Geophysical Research Abstracts
Vol. 19, EGU2017-8794, 2017
EGU General Assembly 2017

Monsoon Forecasting based on Imbalanced Classification Techniques

Pedro Ribera (1), Alicia Troncoso (2), Gualberto Asencio-Cortes (2), Inmaculada Vega (1), and David Gallego (1)

(1) Dept. Physical, Quemical and Natural Systems, Univ.Pablo de Olavide, Seville, Spain, (2) Division of Computer Science, Univ. Pablo de Olavide, Seville, Spain

Monsoonal systems are quasiperiodic processes of the climatic system that control seasonal precipitation over different regions of the world. The Western North Pacific Summer Monsoon (WNPSM) is one of those monsoons and it is known to have a great impact both over the global climate and over the total precipitation of very densely populated areas. The interannual variability of the WNPSM along the last 50-60 years has been related to different climatic indices such as El Niño, El Niño Modoki, the Indian Ocean Dipole or the Pacific Decadal Oscillation. Recently, a new and longer series characterizing the monthly evolution of the WNPSM, the WNP Directional Index (WNPDI), has been developed, extending its previous length from about 50 years to more than 100 years (1900-2007).

Imbalanced classification techniques have been applied to the WNPDI in order to check the capability of traditional climate indices to capture and forecast the evolution of the WNPSM. The problem of forecasting has been transformed into a binary classification problem, in which the positive class represents the occurrence of an extreme monsoon event. Given that the number of extreme monsoons is much lower than the number of non-extreme monsoons, the resultant classification problem is highly imbalanced. The complete dataset is composed of 1296 instances, where only 71 (5.47%) samples correspond to extreme monsoons. Twenty predictor variables based on the cited climatic indices have been proposed, and namely, models based on trees, black box models such as neural networks, support vector machines and nearest neighbors, and finally ensemble-based techniques as random forests have been used in order to forecast the occurrence of extreme monsoons. It can be concluded that the methodology proposed here reports promising results according to the quality parameters evaluated and predicts extreme monsoons for a temporal horizon of a month with a high accuracy.

From a climatological point of view, models based on trees show that the index of the El Niño Modoki in the months previous to an extreme monsoon acts as its best predictor. In most cases, the value of the Indian Ocean Dipole index acts as a second order classifier. But El Niño index, more frequently, or the Pacific Decadal Oscillation index, only in one case, do also modulate the intensity of the WNPSM in some cases.