Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 61 (2015) 422 – 427

Procedia Computer Science

Complex Adaptive Systems, Publication 5
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2015-San Jose, CA

Diagnosing Tropical Cyclone Rapid Intensification Using Kernel Methods and Reanalysis Datasets

Andrew Mercer[a]*, Alexandria Grimes[a]

[a]*Department of Geosciences, Mississippi State University, Mississippi State, MS 39762, USA*

Abstract

Tropical cyclone rapid intensification (RI) continues to be a problem that eludes operational forecasters. Recent work in this area has revealed the value of applying machine learning techniques to classifying storms as RI or non-RI at 24-hours lead time. However, that work showed that differing reanalysis datasets represented the storms in unique ways, offering different discrimination capability and unique predictor sets that are important for RI. The scope of this research is to identify factors important for RI that are consistent among three reanalysis datasets, as these are likely the fields that will provide the greatest discrimination capability. An S-mode rotated principal component analysis was used to formulate unique patterns within RI and non-RI storms, and the resulting RPC scores were used to train a support vector machine classification algorithm that yielded binary RI occurrence output. Base-state meteorological variables (geopotential height, temperature, *u* and *v* wind components, vertical velocity, and relative humidity) at single horizontal levels were tested individually as predictors for the SVM. Base-state fields that were consistently good at discriminating RI events from non-RI events among all three reanalysis datasets were deemed most useful for RI classification and will be considered for future forecast applications.

Keywords: Tropical cyclones; support vector machines; kernel principal component analysis; kernel methods; reanalysis datasets

1. Introduction

Tropical cyclone (hereafter referred to as TC) forecasting continues to be a challenge in operational meteorology. TC forecasts consist of two primary forecast modes, track forecasts and intensity forecasts. Track forecasts are improving with continued updates to operational weather forecast models and the advent of the satellite era, yet

* Corresponding author. Tel.: 1-662-325-3915; fax: 1-662-325-9423.
E-mail address: a.mercer@geosci.msstate.edu

doi:10.1016/j.procs.2015.09.179

intensity forecasts continue to demonstrate limited effectiveness, primarily due to the limited understanding of the mechanisms driving TC intensification and the inability of weather models to render the important underlying thermodynamic processes. Forecasts of rapid intensification (hereafter referred to as RI), which involve a rapid strengthening of the cyclone over a short period of time (typically 24-hours), are of particular concern since most moderately strong (Category 2-3) TCs and all Category 4 and 5 storms undergo RI at some point in their life cycle [1]. Hence, RI forecasting remains a major forecasting challenge.

Initial attempts at forecasting RI in a TC's life cycle have primarily consisted of statistical approaches (i.e. multivariate linear regression [2]). These methods require prior knowledge of the variables relevant for RI, which typically consist of thermodynamic and kinematic fields. In particular, horizontal wind fields are shown to be critical in several kinematic studies on RI [3-5], while other studies reveal the importance of thermodynamic fields (mainly moisture and temperature or equivalent potential temperature) for RI [4, 6, 7]. Blends of kinematic and thermodynamic quantities are utilized as predictors in the most current operational RI classification model, the Statistical Hurricane Intensification Prediction Scheme Rapid Intensification Index (SHIPS-RII [1]). This model utilizes several kinematic variables (vertical wind shear, upper level divergence, potential intensity, and initial maximum sustained wind) and thermodynamic variables (low-level relative humidity, ocean heat content, total precipitable water, inner-core dry-air), as well as a predictor dealing with persistence (previous 12-hour intensification). It is evident from the literature on this topic that while a blend of kinematic and thermodynamic variables is useful, the exact physical mechanisms driving RI are still not well understood.

Recently, dynamic modeling of the TC environment has improved with the implementation of the Hurricane Weather and Research Forecasting (HWRF) model [8]. Recent work [9] showed vast improvements in HWRF's rendering of Hurricane Earl (2010) over previously utilized dynamic models. Eventually, the forecast skill of dynamic models such as HWRF will surpass statistical approaches, but until that time, forecast skill still remains sufficiently low to continue consideration of statistical models. Further, no machine learning techniques have been considered for the RI forecast problem, with the exception of [10], which utilized a single reanalysis dataset.

It is evident from previous work that without a dynamic rendering of the RI process, an optimal suite of predictors useful for discriminating RI/non-RI events is essential. The goal of this project is another look at the predictors useful for RI determination using kernel methods. In particular, three reanalysis datasets' [11-13] renderings of RI and non-RI storms will be formulated on individual base-state meteorological variables with the goal of identifying the variables with the greatest discrimination skill. Section 2 briefly describes the reanalysis datasets, while section 3 discusses the methods employed to discriminate RI and non-RI results. Section 4 shows results, while section 5 provides some discussion and conclusions.

2. Data

Since large-scale predictors were of interest for this work, reanalysis datasets, which consist of gridded three-dimensional large-scale representations of the atmosphere, were used. Three reanalysis datasets were tested, the NCEP/NCAR reanalysis dataset (NNRP – 11), the NCEP-DOE Reanalysis II dataset (DOE – 12) and the 20th century Regional Reanalysis dataset (20th – 13). Details of each dataset are provided in Table 1. Since each reanalysis dataset was formulated using a unique underlying dynamic modeling system, subtle but often important differences in their renderings of the atmosphere exist. As such, any "best classifying predictors" consistent among all reanalysis datasets would likely suggest important variables for discriminating RI and non-RI events. Six base-state meteorological fields for all vertical levels up to 100 mb were retained from each reanalysis dataset: geopotential height, temperature, u and v wind components, vertical velocity, and relative humidity. Each of these fields has high reliability in each reanalysis dataset, as they are primarily based upon observational data instead of model parameterizations, making them useful for the study.

Table 1. Summary of selected reanalysis datasets

Dataset name	Longitude-Latitude Resolution	Vertical Levels	Years available
NCEP/NCAR reanalysis (NNRP – 11)	2.5° x 2.5° (144x73)	17	1948-present
NCEP/DOE reanalysis II (DOE – 12)	2.5° x 2.5° (144x73)	17	1948-present
20th century reanalysis (20th – 13)	2° x 2° (180x91)	24	1871-present