Independent component analysis of MODIS-NDVI data in a large South American wetland

ANDRÉS ANTICO

Centro Regional de Geomática, Facultad de Ciencia y Tecnología, Universidad Autónoma de Entre Ríos, Argentina

Monthly images of Normalized Difference Vegetation Index (NDVI) from the moderate resolution imaging spectroradiometer (MODIS) are used to characterize the spatio-temporal variability of vegetation in a large South American wetland (SAW) (located in the Paraná River floodplain) during the period 2000–2009. While these data do not meet the requirements of classical component extraction techniques (CETs) (e.g. principal component analysis (PCA)), they are suitable for the modern method named independent component analysis (ICA). Hence, ICA is used here to extract three statistically independent modes of inter-annual MODIS-NDVI variability that are successfully interpreted as vegetation responses to hydrological changes. One mode isolates the vegetation response to a severe drought associated with La Niña 2007–2008. Another component reflects the expansion (or contraction) of lagoons owing to high (or low) water level of the Paraná River. The remaining mode captures the vegetation decrease caused by the flood related to El Niño 2006–2007. The results presented here for a particular wetland suggest that ICA of NDVI images is a powerful tool for identifying the physical causes of vegetation changes in other large wetlands.

1. Introduction

Wetlands are important elements of the Earth System when they are considered at global, regional and subregional scales (Mitsch and Gosselink 2007). In a global context, wetlands play an important role in various biogeochemical cycles. For instance, they have a direct impact on atmospheric concentrations of carbon dioxide, methane and sulphur. At regional and subregional scales, wetlands constitute banks of biodiversity, productivity and recruitment. Moreover, these systems act as nutrient traps, moderate the effect of floods, improve water quality and provide places for recreational activities. Thus, it is relevant to improve our understanding of wetland responses to environmental changes.

The Normalized Difference Vegetation Index (NDVI) is a remotely sensed measure of the vigour and amount of vegetation (Tucker 1979, Tucker *et al.* 1991). This index has been widely used to study terrestrial ecosystems and their response to climatic variables (e.g. precipitation) at global and regional scales (e.g. Li and Kafatos 2000, Anyamba *et al.* 2001, Los *et al.* 2001, Gurgel and Ferreira 2003, Lotsch *et al.* 2003a,

---

*Email: aantico@santafe-conicet.gov.ar

Poveda and Salazar 2004, Nagai *et al.* 2007). However, in wetlands, it is not common to use this index for studying vegetation changes. This is probably because the seasonal cycle of NDVI is significantly disrupted by large negative NDVI values when there is a water excess (i.e. water is above the ground or in saturated soils) (Zoffoli *et al.* 2008). Nevertheless, Zoffoli *et al.* (2008) recently showed the potential of NDVI data for gaining insights into the functioning of wetlands. In selected locations of a South American wetland (SAW), Zoffoli *et al.* (2008) analysed NDVI time series derived from coarse resolution (8 km × 8 km) images. They used a correlation coefficient as the primary tool for analysing these time series. Although interesting results were obtained with this simple approach, more detailed studies are needed for increasing our knowledge of large wetlands (Zoffoli *et al.* 2008).

The objective of this letter is to conduct a detailed analysis of the spatio-temporal variability of NDVI with the aim of identifying physical causes of vegetation changes in a large SAW. To achieve this goal, we perform an independent component analysis (ICA) on a sequence of moderate resolution ($0.1^\circ \times 0.1^\circ$) NDVI images of the same SAW studied by Zoffoli *et al.* (2008), and we interpret the extracted components as vegetation responses to hydrological changes. ICA is a modern component extraction technique (CET) that was mainly developed in the fields of signal processing and neural modelling to overcome various limitations of classical CETs (Hyvärinen and 2000). Nevertheless, ICA is also a useful tool for analysing the spatio-temporal distribution of geophysical variables (e.g. Aires *et al.* 2000, Lotsch *et al.* 2003b) and for conducting various types of remote sensing studies (e.g. Lizarazo 2010, Chen *et al.* 2011). At the continental scale, Lotsch *et al.* (2003b) showed that ICA of NDVI images serve to find the physical causes of vegetation variability. However, to the best of our knowledge, ICA has not been used previously in large wetlands for decomposing a sequence of NDVI images into physical meaningful modes of variability. Hence, this study presents ICA of NDVI data as a promising method for studying the ecosystems of large wetlands.

2. Study area

The study area consists of a wetland of subregional extension (300 km long and 50–80 km wide) located in south-eastern South America and associated with the floodplain of the Paraná River (figure 1). Different landscapes and ecosystems coexist in this large wetland (Zoffoli *et al.* 2008). This complex and diverse system evolved from a marine embayment to its present state during the last 6×10^3 years as a consequence of a gradual decrease in sea level (Iriondo 2004). For this reason, this area is usually known as the littoral complex (LC).

According to Kottek *et al.* (2006), the LC has a warm and fully humid temperate climate. The climatological average July (January) temperature is 10°C (27°C), and the climatological monthly rainfall varies between 30 and 130 mm with a primary (secondary) peak in autumn (spring) (National Oceanic and Atmospheric Administration (NOAA) 1991).

The lower Paraná River has climatological seasonal cycles of discharge and water level characterized by a maximum (minimum) in spring (autumn) (Pasquini and Depetris 2007).

3. Data and methods

To characterize vegetation changes between February 2000 and August 2009 in the LC, monthly NDVI images from the moderate resolution imaging spectroradiometer
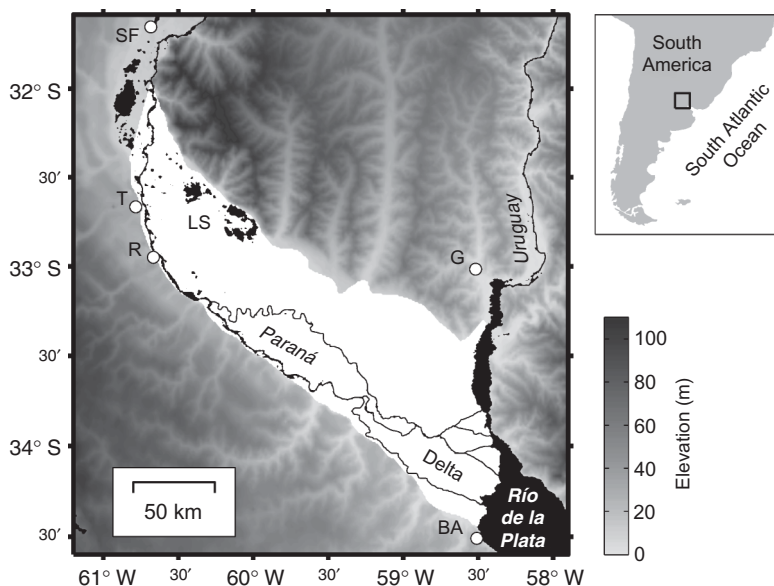
Figure 1. Map showing the littoral complex (LC) (white) and surrounding areas. Main water bodies are shown in black and names of major rivers are given in italics. The relief for land areas is from the digital elevation model ETOPO2v2 (available at http://www.ngdc.noaa.gov/mgg/global/etopo2.html). Elevation values in the LC (not shown) are less than 15 m. Open circles show the locations of the following cities, which are mentioned within this letter: BA, Buenos Aires; G, Gualeguachú; T, Timbúes; R, Rosario; and SF, Santa Fe. Two relevant parts of the study area are also indicated: the lagoon system (LS) and the delta of Paraná river

(MODIS) with a spatial resolution of $0.1^{\circ} \times 0.1^{\circ}$ are used. Because the focus is on inter-annual vegetation variability, the seasonal cycle is removed from the NDVI time series in the following manner. At each grid cell, climatological NDVI values are calculated for each calendar month by averaging the data for that month over the whole study period. Then, monthly anomalies are obtained as deviations from this climatological annual cycle.

Values of skewness and kurtosis of NDVI anomalies significantly depart from 0 in most of the grid cells and, therefore, the probability density function (PDF) of NDVI time series is not Gaussian in most of the LC (figure 2); note that the PDF of these 115-element times series significantly deviates from a Gaussian PDF if the estimates of skewness (or kurtosis) have an absolute value much greater than $\sqrt{15/115} \sim 0.4$ (or $\sqrt{96/115} \sim 1$) (Press *et al*. 1992). Infrequent NDVI values have greater probability to be low than high as shown by the negative values of skewness (figure 2(*a*)). Positive kurtosis values show that the peak of NDVI PDF is more sharp than that of a Gaussian PDF (figure 2(*b*)). Consequently, classical CETs, for example, principal component analysis (PCA), should not be applied to NDVI anomalies since these methods usually require Gaussian input data (Aires *et al.* 2000). This and other limitations of classical CETs are overcome by modern CETs. In particular, ICA is a recently developed CET where statistically independent components (ICs) are extracted from non-Gaussian data (classical CETs such as PCA extract decorrelated components that may not be statistically independent); for a more complete review of ICA, see
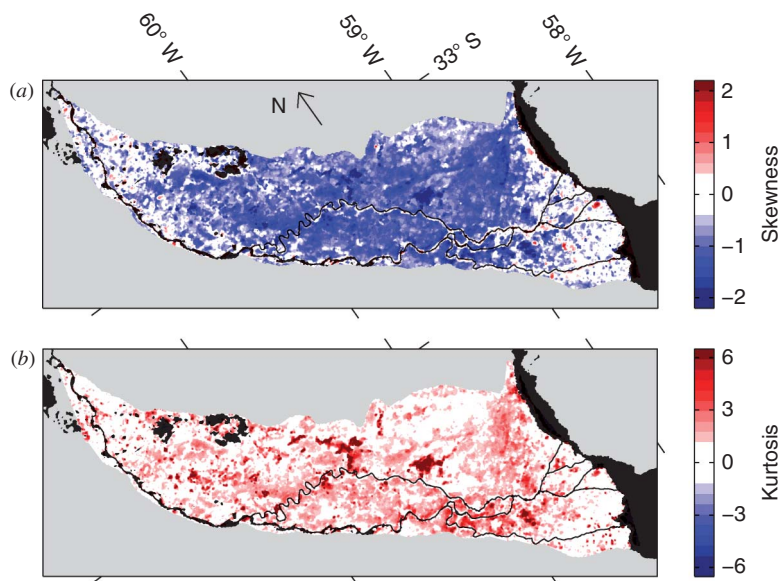
Figure 2. Spatial distributions of (*a*) skewness and (*b*) kurtosis of NDVI anomalies. Skewness and kurtosis estimates whose absolute value is less than 0.4 and 1, respectively, are not shown. That is, estimates of skewness and kurtosis are only shown if they reflect a non-Gaussian behaviour (see section 3).

Hyvärinen and Oja (2000). Therefore, to extract components from NDVI images of the LC, ICA is more appropriate than classical CETs.

It is important to note that several ICA methods exist (Hyvärinen and Oja 2000). In this study, an Infomax ICA algorithm is used because of its simplicity, robustness and fast convergence (Nadal and Parga 1994; Bell and Sejnowski 1995). This algorithm uses a maximum likelihood formulation and the Broyden–Fletcher–Goldfarb–Shanno optimization method (Nielsen 2001). The method assumes linear and instantaneous mixing, a square-mixing matrix and zero noise. Time series of NDVI for every pixel can be expressed as the sum of ICs. The amplitude of each IC varies from location to location according to a mixing matrix that is depicted as a map. In this manner, the data set can be decomposed into a number of ICs equal to the number of grid cells of the NDVI image. For the images used here, this number would be very large ($>10^4$). However, it is common that only a few ICs are meaningful modes of variability (Nadal *et al.* 2000). Hence, to extract only these ICs and to avoid numerical problems, the number of extracted ICs can be decreased by reducing the dimensionality and noise of the input data prior to ICA (Nadal *et al.* 2000). This is usually done through a PCA (or singular value decomposition) truncation so that the largest principal components are only retained. Here, the first five principal components explaining 71% of total variance of NDVI anomalies are retained (i.e. only five ICs are extracted). Hereafter, the kth component from ICA is denoted by ICk. Since the order of the extracted ICs is not meaningful (Hyvärinen and Oja 2000), we order ICs so that the percentage of total variance accounted for a component decreases as k increases (i.e. we order ICs in analogy with principal components).

In floodplain wetlands, water availability mainly depends on the fluvial regime and local rainfall. To interpret ICs of NDVI in terms of changes in these two water inputs, the following two variables are considered: (i) monthly mean water level of Paraná River recorded at Timbúes gauge station (see station location in figure 1) and (ii) monthly total precipitation over the LC. For every month, the latter is estimated as the arithmetic mean of precipitation values from meteorological stations located in the following cities: Buenos Aires, Gualeguaychú, Rosario and Santa Fe (see city locations in figure 1). To make consistent comparisons between ICs extracted from monthly NDVI anomalies and time series of water inputs, the seasonal cycle is also removed from the time series of water level and precipitation in the same manner as for NDVI time series. Accumulated precipitation is considered here as a measure of the ecologically significant water deficit associated with a drought. For a time scale of n months, accumulated precipitation is computed by averaging precipitation anomalies of n consecutive months and by assigning that average to the last month.

To find links between ICs of NDVI and El Niño Southern Oscillation (ENSO), an indicator of this climatic phenomenon should be considered. In this study, the Multivariate ENSO Index (MEI) is used to describe the magnitude and timing of ENSO. This index combines atmospheric and oceanic variables from the equatorial Pacific Ocean and, therefore, it contains more information than other indexes based on a single variable (Wolter and Timlin 1998). It is noted, however, that the conclusions of this study would not change if other indicators of ENSO (e.g. Southern Oscillation Index) are used instead of MEI.

4.   Results and discussion

In this section, ICA results are described and discussed. Only the first three ICs are presented since they can be interpreted as vegetation responses to changes in local rainfall and water level of Paraná River.

IC1 captures a vegetation decrease associated with a drought that occurred in most of the LC between April 2008 and December 2008 (figures 3(*a*) and 4(*a*)). Note that this vegetation minimum almost coincides with the most pronounced and prolonged minimum of rainfall accumulated over 9 months (figure 4(*a*) and 4(*b*)). Note also that this precipitation minimum occurred at the end of the most intense La Niña event of the study period (figure 5(*a*)). Thus, our results are consistent with previous studies that show, at the continental scale, a decrease of precipitation and vegetation during La Niña conditions, and an increase during El Niño conditions, in south-eastern South America (e.g. Los *et al.* 2001; Lotsch *et al.* 2003a).

IC2 reflects changes in vegetation cover associated with variations in Paraná River water level. In the eastern sector of the lagoon system, the amount of vegetation decreases when water level increases and vice versa (figures 3(*b*), 4(*c*) and 4(*d*)). Note that this behaviour is only interrupted in 2006–2007 when the largest peak of water level occurred (figure 4(*c*)). However, the vegetation response to this flood pulse is captured by another IC (see below in this section). It is noted that there are channels that connect the Paraná River with the lagoons. Hence, vegetation changes described by IC2 can be interpreted as expansions (or contractions) of eastern lagoons owing to high (or low) levels of the Paraná River.

IC3 consists of a pulse-like vegetation minimum that mainly occurred in the central part of the LC between March 2007 and July 2007 (figures 3(*c*) and 4(*e*)). This minimum occurred 1–2 months after the largest peaks of time series of precipitation and
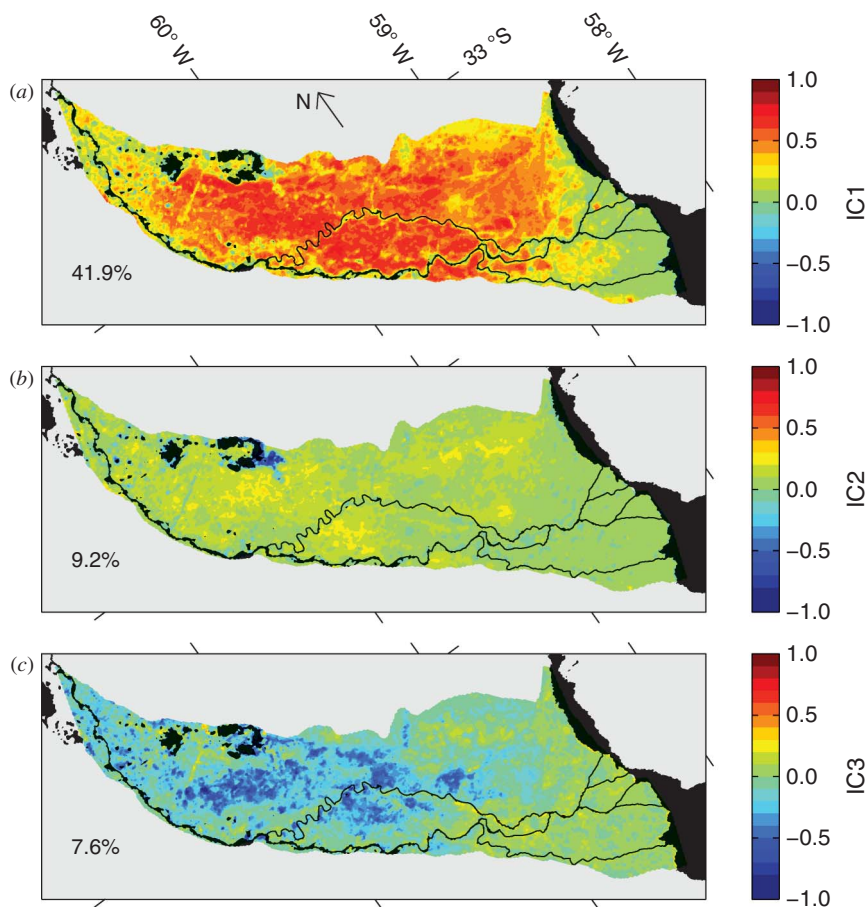
Figure 3. (*a*)–(*c*) Spatial distributions of amplitude of the first three ICs of monthly NDVI anomalies. Each distribution is normalized to a maximum absolute value of unity. The fraction of total variance accounted for each component is shown within the panels.
Note: IC, independent component; NDVI, Normalized Difference Vegetation Index.

water level (figure 4(*f*)). Since these two peaks occurred simultaneously (figure 4(*e*)), it is difficult to assess the role of each of these water sources in generating the flood pulse. Note that the anomalous large values of precipitation and water level occurred at the end of El Niño 2006–2007 (figure 5). Several studies showed that these hydrological conditions are commonly related to El Niño events (e.g. Diaz *et al.* 1998; Los *et al.* 2001; Camilloni and Barros 2003). Therefore, IC3 isolates the effect of a flood pulse, associated with El Niño 2006–2007, on the vegetation cover of the LC.

Interestingly, none of the first three ICs exhibit significant variability in the delta region (figure 3). That is, the seasonal cycle of vegetation remains nearly unchanged from year to year in this part of the LC. This would be the case if there is a dominant water input that does not exhibit significant interannual variability. This type of water source is provided by diurnal and semi-diurnal astronomical tides of the Río De la Plata River which flood the delta region every day (tide tables for Río de la Plata River are available at http://www.hidro.gov.ar/). Thus, our results are in agreement
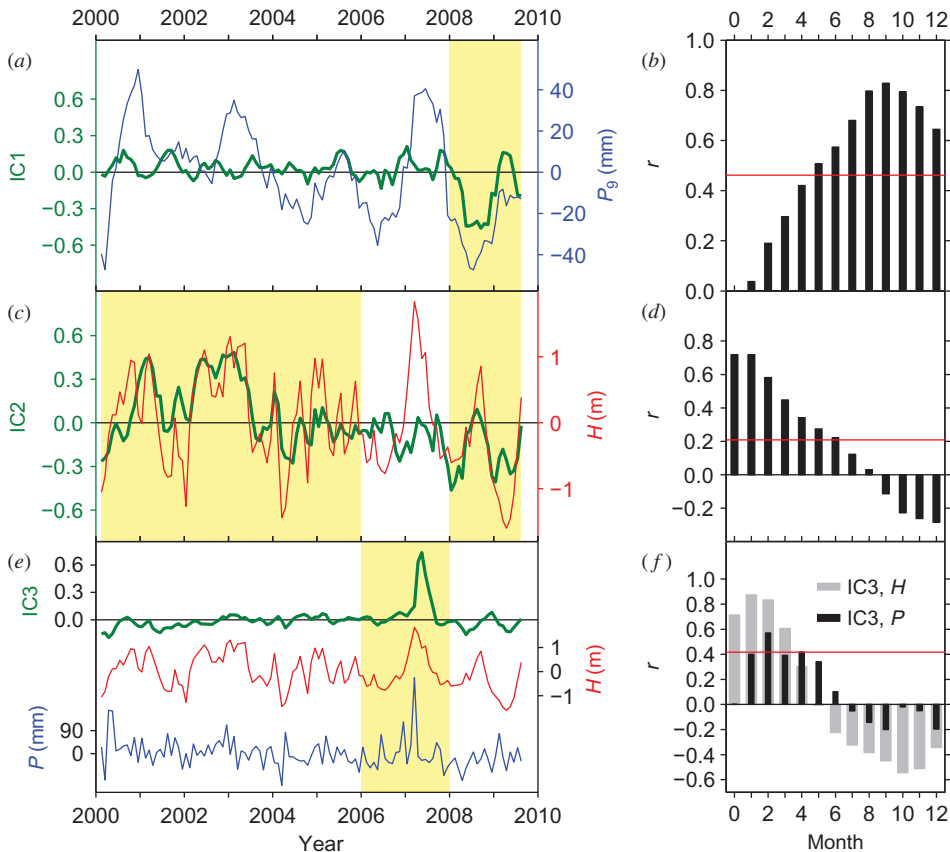
Figure 4. Time series of the first three ICs of monthly NDVI anomalies are shown as thick lines in (*a*), (*c*) and (*e*). In (*a*), the thin line shows accumulated precipitation over 9 months (P_9). In (*c*) and (*e*), thin lines show monthly values of precipitation (*P*) and Paraná River water level (*H*). Panel (*b*) shows values of Pearson's coefficient (*r*) for correlation between time series of IC1 and precipitation accumulated over time intervals varying from 1 to 12 months. Values of *r* for lagged correlation between the following pairs of variables (indicated between brackets) are shown: (*d*) [IC2,H], (*f*) [IC3,H] and (*f*) [IC3,P]; time lags, shown on the abscissa, are only applied to the second variable. The correlation coefficients shown in (*b*), (*d*) and (*f*) are calculated over the shaded periods shown in (*a*), (*c*) and (*e*), respectively. In (*b*), (*d*) and (*f*), *r* values above the horizontal line are significant at the 97.5% level (one-sided *t*-test)
Note: IC, independent component; NDVI, Normalized Difference Vegetation Index.

with Zoffoli *et al.* (2008) who suggested that ENSO-related signals (i.e. interannual variability) are absent in the delta vegetation because of the important effect of tides on the ecosystem of this region.

This study shows that the Infomax ICA algorithm is a useful tool for isolating physical causes of vegetation variability in a large wetland. However, it should be stressed that other ICA algorithms exist (Hyvärinen and Oja 2000). Hence, a possible extension of this work would consist of performing an intercomparison ICA model study to assess the performance of different ICA algorithms for the identification of causes of vegetation changes in wetlands.
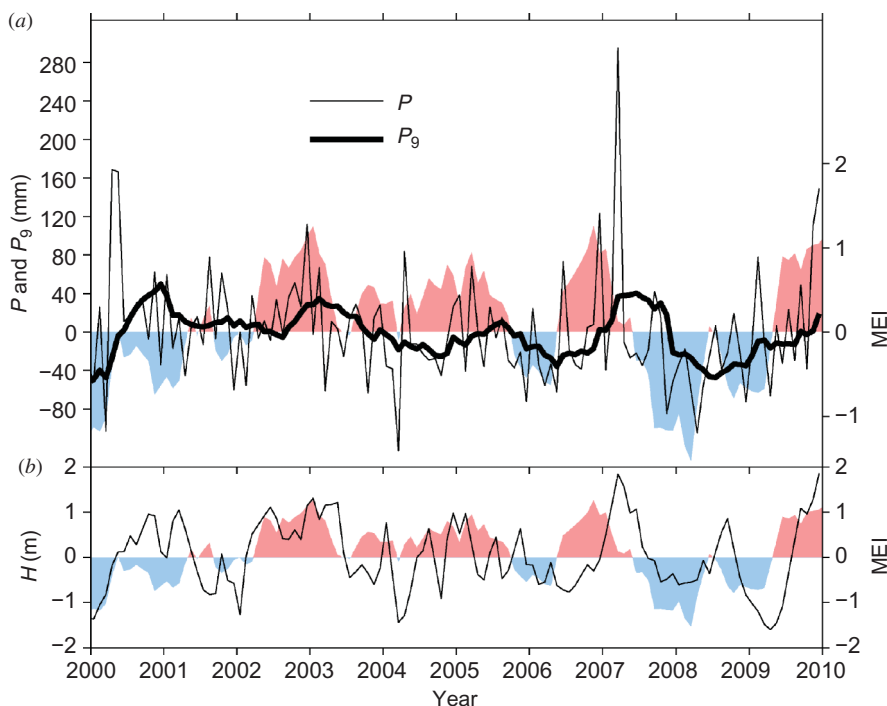
Figure 5.	Time series of monthly anomalies of precipitation (P, thin line in (*a*)), precipitation accumulated over 9 months (P_9, thick line in (*a*)) and Paraná River water level (H, thin line in (*b*)). In (*a*) and (*b*), the time series of MEI is also shown as an area graph. Positive and negative MEI values correspond to El Niño and La Niña conditions, respectively.
Note: MEI, Multivariate ENSO Index.

## 5.	Conclusions

A modern CET named ICA is performed on a 10-year long sequence of monthly MODIS-NDVI images of a large SAW. The main conclusions of this work are the following:

- The NDVI data used in this study do not follow a Gaussian (normal) probability distribution. Consequently, classical CETs, which usually require Gaussian input variables, are not suitable for studying the spatio-temporal variability of these data.
- Since ICA has several advantages over classical CETs (e.g. it does not require Gaussian input data), it succeeds in extracting three physically meaningful components from the NDVI image sequence. Each of these components isolates the effect of a particular hydrological perturbation (e.g. a severe drought associated with La Niña 2007–2008) on the vegetation cover. Thus, ICA of NDVI images provides a detailed and in-depth analysis of the physical causes of vegetation variability in the wetland considered here.
- Although we only applied an ICA method to NDVI data from a particular large wetland of South America, we believe that this modern technique is also a powerful tool for understanding vegetation changes in other wetlands. Hence, we hope that this study provides a useful background for future studies of wetland ecosystems.

Acknowledgements

References

AIRES, F., CHÉDIN, A. and NADAL, J.P., 2000, Independent component analysis of multivariate time series: application to the tropical SST variability. *Journal of Geophysical Research*, **105**, pp. 17437–17455.

ANYAMBA, A., TUCKER, C.J. and EASTMAN, J.R., 2001, NDVI anomaly patterns over Africa during the 1997/98 ENSO warm event. *International Journal of Remote Sensing*, **22**, pp. 1847–1859.

BELL, A. and SEJNOWSKI, T.J., 1995, An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, **7**, pp. 1129–1159.

CAMILLONI, I.A. and BARROS, V.R., 2003, Extreme discharge events in the Paraná River and their climate forcing. *Journal of Hydrology*, **278**, pp. 94–106.

CHEN, F., GUAN, Z., YANG, X. and CUI, W., 2011, A novel remote sensing image fusion method based on independent component analysis. *International Journal of Remote Sensing*, **32**, pp. 2745–2763.

DIAZ, A.F., STUDZINSKI, C.D. and MECHOSO, C.R., 1998, Relationships between precipitation anomalies in Uruguay and Southern Brazil and sea surface temperature in the Pacific and Atlantic Oceans. *Journal of Climate*, **11**, pp. 251–271.

GURGEL, H.C. and FERREIRA, N.J., 2003, Annual and interannual variability of NDVI in Brazil and its connections with climate. *International Journal of Remote Sensing*, **24**, pp. 3595–3609.

HYVÄRINEN, A. and OJA, E., 2000, Independent component analysis: algorithms and applications. *Neural Networks*, **13**, pp. 411–430.

IRIONDO, M., 2004, The littoral complex at the Paraná mouth. *Quaternary International*, **114**, pp. 143–154.

KOTTEK, M., RIESSER, J.G., BECK, C., RUDOLF, B. and RUBEL, F., 2006, World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, **15**, pp. 259–263.

LI, Z. and KAFATOS, M., 2000, Interannual variability of vegetation in the United States and its relation to El Niño/Southern Oscillation. *Remote Sensing of Environment*, **71**, pp. 239–247.

LIZARAZO, I., 2010, Fuzzy image regions for estimation of impervious surface areas. *Remote Sensing Letters*, **1**, pp. 19–27.

LOS, S.O., COLLATZ, G.J., BOUNOUA, L. and SELLERS, P.J., 2001, Global interannual variations in sea surface temperature and land surface vegetation, air temperature, and precipitation. *Journal of Climate*, **14**, pp. 1535–1549.

LOTSCH, A., FRIEDL, M.A., ANDERSON, T. and TUCKER, C.J., 2003a, Coupled vegetation-precipitation variability observed from satellite and climate records. *Geophysical Research Letters*, **30**, pp. 1774–1777.

LOTSCH, A., FRIEDL, M.A. and PINZÓN, J., 2003b, Spatio-temporal deconvolution of NDVI image sequences using independent component analysis. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, pp. 2938–2942.

MITSCH, W.J. and GOSSELINK, J.G., 2007, *Wetlands*, 4th ed. (Hoboken, NJ: John Wiley & Sons).

NADAL, J., KORUTCHEVA, E. and AIRES, F., 2000, Blind source separation in the presence of weak sources. *Neural Networks*, **23**, pp. 589–596.

NADAL, J.P. and PARGA, N., 1994, Nonlinear neurons in the low-noise limit: a factorial code maximizes information transfer. *Network: Computation in Neural Systems*, **5**, pp. 565–581.

NAGAI, S., ICHII, K. and MORIMOTO, H., 2007, Interannual variations in vegetation activities and climate variability caused by ENSO in tropical rainforests. *International Journal of Remote Sensing*, **28**, pp. 1285–1297.

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION (NOAA), 1991, *Climates of the World* (Asheville, NC: National Environmental Satellite, Data, and Information Service, National Climatic Data Center).

NIELSEN, H.B., 2001, *UCMINF – An Algorithm for Unconstrained, Nonlinear Optimization*, Technical Report IMM-TEC-0019, IMM (Copenhagen: Technical University of Denmark).

PASQUINI, A.I. and DEPETRIS, P.J., 2007, Discharge trends and flow dynamics of South American rivers draining the Southern Atlantic seaboard: an overview. *Journal of Hydrology*, **333**, pp. 385–399.

POVEDA, G. and SALAZAR, L.F., 2004, Annual and interannual (ENSO) variability of spatial scaling properties of a vegetation index (NDVI) in Amazonia. *Remote Sensing of Environment*, **93**, pp. 391–401.

PRESS, W.H., TEUKOLSKY, S.A., VETTERLING, W.T. and FLANNERY, B.P., 1992, *Numerical Recipes in Fortran* (New York: Cambridge University Press).

TUCKER, C.J., 1979, Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, **8**, pp. 127–150.

TUCKER, C.J., NEWCOMB, W.W., LOS, S.O. and PRINCE, S.D., 1991, Mean and interyear variation of growing-season normalized difference vegetation index for the Sahel 1981–1989. *International Journal of Remote Sensing*, **12**, pp. 1133–1135.

WOLTER, K. and TIMLIN, M.S., 1998, Measuring the strength of ENSO events – How does 1997/98 rank? *Weather*, **53**, pp. 315–324.

ZOFFOLI, M.L., KANDUS, P., MADANES, N. and CALVO, D.H., 2008, Seasonal and interannual analysis of wetlands in South America using NOAA-AVHRR NDVI time series: the case of the Parana Delta Region. *Landscape Ecology*, **23**, pp. 833–848.