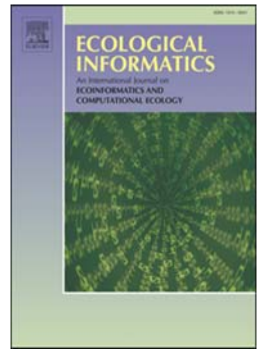

Evaluating Temporal Aggregation for Predicting the Sea Surface Temperature of the Atlantic Ocean

Rebecca Salles, Patricia Mattos, Ana-Maria Dubois Iorgulescu, Eduardo Bezerra, Leonardo Lima, Eduardo Ogasawara

PII: S1574-9541(16)30181-9
DOI: doi: 10.1016/j.ecoinf.2016.10.004
Reference: ECOINF 715

To appear in: *Ecological Informatics*

Received date: 15 March 2016
Revised date: 8 October 2016
Accepted date: 14 October 2016

Please cite this article as: Salles, Rebecca, Mattos, Patricia, Iorgulescu, Ana-Maria Dubois, Bezerra, Eduardo, Lima, Leonardo, Ogasawara, Eduardo, Evaluating Temporal Aggregation for Predicting the Sea Surface Temperature of the Atlantic Ocean, *Ecological Informatics* (2016), doi: 10.1016/j.ecoinf.2016.10.004

Evaluating Temporal Aggregation for Predicting the Sea Surface Temperature of the Atlantic Ocean

Rebecca Salles[a], Patricia Mattos[a], Ana-Maria Dubois Iorgulescu[a], Eduardo Bezerra[a], Leonardo Lima[a], Eduardo Ogasawara[a]

[a]*Federal Center of Technological Education of Rio de Janeiro (CEFET/RJ)*

Abstract

Extreme environmental events such as droughts affect millions of people all around the world. Although it is not possible to prevent this type of event, its prediction under different time horizons enables the mitigation of eventual damages caused by its occurrence. An important variable for identifying occurrences of droughts is the sea surface temperature (SST). In the Tropical Atlantic Ocean, SST data are collected and provided by the Prediction and Research Moored Array in the Tropical Atlantic (PIRATA) Project, which is an observation network composed of sensor buoys arranged in this region. Sensors of this type, and more generally Internet of Things (IoT) sensors, commonly lead to data losses that influence the quality of datasets collected for adjusting prediction models. In this paper, we explore the influence of temporal aggregation in predicting step-ahead SST considering different prediction horizons and different sizes for training datasets. We have conducted several experiments using data collected by PIRATA Project. Our results point out scenarios for training datasets and prediction horizons indicating whether or not temporal aggregated SST time series may be beneficial for prediction.

Keywords: prediction models, time series, sea surface temperature, Atlantic Ocean, temporal aggregation

1. Introduction

There are evidences in the literature pointing out to the existence of a causal relationship between the tropical southwestern Atlantic Sea Surface Temperatures (SST) and the occurrence of extreme weather events [13, 19, 32, 35, 42, 54]. In particular, the variability of Tropical Atlantic Ocean SST has a strong influence on the distribution of precipitations in Southern Tropical America [38], including the northeastern of Brazil [23] and southern-western of Amazon region [42, 62].

Among such extreme events, droughts are very relevant because of their economic impact. The causes for severe droughts in northeastern Brazil have been studied for a long time and currently there is an understating that main factors contributing to the occurrence of this natural disaster are related to SST, including both the El Niño Southern Oscillation (ENSO) and the Intertropical Convergence Zone (ITCZ) [15, 24, 33, 35].

In Brazil, records of droughts and its socioeconomic impacts are dated since the beginning of Portuguese colonization, but it was in the 17th century that governments began to take initiatives to mitigate their effects [24]. It is estimated that 32.8 million people have been affected by drought in the last thirty years, which lead to a loss of approximately $2.4 billion for the country [16].

Besides, other extreme events such as Hurricane Catarina [54], in southern Brazil, in 2004 draw the attention of researchers on the role played by SST variation in the Tropical South Atlantic. Catarina, which affected 150 thousand people, was the first hurricane ever registered in the south Atlantic and occurred in conditions very different from those usually observed.

There are also strong indications in literature that the SST has an extremely important role in other phenomena that occur in the tropical Atlantic Ocean, among them are: (i) the process that gives rise to tropical cyclones in the Atlantic Ocean [54]; (ii) the rainfall in the Amazon region [19]; (iii) the Amazon vegetative volume [13]; (iv) the carbon sequestration in the ocean [21].

From the aforementioned characteristics, it is clear that monitoring south Atlantic SST time series and having a more accurate model of its evolution under different prediction horizons could allow Brazilian government and society to better prepare themselves to droughts or floods in northeastern Brazil and the Amazon basin [58]. Although there are many works and models that focus on short-term SST prediction [2, 25, 32, 61], in this paper our goal is to analyze the usage of temporal aggregation for SST prediction under: (i) different prediction horizons and (ii) different sizes for training datasets.

Given the great diversity of extreme events and their commonly unstable nature, the analysis of (i) is often an important asset to studies that focus on the prediction of such events, since the desired prediction horizon depends on the characteristics of each natural event. The latter context (ii) is important since SST data are collected from different sources, such as sensors, which are not resilient to failures. Specially, after the advent of Internet of Things (IoT) [39], many IoT devices becomes vulnerable to communication interferences and instruments/sensors malfunction [4, 9, 10, 39, 56]. In the event of device failure, a problem that may occur is a significant lack of observations in SST time series. Addressing this is important to develop accurate prediction models [45]. Two main approaches are commonly applied: imputation and prediction using subsequences. Impu-