2002; Kandus et al., 2001; Solbø & Solheim, 2004).

Principal Component Analysis (PCA) permits processing images with a greater number of spectrum bands and obtaining information from multi-temporal data (Crosta, 1992; Sato et al., 2011). PCA provides a new set of variables, simplifies the description of the data set, and analyses the structure of the observations and variables (Abdi & Williams, 2010). It proves to be a promising method to detect the dynamics and changes in land cover such as in case of floods (Dutsenwai et al., 2016).

When mapping floods on the basis of remote sensing tools, the more comprehensive the time series of images in the PCA is, the better the results are as water can be mapped even in vertical target areas, such as densely vegetated areas (Solbø & Solheim, 2004; Santoro et al., 2015; Graosque, 2018). Thus, this paper aims at mapping flood areas employing Principal Component Analysis in the area surrounding the city of Santa Fe, Argentina.

Area of study

The area of study encompasses a section of the Paraná river which divides the two Argentinian provinces Santa Fe and Entre Ríos and borders on the cities of Santa Fe and Paraná. In the area, the floodplain comprises urban and rural areas resulting in urban parts of Santa Fe being flood-prone.

The analysed region (fig. 1) was classified as urban areas, rural areas, humid areas or wooded areas, in addition to permanent water bodies such as the Paraná river (Benzaquén & Argentina, 2013).

Methodology

Radar images of the satellite Sentinel-1B (S1B) from 2016 were analysed to map flood-prone areas in the region. The Synthetic Aperture Radar (SAR) images are C-band, Vertical Horizontal (VH) polarised and were acquired in InterTerometric Wide Swath (IW) mode. The softwares SNAP and ArcMap 10.2 were used to process the images.

The table outlines the 2016 images and the respective level of the Paraná river in the city of Santa Fe. The alert level for major flooding of this river in Santa Fe is 5.3 metres (Table 1). The river’s average level is 3.6 metres (Centro de Informaciones Meteorológicas, 2018).
Prior to the Principal Component Analysis, the images were calibrated, filtered (speckle), and complemented with orbital information and terrain correction. Furthermore, all images were georeferenced based on the 22nd October 2016 image using 2,000 reference points automatically extracted by the software SNAP. The same software generated the principal component images, which include, among others, one image without and another with the effects of floods, both after pre-processing (fig. 2).

The validation of the flooded area was restricted to the flood maps developed for the city of Santa Fe in 2003 (International Disasters Charter, 2019) since fieldwork was not feasible. For comparison purposes, the thresholding method was applied to the image with the highest water level, namely the one from 7th February 2016 (Graosque, 2018).

**Results**

Targets with variations in pixel intensity between images with and without floods were used to delineate the flooded areas. The area covered by water varies according to the image sequence facilitating the identification of flooded areas while averting the risk of misinterpreting permanent water bodies.

The principal component image that represents the flooded areas best is PC 2. The images from February, March and April (images with floods) contribute to the principal component (Table II).
In PC 2 the dark-colored pixels indicate major variations; those in light grey or white suggest no major pixel changes. The Setubal lagoon is a prime example of a permanent water body represented in light grey (fig. 3-B). Around the water bodies, however, dark areas appear indicating changes in pixel intensity between the twelve images. These dark areas are mainly located in proximity to the Salado river, to humid areas, and along the shore of the Setubal lagoon.

Image PC 1 concentrates 97% of the eigenvalues and represents the pixels all images share. All images maintain a uniform contribution since the features of the targets are continuous throughout the dates (Henebry, 2014). PC 1 is very similar to the image of the area of study before applying the method making it difficult to differentiate between water pixels and non-water pixels (fig. 3-A). Consequently, PC 1 is not ideal for flood evaluation.

PC 2, 3 and 4 emphasise information that is less common compared to PC 1. Among these, PC 3 and 4 show variations in image contributions and highlight specific targets such as rural areas (fig. 3-C and D). These two PCs combined represent 0.8% of the eigenvalues. Moreover, the main contributing images for these two PCs are images with and without floods which drastically reduces the chances of identifying a specific threshold for flood water.

PC 2 and 3 concentrate the greatest water body variation (Gómez-Palacios et al., 2017). PC 2, however, contains three times more eigenvalues than PC 3 qualifying it as the main PC for flood analysis. Furthermore, PC 2 illustrates how the images that contribute most eigenvalues are exactly those dated to events with water levels above the alert level (5.3 metres) indicating a variation in flood-related pixels in this principal component.

PC 2 contributed 1.4% of the eigenvalues, and targets with the highest variation show dark pixels while targets with little variation are represented by lighter pixels (fig. 3-B). The Setubal lagoon is a permanent water body with minimum pixel variation. Consequently, this water body can be consulted to define the threshold for what is water and what is not. Pixel intensity may vary in urban and rural targets; however, urban targets need to be disregarded since this method and the applied tools are inappropriate to map floods in such targets (Solbe & Solheim, 2004; Henry et al., 2006). Regarding vegetation, the crop and the season of every area must be taken into consideration (Santoro & Wegmüller, 2014).
The flood areas mapped in PC 2 are mainly located in humid target areas where vertical vegetation is lower (Wu et al., 2014). Wooded and rural areas, however, were also partially identified as flood areas. Although vegetation cover prevents the identification of water depth in these targets, the pixel intensity varies between the images with and without floods. PC 2 shows the pixel variation in flood targets best (fig. 4).

The flood map from April 2003 (International Disasters Charter, 2019) was used to validate the results surrounding the city of Santa Fe including areas of the Salado river (fig. 5). In comparison with this map, the use of PC 2 achieved an accuracy of 54.37%. Based on the results, 93.03% of the mapped area is considered flood-prone. In less urbanised areas comprising more wooded and humid areas, the accuracy was calculated based on the flood area developed by Graosque (2018) because Argentina floods (2019) only mapped the city of Santa Fe. In wooded and humid areas, the PC 2 method achieved an accuracy of 70.26%.

The flood area could be mapped with only one image (fig. 4-B); however, the comparison with the technique of Principal Component Analysis shows that the thresholding method with a single image did not fully map flood-prone areas (fig. 4). Technically, PCA also uses a threshold to identify water and non-water which is, however, defined by multiple images. The observed target threshold using a single image was -22.5 [db] while it was 0.0 [db] using PCA in PC 2 (Goméz-Palacios et al., 2017; Graosque, 2018). The change in threshold values is due to the fact that in PC 2 the image basically distinguishes between pixels that varied and those that did not undergo any variation.

In addition to diverging thresholds, the differences between the methods using a time series of images and those using a single image are reflected in vertical targets such as vegetation. Mapping efforts applying the thresholding method with a single image are only recommended for flat areas (Malnes et al., 2002; Solbø & Solheim, 2004; Tuan & Duong, 2009). Analysing a time series of multiple images and hence identifying variations in pixel intensity makes it possible to map water in targets with trees, for example.

Since these tools are inappropriate to analyse floods in urban areas (Solbe & Solheim, 2004), urban Santa Fe and Santo Tomé were excluded from mapping (fig. 5).
Conclusions and recommendations

Principal Component Analysis proves to be a viable option to map flood areas remotely because in addition to the precision of the result, the threshold between water and non-water is identified more precisely for horizontal as well as vertical targets. Only PCA (mainly in PC 2) permits to evaluate in which images the water level was elevated; however ideally, this information should be available before applying the method because the images that are expected to provide a major contribution to PC 2 would already be known. Measuring water levels in the field could provide such information.

Another distinctive feature is, however, that the method is not applicable to map floods in urban areas. For this reason, the urban areas of Santa Fe and Santo Tomé were not included in the mapping process.

Moreover, the use of PCA to map floods in rural areas may show divergence in the final result due to the types of crops and the season during which the images were taken. In this case, information on the area is required for the flood analysis.

References


Gómez-Palacios, D., Torres, M. A., & Reinoso, E. (2017). Flood mapping through principal component analysis of multitemporal satellite imagery considering the alteration of water spectral properties due to turbidity conditions. Geomatics, Natural Hazards and Risk, 8(2), 607-623. DOI: https://doi.org/10.1080/19475705.2016.1250115


Rahman, M. R., & Thakur, P. K. (2018). Detecting, mapping and analysing of flood water propagation using synthetic aperture radar (Sar) satellite data and GIS: A case study from the Kendrapara District


