ARTICLE IN PRESS

Journal of Hydrology xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Verification of the skill of numerical weather prediction models in forecasting rainfall from U.S. landfalling tropical cyclones

Beda Luitel [a], Gabriele Villarini [a,*], Gabriel A. Vecchi [b]

[a] IIHR-Hydroscience & Engineering, The University of Iowa, Iowa City, IA, USA
[b] NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Tropical cyclones
Rainfall
Numerical weather predictions
North Atlantic
Short-term forecasting

ABSTRACT

The goal of this study is the evaluation of the skill of five state-of-the-art numerical weather prediction (NWP) systems [European Centre for Medium-Range Weather Forecasts (ECMWF), UK Met Office (UKMO), National Centers for Environmental Prediction (NCEP), China Meteorological Administration (CMA), and Canadian Meteorological Center (CMC)] in forecasting rainfall from North Atlantic tropical cyclones (TCs). Analyses focus on 15 North Atlantic TCs that made landfall along the U.S. coast over the 2007–2012 period. As reference data we use gridded rainfall provided by the Climate Prediction Center (CPC). We consider forecast lead-times up to five days. To benchmark the skill of these models, we consider rainfall estimates from one radar-based (Stage IV) and four satellite-based [Tropical Rainfall Measuring Mission - Multi-satellite Precipitation Analysis (TMPA, both real-time and research version); Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN); the CPC MORPHing Technique (CMORPH)] rainfall products. Daily and storm total rainfall fields from each of these remote sensing products are compared to the reference data to obtain information about the range of errors we can expect from "observational data." The skill of the NWP models is quantified: (1) by visual examination of the distribution of the errors in storm total rainfall for the different lead-times, and numerical examination of the first three moments of the error distribution; (2) relative to climatology at the daily scale. Considering these skill metrics, we conclude that the NWP models can provide skillful forecasts of TC rainfall with lead-times up to 48 h, without a consistently best or worst NWP model.

1. Introduction

North Atlantic tropical cyclones (TCs) are responsible for significant societal and economic impacts. Over the 1900–2005 period, the average annual normalized damage associated with TCs in the continental United States is about $10 billion (value normalized to 2005 monetary value; Pielke et al., 2008). Overall, this damage accounts for close to half (Table 1 in Smith and Katz (2013)) of the total weather and climate disasters over the period of 1981–2011, much more than the damage associated with any other type of weather related disasters.

TCs are associated with multiple hazards, including strong winds, storm surges, heavy rainfall and flooding. While the effects of winds and surge are mostly felt along the coastal areas near the landfall location, heavy rainfall and flooding are responsible for significant damage over much larger areas, even hundreds of kilometers from the coast. More than 50% of the fatalities associated with TCs between 1970 and 2004 were caused by fresh water flooding (http://www.nws.noaa.gov/os/water/ahps/pdfs/Inland-FloodBrochure7F.pdf). Over the period 1963–2012, Rappaport (2014) showed that almost 50% of the U.S landfalling TCs have at least one fatality related to rain. Hurricane Ivan (2004) alone accounted for two-thirds of the total flood insurance payments made by the federal government in that year, impacting 23 different states (Czajkowski et al., 2013).

U.S. landfalling TCs are responsible for major flood events over large areas east of the Rocky Mountains, in particular along the eastern and central United States and along the coastal regions on the Gulf of Mexico (Villarini and Smith, 2010, 2013; Villarini et al., 2011, 2014). Although precipitation directly associated with TCs is less than 25% of the annual precipitation even in the most affected regions, the impacts can be extremely significant (e.g., Kunkel et al., 2010; Jiang and Zipser, 2010; Barlow, 2011).

* Corresponding author at: IIHR-Hydroscience & Engineering, The University of Iowa, 306 C. Maxwell Stanley Hydraulics Laboratory, Iowa City, 52242 IA, USA.
E-mail address: gabriele-villarini@uiowa.edu (G. Villarini).

http://dx.doi.org/10.1016/j.jhydrol.2016.09.019
0022-1694/© 2016 Published by Elsevier B.V.

Despite these negative socio-economic impacts, landfalling TCs have also been found to play a significant role as "drought busters" (e.g., Elsberry, 2002; Maxwell et al., 2012, 2013; Kam et al., 2013). Torrential rainfall associated with TCs occasionally can have the effect of breaking a prolonged drought by recharging reservoirs and elevating soil moisture. In these situations TC rainfall mitigates one environmental stressor, even as the potential for damage associated with the extreme rainfall, high-speed wind and ocean surge remain (e.g., Kam et al., 2013; Maxwell et al., 2013; Khouakhi and Villarini, 2016). Because rainfall associated with TCs has both significant positive and negative impacts on our society, it is critical that we understand how skillful current forecasting systems are in predicting rainfall associated with these storms to help us improve our preparedness and mitigation efforts.

Numerical Weather Prediction (NWP) models provide forecasts of a number of weather-related variables (e.g., precipitation, temperature at different levels) for different lead-times (e.g., Lorenc, 1986; Bougeault et al., 2010). However, quantitative information about the skill of NWP models in forecasting TC rainfall is still limited (Marchok et al., 2007; Mohanty et al., 2014). For a skillful prediction of TC rainfall, the models must predict the strength and distribution of the rainfall rate and wind fields together with the track and intensity of the TC system (see Halperin et al. (2013) for a discussion on the genesis forecasting of North Atlantic TCs). Therefore, precipitation forecasts from NWP models in general, and for TCs in particular, are inherently uncertain and subject to three types of error: localization, timing and intensity of precipitation events (e.g., Marchok et al., 2007). In this study, our goal is to evaluate the skill of NWP models in forecasting TC rainfall by quantifying their errors with respect to a reference (rain gauge-based) dataset. Moreover, five additional "observational" (remote sensing-based) datasets are also compared to the reference dataset: the skill of the NWP models in forecasting TC rainfall is quantified for different lead-times, and discussed and interpreted with respect to the performance of these "observational" products.

In this paper, the description of data and methodology is provided in Section 2, followed by results and discussion in Section 3. Section 4 summarizes the main points of the study and concludes the paper.

2. Data and methodology

We use the Climate Prediction Center (CPC) Unified Gauge-Based Analysis of Daily Precipitation over the continental United States. These data represent daily accumulations and are obtained by interpolating rain gauge measurements from a number of different networks and sources: the National Oceanic and Atmospheric Administration (NOAA)'s National Climate Data Center (NCDC) daily COOP stations, daily accumulations from hourly precipitation datasets, and the CPC dataset (it includes data from River Forecast Centers and 1st order stations). The spatial resolution is 0.25-decimal degree over the continental United States. There are different quality control steps that are implemented to remove duplicate and overlapping stations, buddy checks are used to eliminate extreme values, and standard deviation checks are used to compare the daily precipitation data against a daily climatology (Higgins et al., 2000). For the North Atlantic TC track information (date, time, latitude and longitude of all recorded storms with a 6-h resolution) we use the NOAA-Hurricane Research Division's Hurricane Database (HURDAT-2; Landsea and Franklin, 2013).

We evaluate the forecast rainfall produced by five state-of-art NWP models: European Centre for Medium-Range Weather Forecasts (ECMWF; Buizza et al., 2007), UK Met Office (UKMO; Bowler et al., 2008), National Centers for Environmental Prediction (NCEP; Toth and Kalnay, 1997), China Meteorological Administration (CMA), and Canadian Meteorological Center (CMC; Houtekamer et al., 2009). Data for NWP models have been archived from the THORPEX Interactive Grand Global Ensemble (TIGGE; Bougeault et al., 2010).

To benchmark the skill of these NWP models, we consider rainfall estimates from five remote sensing products (one ground based radar and four satellite-based rainfall products). Stage IV multi-sensor precipitation dataset is produced by NOAA-NCEP (Lin and Mitchell, 2005). It has ~4-km and hourly resolution, and is obtained by merging ground-based radars across the United States, and rain gauge measurements are used to perform bias correction. The four satellite-based rainfall products we use are: Tropical Rainfall Measuring Mission - Multi-satellite Precipitation Analysis [TMPA; both real-time (TMPA_RT) and research version (TMPA_RV); Huffman et al., 2010]; Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN; Sorooshian et al., 2000); CPC MORPHing Technique (CMORPH; Joyce et al., 2004). These satellite-based products have a resolution coarser than Stage IV (3-hourly and 0.25-degree), and TMPA-research version is the only product for which a monthly bias correction with respect to rain gauges is applied.

The rainfall products have different spatial resolutions. Stage IV has the highest spatial resolution (~4 km), the four satellite-based remote sensing products and the CPC data have 0.25×0.25 degree lat/lon spatial resolution, whereas all the NWP model output we have explored are on a 0.5 degree lat/lon spatial resolution grid (the default resolution in the TIGGE archive). Because our focus is on the NWP models, all the products were regridded into 0.5-degree resolution to establish uniformity in the analysis, with the expectation that both the agreement between observational estimates and the forecasting skill will increase as we coarsen the spatial resolution (e.g., Lavers and Villarini, 2013). Results are based only on rainfall over land.

We focus on the evaluation of the precipitation associated with North Atlantic TCs that affected the continental United States over the period 2007–2012, and examine 15 storms that came within 500 km of the coast of the United States. We consider TC-rainfall the rainfall that occurred within a 500-km buffer around the center of circulation of a given storm. We compare the storm total rainfall obtained from the CPC data (our reference) against the rainfall forecasts from the five NWP models. To quantify how close (or far) these forecasts are from the reference data, we also use rainfall estimates from the five remote sensing products to give a range of potentially acceptable results. The availability of these remote sensing products allows us to complement the "absolute" evaluation of the performance of the NWP models with an assessment that is "relative" to what obtained when using "observational" records. For instance, we could use the correlation coefficient as skill metric and obtain a value of 0.4 between CPC and forecast rainfall. While we can interpret this number in an absolute sense (i.e., 0.4 on a scale from −1 to +1), we can also interpret it in a relative sense: the 0.4-correlation value has a different interpretation if the range of values we get from the remote sensing products is between 0.9 and 0.95, rather than between 0.3 and 0.45. Therefore, it will provide us with an additional way of benchmarking the NWP models.

Here we use Hurricane Irene (2011) as an example of our approach and methodology (Fig. 1; the results for the other 14 storms are in Supplementary Figs. S1–S14). Our approach is to examine the skill of the forecasts starting from the first time the center of circulation of the storm is within 500 km from the U.S coastline. We will refer to this as the "0-h lead-time." In the example in Fig. 1, the "0-h lead-time" represents the storm total rainfall from 27 August 2011 at 12 UTC to 31 August 2011 at 12 UTC. We will refer to the "12-h lead-time" the forecast for the period from 27 August 2011 at 12 UTC to 31 August 2011 at 12 UTC initialized