Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 114 (2017) 314–323

Procedia Computer Science

www.elsevier.com/locate/procedia

Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems, CAS October 30 – November 1, 2017, Chicago, Illinois, USA

Reducing Tropical Cyclone Prediction Errors Using Machine Learning Approaches

Michael B. Richman[a]*, Lance M. Leslie[a], Hamish A. Ramsay[b], Philip J. Klotzbach[c]

[a]*School of Meteorology, University of Oklahoma, Norman, OK 73072 United States*
[b]*School of Earth, Atmosphere and Environment, Monash University, Melbourne, 3800 Australia*
[c]*Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523 United States*

Abstract

Tropical cyclones (TCs) in the North Atlantic region are predictions for the June 1-November 30 season using predictors from Atlantic, Indian and Pacific Oceans sea surface temperature anomalies. Here, the aim is to reduce TC seasonal prediction errors and is realized by applying support vector regression (SVR) to an initial predictor pool and using the model prediction errors to iteratively identify additional attributes that reduce those errors. Prediction errors from this approach are compared with those from an existing, statistical seasonal prediction model, developed at Colorado State University (CSU). The SVR approach was optimized using attribute selection with wrapper selection techniques and by testing various kernels over a range of complexity parameter. Results of the comparison between seasonal SVR and the CSU TC model indicate that proper attribute selection lowers prediction errors significantly. Compared with the CSU model, the SVR model increases correlations between prediction and observed annual TC count from 0.62 to 0.83; the mean absolute error (MAE) is reduced from 2.9 to 1.8 and the root mean squared error (RMSE) drops from 3.8 to 2.7. Furthermore, the approach of using prediction errors to improve machine learning models is flexible and can be adapted readily to other TC basins.

Peer-review under responsibility of the scientific committee of the Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems.

Keywords: Machine Learning; Support Vector Regression; Kernels; Tropical Cyclones; Prediction Errors; Prediction; Attribute Selection

* Corresponding author. Tel.: +1-405-325-1853; fax: +1-405-325-7689.
E-mail address: mrichman@ou.edu

1877-0509 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the scientific committee of the Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems.
10.1016/j.procs.2017.09.048

1. Introduction

Tropical cyclones (TCs) rank high among the most devastating natural hazards. They are rotating low pressure tropical weather phenomena characterized by destructive winds and intense rainfall. When approaching coastal regions TCs are at their most threatening as, in addition to the strong winds and heavy rainfall, they can generate flooding from a combination of both massive rainfall amounts over land and storm surges generated over the oceans. TCs are an annual threat to many countries, with the potential to cause massive human casualties, damage to infrastructure, loss of crops and livestock, and irrevocable harm to ecological systems.

Given the immense socio-economic impact of TCs, especially those near land or making landfall, much research effort has been made to provide timely and skillful seasonal predictions of the number (and frequently the intensity and landfall locations) of TCs occurring in the TC seasons of the many regions affected by these storms. TC seasonal predictions for the North Atlantic region, based on statistical modeling, are issued by Colorado State University (CSU). These predictions commenced in 1984 [1], and have been posted every season from 1984 to the present. The statistical models used in prediction mode by CSU will be referred to hereafter as the CSU model. The oceanic and atmospheric predictors have changed since 1984 and the changes are summarized by [2]. An important aspect of the CSU model and the machine learning model developed here is that both utilize significant statistical relationships established between the total North Atlantic seasonal TC numbers and sea surface temperature (SST) anomalies (SSTA) as discussed, for example, by [3]. The SSTA are measured relative to long-term monthly mean SST values (e.g., the 1981-2010 mean SST). More recently, CSU's statistical models predict Accumulated Cyclone Energy (ACE; an integrated metric that accounts for frequency, intensity and duration of TCs), and the prediction for the number of TCs is derived from a linear regression between historical TCs and ACE.

Additional CSU predictions are generated during the TC season, but not addressed here, as our focus is on predictions issued before or at the beginning of the North Atlantic TC season (June 1). The CSU model predictions are used as the baseline relative to which the machine learning model developed in this study is compared. The CSU predictions were chosen primarily because of their lengthy prediction record of 32 years. It is noted also that the present study focusses only on the seasonal count of all TCs and does not provide separate count predictions of those TCs that reach hurricane strength or ACE.

2. Data and Methodology

2.1. Data

As mentioned in the Introduction, the TC season in the North Atlantic Ocean begins on June 1 and ends on November 30, and one of the CSU TC seasonal predictions is issued around June 1 for the upcoming TC season. Here, the CSU TC prediction issued around June 1 is compared to the observed number of TCs for the period 1984 to 2015, with the differences between the predictions and observations defining the CSU prediction errors. Monthly SST data are obtained from the Hadley Centre Sea Ice and Sea Surface Temperature dataset (HadISST; [4]) for the 32 year period 1984-2015, and anomalies are computed relative to the 1981-2010 monthly climatology.

Rank correlations were formed between the vector of annual total TCs and the set of monthly average SSTAs for January through May 1984-2015. The monthly rank correlations at every location in the global oceans between 60°N and 60°S were isoplethed to define monthly spatial fields. Rank correlation was used, rather than the more traditional Pearson correlation, as it measures nonlinear relationships in the values that are linear in the ranks (i.e., monotone relationships). Rank correlation also helps to reduce the impact of outliers, for example, the 2005 Atlantic hurricane season where 28 TCs were observed. Rank correlation coefficients with the most statistical significance and spatial coherence were represented as rectangles (Fig. 1). Each rectangle is treated as an attribute, lettered by basin and numbered (e.g., A1 for Atlantic region 1). For each rectangle, all gridpoints were averaged to obtain the mean monthly SSTA. These attributes served as the basis for a predictor pool, along with the trends in SSTAs for each rectangle for April to May. These two-way interactions between rectangles is defined as the multiplication of the SSTA between all sets of two boxes (e.g., A1*P1). The process and rationale of defining interaction terms in this manner is outlined in [5]. This combination of 5 months of candidate attributes times 6 attributes per month, 6 trend attributes and 75 interaction attributes results in 111 candidate attributes.