Seasonal Forecasting of Tropical Cyclone Landfall Using ECMWF's System 4

DAN L. BERGMAN

Third Swedish National Pension Fund, and Department of Meteorology, Stockholm University, Stockholm, Sweden

LINUS MAGNUSSON

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

JOHAN NILSSON

Department of Meteorology, Stockholm University, Stockholm, Sweden

FREDERIC VITART

European Centre for Medium Range-Weather Forecasts, Reading, United Kingdom

(Manuscript received 21 February 2018, in final form 26 June 2019)

ABSTRACT

A method has been developed to forecast seasonal landfall risk using ensembles of cyclone tracks generated by ECMWF's seasonal forecast system 4. The method has been applied to analyze and retrospectively forecast the landfall risk along the North American coast. The main result is that the method can be used to forecast landfall for some parts of the coast, but the skill is lower than for basinwide forecasts of activity. The rank correlations between forecasts issued on 1 May and observations are 0.6 for basinwide tropical cyclone number and 0.5 for landfall anywhere along the coast. When the forecast period is limited to the peak of the hurricane season, the landfall correlation increases to 0.6. Moreover, when the forecast issue date is pushed forward to 1 August, basinwide tropical cyclone and hurricane correlations increase to 0.7 and 0.8, respectively, whereas landfall correlations improve less. The quality of the forecasts is in line with that obtained by others.

1. Introduction

Tropical cyclones (TCs) are among the most destructive geophysical phenomena on Earth (Emanuel 2005; Pielke et al. 2008; Blake et al. 2011). They cause extensive damage and result in catastrophic loss of life. Efforts to forecast tropical cyclone activity on a seasonal time scale were pioneered by Nicholls (1979) and Gray (1984). In 1984, Gray demonstrated a link between hurricane activity and slow irregular climate modes, most notably El Niño–Southern Oscillation (ENSO). The persistence and predictability of such slow modes form an important basis for seasonal hurricane forecasting today. However, the natural variability of the climate system is substantial, and

many weather patterns are governed by inherently chaotic dynamics. As a result it is challenging to forecast hurricane activity on a seasonal time scale (Camargo et al. 2010; Barnston et al. 2012; Duan and Wei 2013).

The European Centre for Medium-Range Weather Forecasts (Vitart and Stockdale 2001; Molteni et al. 2011), the Met Office (Camp et al. 2015; MacLachlan et al. 2015), the National Oceanic and Atmospheric Administration (Saha et al. 2014), and Colorado State University (Klotzbach and Gray 2009) among others produce seasonal forecasts of basinwide tropical cyclone activity. However, despite the great societal interest in preparing and warning for potential damages ashore, neither ECMWF, nor the Met Office, nor NOAA, publish seasonal forecasts of landfall risk. This is linked to difficulties in producing forecasts of sufficient quality and usefulness. The continuous development of numerical coupled atmospheric–ocean prediction models by ECMWF and others has now reached a stage where

Denotes content that is immediately available upon publication as open access.

Corresponding author: Dan L. Bergman, dan.bergman@ap3.se

it is worth reexamining the possibility to forecast landfall risk at the seasonal time scale. Work along these lines has been reported by Vecchi et al. (2014) pointing toward the possibility to forecast TC activity on finer than basinwide scales and Manganello et al. (2016) demonstrating increased skill in forecasting TC activity in land-adjacent areas using high-resolution models. Also Camp and Caron (2017) and Murakami et al. (2016b) have reported similar results (see section 4).

Ideally, we would like to forecast tropical storm and hurricane landfall, intensity, size, and forward speed as well as other parameters critical for damage and destruction. However, it is natural (and challenging enough) to start by forecasting the risk of the landfall itself because it is a key factor impacting the extent of the damage caused.

The present paper reports the results of an investigation of the possibility to use ECMWF's seasonal forecast system 4 (System 4) to forecast TC and hurricane landfall risk along different segments of the North American Atlantic and Gulf coasts, with a particular focus on the U.S. part of the coast. For this purpose, a method has been developed to forecast landfall using the ensembles of storm tracks generated by System 4, and the quality of the forecasts has been analyzed. All Caribbean islands including Cuba and Hispaniola were excluded from the analysis.

The paper is organized as follows: first, in section 2, we describe the main data sources and the method developed; next, in section 3, we describe the results of the forecasts and their quality; and, in section 4, we discuss the results and draw some conclusions regarding the method and its applicability.

2. Method and data

In this section, the method developed to forecast landfall risk is described. The method uses System 4's ensembles of storm tracks and historical data from HURDAT2 (Landsea and Franklin 2013) to construct the forecasts. We will use the term LF+ to denote the method developed here. (L refers to landfall, F to forecast, and + to the fact that the method depends on and can be viewed as an extension of System 4.) We will refer to System 4 together with LF+ as S4LF+.

The main components of the method are (i) mapping of System 4 wind speeds to corresponding real wind speeds, (ii) determination of track extent and landfall, and (iii) calculation of forecasts of different variables, for example the number of expected landfalls along a segment of the coast.

a. Observation and historical record

For the North Atlantic (including the Gulf of Mexico and the Caribbean) NOAA has compiled data on all

known tropical cyclones dating back to 1851 (Landsea and Franklin 2013). We have used their most recent database, HURDAT2 (access date 4 June 2015), for construction and validation of our forecasts.

The number of tropical storms and hurricanes forming over the Atlantic and Gulf of Mexico varies considerably from year to year as does their intensity, duration, and the number that makes landfall. Figure 1 shows the number of tropical cyclones and hurricanes formed (first row), the activity measured as cyclone days (second row), and the number of landfalls in the conterminous United States between the points c and g in Fig. 2 (third row) for each year during the period 1900–2014. As can be seen, there is a considerable interannual variability in the historical record. Increasing trends in TC and hurricane number and days can be observed. These trends are consistent with underreporting or missing data before systematic aircraft reconnaissance in the 1950s and the launch of weather satellites in the 1960s (see e.g., Vecchi and Knutson 2011, Villarini et al. 2012, and Landsea et al. 2010). The landfall data is more reliable, and no statistically significant trend can be found.

To design and assess the forecast method, we need climatological estimates of the mean and standard deviation of the relevant variables (number, days, and landfall). To obtain such estimates we need to select time periods that yield robust and representative estimates. We have selected two different time periods: 1950–2014 for number and days and 1900–2014 for landfall. Table 1 shows the estimated means and standard deviations. While the satellites made the record nearly complete post 1970, the improved quality of the record has to be weighed against the smaller sample size as well as poorer sampling of for example the Atlantic multidecadal oscillation (AMO) and other slow dynamics affecting the frequencies of tropical cyclones, hence we have chosen the longer time periods. For days and number we did not include the 1940s because there are indications that the quality of the record improved from about 1950 (Vecchi and Knutson 2011). For landfall we did not include the later part of the 1800s for the same reason (Landsea et al. 1999).

Clearly the time series in Fig. 1 are nonstationary, but if we nevertheless assume that the observed number of TCs for each year during the selected time period represents a random draw from an unknown but constant probability distribution representing the “true climatology”, then the sample mean and standard error represents our statistically best estimates of the distributions mean and standard deviation. In a way, this is the simplest possible hypothesis: we assume that we know nothing about the state of the AMO, the ENSO, or climate

FIG. 1. Annual observed TC and hurricane number, days, and U.S. landfall as recorded in HURDAT2 (red), rolling 10-yr averages (blue), and long-term averages (black).

change, et cetera. Moreover, under this hypothesis, we can use bootstrapping to estimate confidence intervals for the mean and standard deviation (Press et al. 1992). Table 1 compiles the estimated means and standard deviations for the key TC- and hurricane-related variables. The width of the confidence intervals quantifies in the uncertainty associated with the historical means and standard deviations. We will use some of these statistics to calibrate our forecast model, and the confidence intervals will provide an indication of the size of the systematic errors (or biases) that we can expect when scoring the forecasts against observation (cf. Table 1 and Table 2).

b. ECMWF's seasonal forecast system

As mentioned above, ECMWF's seasonal forecast system 4 forms the basis for the method to forecast landfall that we describe here (Molteni et al. 2011). System 4 was operational between November 2011 and October 2017 and is the result of extensive research and development over the last three decades at ECMWF.

In brief, System 4 has been developed and designed to forecast global atmospheric as well as oceanic conditions

on a seasonal time scale up to about a year. The system has two main components, an oceanic model, NEMO,¹ and an atmospheric model, ECMWF's Integrated Forecasting System (IFS) Cycle 36r4. The two model components have been coupled² to capture atmospheric and oceanic interactions and phenomena that can be important for hurricane forecasts, ENSO being one example (Gray 1984; Bove et al. 1998).

The so-called ORCA1 grid configuration is used for the oceanic model. This configuration has a horizontal resolution of about $1^\circ \times 1^\circ$ in the midlatitudes (equatorially refined) and in the vertical 42 levels are used, 18 of which are in the first 200 m. The spectral resolution of the atmospheric model is T255 in the horizontal, and gridpoint calculations are done on a reduced Gaussian N128 grid, which corresponds to a resolution of about 0.7° . In the vertical, 91 levels are used with the top level at 0.01 hPa. The time step in the atmospheric model is

¹ Nucleus for European Modeling of the Ocean.

² A version of the OASIS3 coupler (developed at CERFACS) has been used.

FIG. 2. Points showing the division of the North American coastline into segments. Statistics on the segments are shown in Table 5.

45 min, the time step in the oceanic model is 60 min, and the models are coupled every three hours.

In this study, retrospective forecasts (so-called reforecasts) and forecasts³ for the period 1981–2014 have been used. For each year an ensemble of 51 reforecasts starting on 1 May and an ensemble of 51 reforecasts starting on 1 August each spanning seven months ahead have been used. See Molteni et al. (2011) for details.

c. The ECMWF tracker

To identify and track tropical cyclones from the model fields forecasted by System 4, a so-called *objective tracker* has been used. [The model cyclone tracks and related climatology are described in Bergman et al. (2017, 9–10 and 29–37).] Briefly, the ECMWF tracker at each time step first identifies potential tropical cyclones using the criteria developed by Vitart et al. (1997), and then it associates cyclones at different time steps with each other using the criteria developed by Van der Grijn et al. (2005), thereby building the track.

Key criteria used in identifying potential tropical cyclones are 1) the presence of a local maximum in the vorticity, 2) the presence of a warm core, and 3) a maximum surface wind speed of at least 25 kt (13 m s^{-1}). The

warm core is identified by the presence of a maximum of temperature averaged between 200 and 500 hPa above the center of the cyclone and by the presence of a maximum of thickness between 850 and 200 hPa. The distance between the center of the warm core and the center of the cyclone (defined by the local minimum in sea level pressure) must not exceed 8° latitude; from the center of the core, the temperature must decrease by at least 0.5°C in all directions within a distance of 10° latitude; and the thickness between 850 and 200 hPa must decrease by at least 50 m. Moreover, the local vorticity maximum should be larger than $3.5 \times 10^{-5} \text{ s}^{-1}$ at 850 hPa and the maximum and the 25-kt winds ($1 \text{ kt} \approx 0.51 \text{ m s}^{-1}$) must fall within an 8° circle centered on the middle of the cyclone. The threshold of 25 kt was chosen to get the most realistic total climatological number of tropical cyclones over the Atlantic in the model reforecasts. Only cyclones meeting the above criteria at least three times (time step 12 h) during their life are kept. The warm core criterion needs only to be fulfilled once during the life of the cyclone (see Vitart et al. 2011 for further details).

d. Intensity calibration

While System 4 generates tropical depressions with a warm core and sufficient strength to make them dynamically similar to observed tropical cyclones, the resolution of System 4 is not sufficient to fully capture the intense winds near the center of a hurricane. This and other limitations make forecasts of cyclone frequency, landfall, and in particular cyclone intensity challenging. Care must be exercised when using System 4 for this purpose.

We have devised a statistical quantile-quantile (Q-Q) approach to map the intensity obtained by applying the tracker to the System 4-generated fields onto a “true” intensity. The result is that, at each point in time, the cyclones with a System 4 wind speed of 24 kt or above will be regarded as tropical cyclones and those with a wind speed of 31 kt or above will be regarded as hurricanes. See Bergman et al. (2017; their appendix A) for further details.

e. Landfall determination

To determine landfall, we have constructed a global land mask with a $0.05^\circ \times 0.05^\circ$ resolution, from the Group for High-Resolution Sea Surface Temperature’s (GHRSSST) high-resolution land mask.⁴ For each point within 500 km of land we have calculated the shortest distance to land, to detect cyclones that pass land close

³ For the last three years, 2011–14 real-time data were used.

⁴ See GHRSSST for more information <https://www.ghrsst.org/products-and-services/tools/navo-ghrsst-pp-land-sea-mask/>.

TABLE 1. Mean and standard deviation for TC and hurricane number, days, and U.S. landfall estimated from HURDAT2. “Best” denotes the corresponding best estimate, and lower and upper are the bounds of the 95%, two-sided confidence intervals.

		Mean			Standard deviation		
	Period	Best	Lower	Upper	Best	Lower	Upper
TC							
Number	1950–2014	10.8	9.9	12.0	4.2	3.6	5.6
Days	1950–2014	56	50	64	27	24	33
U.S. landfall	1900–2014	4.0	3.6	4.5	2.4	2.1	2.9
Hurricane							
Number	1950–2014	6.1	5.5	6.8	2.7	2.2	3.4
Days	1950–2014	23	20	27	14	12	16
U.S. landfall	1900–2014	2.0	1.7	2.3	1.7	1.5	2.0

enough to make damage even though the eye of the cyclone does not cross the coastline. The Caribbean islands are not included in this study, nor is the continent south of 10°N or north of 52°N.

To make the landfall statistics as comparable as possible, the same land mask and criteria for landfall were applied to the forecasts and the historically observed tracks (HURDAT2). The track data points consist of the position, maximum wind speed, and minimum pressure. Each point is determined to either be over land or over sea, using the position and the land mask. Note that points within 36.2 km of land are regarded to be over land; as, even if the center of the cyclone does not cross the coastline, damaging winds are likely to occur ashore if a hurricane is within 1.5 times the radius of maximum winds from land. The average radius of maximum winds for landfalling U.S. hurricanes has been taken from [Blake et al. \(2011\)](#). A detailed description of the algorithm used to identify landfall can be found in [Bergman et al. \(2017\)](#). Figure 3 shows examples of tracks and landfall points identified by this algorithm. The significant differences between the tracks generated by S4LF+ and those of HURDAT2 illustrates both the chaotic (on a seasonal time scale hard to predict) steering patterns as well as the forecast model limitations.

f. Forecast calculation

In the previous sections, we have described the algorithms and methods developed to process the tracks, estimate the cyclone intensity, and determine landfall. Here we describe the last step, that is, the method to calculate the values of the variables that we seek to forecast. We will refer to this step as the *forecast calculation*.

The main variables we seek to forecast are TC and hurricane number, days, and landfall. Landfall will be predicted for different segments of the North American coast, the main segments being the full coast (North American landfall) and the U.S. part of the coast (U.S.

landfall), but other shorter segments will also be considered (see [Fig. 2](#) and [Table 5](#)). If two TCs occur at the same time both will count toward the number of TC days. And, if a TC makes multiple landfalls, each landfall will count, provided that they are separated by at least 300 km. The variables for hurricane strength systems are named analogously. We have chosen the number of TC days, not the accumulated cyclone energy (ACE), to estimate the total basinwide activity of a given season, because the wind speed estimated by the tracker is subject to substantial uncertainty (see [Bergman et al. 2017, 29–30](#)), an uncertainty that is amplified by the quadratic dependence of the ACE on wind speed.

The landfall forecasts for System 4 are calculated for the 34-yr period between 1981 and 2014, which is the full time period for which ensemble reforecast data was available. We will refer to this 1981–2014 period as the reforecast period, and all time averages of System 4 model data will be taken over this period. The HURDAT2 observational data from the period 1900–2014 is used to estimate the expected means of the distributions of the variables that we seek to forecast. To obtain robust estimates of the means, two different time periods are used: 1950–2014 for the basinwide variables (number and days) and 1900–2014 for the landfall variables. The choice of these time periods is discussed in [section 2a](#). The Atlantic hurricane season officially starts on 1 June and ends on 31 November. Note that our use of 1 May instead of 1 June as start date of the season is nonstandard, but of little significance since few cyclones form in May.

To describe the forecast calculation, we use some notations: Let t denote the year, $X(t)$ the variable we seek to forecast (e.g., the number of TCs that make landfall), $x(t)$ the observed value of the variable (i.e., the outcome), and $\hat{x}_i(t)$ the value of the variable according to our forecast model's i th ensemble member. Note that we regard $X(t)$ as a stochastic variable and even a perfect forecast model will sometimes yield a

TABLE 2. Mean forecast and observation, their difference (systematic error), and the bias corrected RMSE for the period 1981–2014 for key TC and hurricane variables.

Variable	TC				Hurricane			
	Forecasted	Observed	Difference	RMSE	Forecasted	Observed	Difference	RMSE
Number	10.8	11.9	−1.1	4.1	6.5	6.4	0.1	2.6
Days	45	63	−18	26	16	23	−7	12
North American landfalls	8.7	7.6	1.1	4.1	3.7	3.1	0.6	2.6
U.S. landfalls	4.0	4.0	0.0	3.0	2.2	1.7	0.5	2.1

forecast $F[X(t)]$ that differs from the subsequently observed outcome $x(t)$ due to the chaotic dynamics embedded in the climate system. Moreover, we denote time averages by $\bar{x}$ and ensemble means by $\langle \hat{x}(t) \rangle$. Note that time averages are taken over three different time

periods (see the previous paragraph). There are 51 ensemble members both for the reforecasts starting on 1 May and for those starting on 1 August. For each season and variable, the basic idea is to construct the forecast $F[X(t)]$ by rescaling ensemble

FIG. 3. HURDAT2 and S4LF+ tracks for the hurricane season 2003. Genesis points are marked by yellow, termination points by red, and landfall points by blue dots. For S4LF+ 50 randomly selected tracks are shown.

mean of the season using one of two factors: f_b for the basinwide variables and f_l for the landfall variables. The former is computed as

$$f_b = \frac{\bar{x}_b}{\langle \hat{x}_b \rangle}, \quad (1)$$

where x_b is the TC number variable, $\bar{x}_b$ its historically observed average over the 1950–2014 period, and $\langle \hat{x}_b \rangle$ the 1981–2014 average of the ensemble mean TC number. Similarly, f_l is computed as

$$f_l = \frac{\bar{x}_l}{\langle \hat{x}_l \rangle}, \quad (2)$$

where x_l is the U.S. TC landfall variable, $\bar{x}_l$ its historically observed average over the 1900–2014 period, and $\langle \hat{x}_l \rangle$ the 1981–2014 average of the ensemble mean number of U.S. TC landfalls.

Each forecast is then constructed as

$$F[X(t)] = f \times \langle \hat{x}(t) \rangle, \quad (3)$$

where f denotes f_b or f_l depending on the variable forecasted and $\langle \hat{x}(t) \rangle$ denotes the ensemble mean for season t for the variable that we seek to forecast.

By scaling the ensemble means in this fashion, the forecasts become more robust, especially the hurricane and landfall forecasts. This is key, especially when forecasting landfall for shorter segments of the coastline where landfall is infrequent.

Another consequence of the scaling, is that we introduce a systematic error or bias when scoring the forecast against observation for the reforecast period. In the case of the TC number forecast, the bias is exactly equal to the difference between the observed averages for the 1950–2014 and the 1981–2014 periods. Similarly for the U.S. TC landfall forecast, the bias is exactly equal to the difference between the observed averages for the 1900–2014 and the 1981–2014 periods. For the other variables the bias is impacted by other factors as well.

This approach is similar, but not identical, to the approach used by the ECMWF in the operational forecasts of basinwide TC and hurricane frequency. The difference being our choice of factors and longer time periods for the historical averages that we expect to yield better forecasts going forward (again, cf. section 2a).

One may also note that this method does not make any use of the information present in the distribution of the ensemble of reforecasts other than that captured by the mean, for example it does not make use of the minimum, the maximum, the dispersion, or the skewness.

3. Results

All the forecasts that we describe in this section have been made using our System 4–based forecast method, LF+, described in section 2. The forecasts have been issued on 1 May for the period 1 May–30 November (i.e., the official hurricane season plus May), except in section 3f where the impact of changing the lead time and the forecast period is described. Except where otherwise is explicitly stated, only forecasts and statistics for the period 1981–2014 are discussed in this section.

a. Mean error and RMSE

To obtain an overview of the forecasts, one may compare time series of the forecasts to observations (HURDAT2). Figure 4 shows such time series for two key basinwide variables—TC number and TC days—as well as for two key landfall variables—TC North American landfall and TC U.S. landfall.

First, we may note that the variability in the observations is considerable and contains several extreme years with a high number of systems, a large number of TC days, and a large number of landfalls. A significant part of the variability is driven by the chaotic dynamics embedded in the climate system. The variability in the forecasts is smaller. This is expected because the forecasts are based on ensemble means, which suppresses the impact of the chaotic dynamics.

Although less apparent in the figure, there are, by construction, systematic errors in the forecasts when evaluated against observation for the 1981–2014 period. The systematic errors arise largely from the fact that the forecasts are calibrated using long-term averages of HURDAT2 observations (the time periods used were 1900–2014 in case of landfall and 1950–2014 in case of number and days variables), whereas the systematic error is computed for the time period 1981–2014 (that is the full time period for which we had access to System 4 forecasts). For example, the observed TC number was 11.9 for the 1981–2014 period and 10.8 for the 1950–2014 period, giving rise to an average forecast of 10.8 and a systematic error of 1.1 TCs per year. We accept this systematic error or bias in order to gain robustness (see sections 2a and f).

Table 2 shows the mean forecasts and observations as well as the systematic (mean) errors and the bias corrected root-mean-square errors (RMSEs) in the forecasts. One may note that, the forecasted number of TC days is lower than observed. Part of this is by construction as described above, and part of it arises from a too short average life time of the model cyclones (4.2 days) compared to observation (5.3 days). For the number of hurricanes and the number of U.S. landfalling

FIG. 4. Time series of forecast (red) and observation (blue) for TC number, days, North American landfall, and U.S. landfall. Shaded areas show the forecast spread (5th, 16th, 84th, and 95th percentile) in the forecast ensemble.

TCs the systematic error is small, however this does not necessarily mean that we will be better at forecasting these variables, rather it is the result of coincidental overlaps of different historical averages.

The bias corrected RMSE provides another view of the error associated with the forecasts; in the case of TC number the RMSE is 4.1 and in the case of hurricane number the RMSE is 2.6. Clearly the RMSEs are larger than the systematic errors in both cases; nevertheless, the relative RMSEs (that is RMSEs divided by the means of the corresponding observations) are 35% and 41% for TCs and hurricanes, respectively. The relative RMSE in the forecast for North American TC landfall is 48%, which is comparable to the relative RMSE in the basinwide forecasts of TC and hurricane number and days. This is encouraging because the accuracy in seasonal forecasting of basinwide activity has been judged sufficient to motivate the regular dissemination of forecasts by ECMWF, NOAA, and the Met Office. The large relative RMSEs are in part the result of the chaotic nature of the climate system, which introduces a considerable degree of unpredictable variability. For the landfall forecasts the relative RMSEs are significantly higher than for the basinwide variables. While discouraging, this is in part expected due to the influence of short-lived atmospheric steering conditions on

hurricane and TC landfall. Clearly, understanding the chaotic and unpredictable part of the climate system is important when assessing the quality of the forecasts and what room there may be for improving them. We discuss this further and attempt to quantify the unpredictable part of the variability in [Bergman et al. \(2017\)](#).

b. Correlation

In this section we look closer at the forecasts and how we may quantify any skill associated with them. Before jumping into statistical analysis, it may be useful to build an intuition regarding the strength of the association between forecast and observation (outcome) by plotting them against each other in the key cases. [Figure 5](#) shows such plots for the two key basinwide variables—TC number and TC days—and the two key landfall variables—North American and U.S. TC landfall.

The main statistic that we have used to quantify the skill of the forecasts is the Spearman rank correlation ρ . We will base our analysis on the Spearman rather than the Pearson correlation r , because we expect it to suppress the impact of outliers and be more robust given the limited sample size, the associated noise, and the nonnormal fat-tailed distributions of the variables. One should however note that we plot the variables

FIG. 5. Scatterplots of observation vs forecast for TC number, days, North American landfall, and U.S. landfall. The one-to-one line is solid, and the regression line is dashed.

themselves, not their ranks, in Fig. 5, hence the slope of the regression line is more closely associated with the Pearson than the Spearman correlation. For comparison we show both Spearman and Pearson correlations and associated p values for four of the key variables forecasted in Table 3. The differences between the two are less than 0.1 in the cases where the correlations are significantly different from zero at the 99% level. In the following we will only use the Spearman correlation except where the other is explicitly stated.

As can be seen in the table, the rank correlations for TC and hurricane number as well as for TC and hurricane days are close to 0.6 with p values below 0.001 in all cases. For TCs making landfall anywhere along the North American coast, $\rho = 0.53$ with a p value of 0.001. Thus, in these cases we have a clear association between forecast and observation. In contrast, one may

note that the rank correlation is only weakly positive (0.27) for U.S. TC landfall and the associated p value is 0.13. For hurricane landfall along the U.S. coast, the rank correlation is even lower and nonsignificant. (For landfall statistics on different segments of the coast, see section 3d.)

Because there are fewer hurricanes than tropical cyclones and because of the limited forecast model resolution, one may ask if it would not be better to use the tropical cyclone forecasts to forecast hurricane activity and landfall. A comparison of the correlation between the TC number forecast and the observed hurricane number ($\rho = 0.52$, p value = 0.0015) and the correlation between hurricane number forecast and observation ($\rho = 0.55$, p value = 0.0007) indicates that this is not the case for hurricane activity in the basin. The correlation between the TC North American

TABLE 3. Spearman rank ρ and Pearson r correlations for forecasts of key TC and hurricane variables for the period 1981–2014. The p values are two sided.

Variable	TC				Hurricane			
	ρ	p value	r	p value	ρ	p value	r	p value
Number	0.57	0.0004	0.54	0.001	0.55	0.0007	0.56	0.001
Days	0.61	0.0001	0.54	0.001	0.61	0.0001	0.56	0.001
North American landfall	0.53	0.001	0.45	0.008	0.29	0.10	0.21	0.23
U.S. landfall	0.27	0.13	0.08	0.63	0.11	0.55	0.03	0.85

landfall forecast and the observed hurricane North American landfall ($\rho = 0.35$, p value = 0.10) is somewhat higher than the correlation between the direct forecast of hurricane North American landfall and observation ($\rho = 0.28$, p value = 0.04), indicating that one may be able to gain some forecast skill by using all model TCs in forecasting landfall rather than only those with the highest wind speeds corresponding hurricane strength, but the difference is small and may not be significant.

In summary, the data and analyses described show that there is a moderate association between forecast and observation for North American TC landfalls comparable in strength to the associations for the basinwide variables TC number and TC days; however, no clear association can be spotted in the case of U.S. landfall. For hurricane strength systems, the associations for the basinwide variables are equally strong as for the TC strength systems, but no clear associations are present for the landfall variables.

c. Contingency

Another basic way to quantify the skill of the forecasts is by classification of each year either as a low, normal, or high year as follows: For example, out of the 34 years considered, the 11 years with the lowest number of North American storm landfalls are placed into the low category, the 12 next into the normal category, and the 11 last into the high category. Similarly, the forecasts are ranked from lowest to highest and the 11 with the lowest rank are classified as low, the next 12 as average,

and the last 11 as high. NOAA issues seasonal forecasts for basinwide activity with a similar classification.

Using this classification, we can easily compare the number of correct and incorrect forecasts for each one of the categories (see Table 4). For example, in the case of North American TC landfall, in total 14 out of the 34 forecasts were correct and the remaining 20 were incorrect. As a comparison, if we were to issue 34 forecasts randomly, we would on average expect 11.4 to be correct. However, only once was a low North American landfall season forecasted when it turned out to be a high season. Thus, when we forecast a low season we can with a relatively high degree of confidence assume that it will not turn out as a high season; and, similarly, when we forecast a high season, we may be rather confident that it will not turn out as a low season. If our forecast points to a normal season, the outcome is highly uncertain. In contrast, if we look at U.S. TC landfall forecasts, high and low forecasts, are more often completely wrong.

When we calculate the skill score $S = (C - E)/(T - E)$, where C is the number of correct forecasts, T the total number, and E the number expected to be correct by chance (Panofsky and Brier 1958) for our forecasts of TC number, days, North American landfall, and U.S. landfall we obtain 0.38, 0.29, 0.12, and 0.07, respectively. Similarly we obtain 0.34, 0.34, 0.16, and -0.02 for the corresponding hurricane forecasts. A random forecast would score 0 and a perfect forecast 1. The fact that the North American landfall forecasts rarely are

TABLE 4. Forecasts and observation of low, normal, and high seasons.

		Forecast			Forecast		
		Low	Normal	High	Low	Normal	High
		Number			Days		
Observed	Low	9	2	0	7	3	1
	Normal	0	6	6	3	5	4
	High	2	4	5	1	4	6
		North American landfall			U.S. landfall		
Observed	Low	6	4	1	6	2	3
	Normal	4	3	5	3	4	5
	High	1	5	5	2	6	3

FIG. 6. Average number of landfalls per year along different segments of the North American coastline. Blue bars show the numbers forecasted for the 1981–2014 period, red crosses show the average numbers observed for the 1900–2014 period, and the 95% confidence intervals show the uncertainty in the historical averages. See Fig. 2 for the locations of the start and end points of the segments.

completely off as discussed in the previous paragraph is not captured by this metric.

d. Landfall along different segments

To better understand the method's ability to forecast landfall, the North American coastline was divided into different segments. Figure 2 shows the points along the coastline that were used to define the segments, the longest segment (points a–h) corresponds to what we call the North American coast. By including points b and e we can compare the activity in the Gulf to that of the East Coast. Because many insurance contracts only cover the state of Florida or the U.S. part of the coast, the points c, d, f, and g were introduced.

First, we may consider the observed average numbers of TC and hurricane landfalls along the different segments for the period 1900–2014. The red crosses in Fig. 6 show these averages and the red 95% confidence intervals the associated uncertainty in the historical averages estimated by resampling the observed annual number of landfalls for each segment under the hypothesis discussed in the end of section 2a. The uncertainty is the largest in absolute terms for North American TC landfall (about ± 1 TC), however, on a relative basis, the uncertainty is larger for the shorter segments.

Second, we may compare the forecasts to observation. In the case of TCs, the forecasts (blue bars in the figure) are, with a few exceptions, near or within the confidence intervals. The main exceptions are for North American landfall and landfall south of the

United States where the forecasts are markedly higher than the observations. One main driver behind this can be the formation of a number of unphysical low pressure systems just north of South America early in the season in the System 4 forecast model as discussed by Manganello et al. (2016). It may also be due to underreporting of weak TC landfall outside the United States during the early parts of the previous century, suppressing the 1900–2014 HURDAT2 average. While we believe the landfall record for the United States to be reliable from 1900 and onward, we are less certain about the quality of the record south of the United States.

In the case of hurricanes, one may note that the forecasted numbers of landfalls in Canada (segment g–h) and along the East Coast (segment e–g) are higher than observed. This is likely due to model misclassification of TC strength systems as hurricanes, and it may also be due to misclassification of a number of nontropical storm systems moving off the U.S. mainland and curving back to make landfall thereby inflating the landfall count in this region [see System 4 model climatology in Bergman et al. (2017), their appendix B].

Moving on to the skill and error associated with the forecasts when scored against observation during the reforecast period 1981–2014, one may first note from Table 5 that, as expected, the RMSEs are larger than the systematic errors and, relatively speaking, tend to be higher, the shorter the segment and the lower the average number landfalls. In terms of rank correlations, we have noted in section 3b that the skill is higher for

TABLE 5. Mean forecast and observation, their difference (systematic error), the bias corrected RMSE, and the Spearman rank correlation ρ for different segments of the North American coastline. The start and end-points of the segments are shown in Fig. 2. The p value is two sided.

						Correlation	
	Segment	Forecasted	Observed	Difference	RMSE	ρ	p value
TC							
a–h	North America, $10^{\circ}\text{N} < \text{lat} < 52^{\circ}\text{N}$	8.7	7.6	1.1	4.1	0.53	0.001
c–g	United States	4.0	4.0	0.0	3.0	0.27	0.13
a–c	Lat 10°N up to United States	3.0	2.0	1.1	1.9	0.48	0.004
b–e	Gulf to Florida tip	3.6	3.7	−0.1	2.7	0.45	0.01
c–e	U.S. part of Gulf to Florida tip	1.9	2.4	−0.4	2.0	0.24	0.17
e–g	Florida tip to Canada	2.1	1.6	0.4	1.6	0.11	0.53
g–h	Canada to lat 52°N	1.6	1.6	0.0	1.4	0.31	0.07
d–f	Florida	1.3	1.4	−0.1	1.4	0.29	0.10
c–d	U.S. part of Gulf except Florida	1.2	1.4	−0.2	1.6	−0.18	0.30
f–g	U.S. East Coast north of Florida	1.5	1.2	0.3	1.2	0.21	0.23
Hurricane							
a–h	North America, $10^{\circ}\text{N} < \text{lat} < 52^{\circ}\text{N}$	4.2	3.1	1.1	2.6	0.29	0.10
c–g	United States	2.2	1.7	0.5	2.1	0.10	0.56
a–c	Lat 10°N up to United States	0.6	0.9	−0.3	1.2	0.36	0.03
b–e	Gulf to Florida tip	1.4	1.6	−0.2	1.7	0.34	0.05
c–e	U.S. part of Gulf to Florida tip	0.9	0.9	0.0	1.2	0.15	0.38
e–g	Florida tip to Canada	1.3	0.8	0.6	1.2	0.04	0.82
g–h	Canada to lat 52°N	1.4	0.5	0.8	0.7	0.36	0.04
d–f	Florida	0.6	0.4	0.2	0.9	0.14	0.44
c–d	U.S. part of Gulf except Florida	0.5	0.6	−0.1	0.9	0.17	0.34
f–g	U.S. East coast north of Florida	1.1	0.6	0.4	0.9	0.10	0.57

the full North American coast (segment a–h) than for the U.S. part alone (segment c–g). Interestingly, for the shorter segments the highest skill is found for the two of the southernmost segments of coast—the coast south of the United States (segment a–c) and for the Gulf (segment b–e). No significant skill can be found for the U.S. part of the East Coast (segment e–g). Part of the explanation for this may lie in a stronger influence of ENSO in the former regions (see section 3e below). Also, one needs to take a conservative view of the significance levels because multiple hypotheses are being tested.

In conclusion, the average forecasts for the different segments often fall outside the confidence intervals estimated from observations, indicating that the model's distribution of landfall locations differs significantly from the historical record. With two notable exceptions—the coast south of the United States (segment a–c) and the Gulf (segment b–e)—the correlations are generally speaking not significant at the 95% level, including for segment e–g that corresponds to the entire U.S. East Coast.

e. ENSO link to landfall

Gray (1984) showed that El Niño conditions reduce TC activity in the North Atlantic, suggesting a strengthening of the upper-level westerlies over the near-equatorial

parts of the basin as the main mechanism. More recently Kossin et al. (2010), Murakami et al. (2016a), and Boudreault et al. (2017) have investigated the impact of ENSO, the AMO, and other factors on TC activity in different parts of the North Atlantic and the Gulf of Mexico using clustering techniques. Murakami et al. (2016a) found the strongest rank correlation, -0.5 , between the ENSO and the activity in the Gulf. Bove et al. (1998) showed that the ENSO phase impacts the U.S. landfall rates and Landsea et al. (1999), Smith et al. (2007), and Klotzbach (2011) showed that the ENSO phase also modulates TC landfall rates regionally, including in the Gulf, Florida, and along the East Coast.

In line with Murakami et al. (2016a), we have computed the rank correlations between the observed Niño-3.4 SST index (averaged over July–November) and the observed number TC landfalls along different segments (see Table 6). The correlations are negative and significant (p values less than 0.02) for the full North American coastline (segment a–h), the coast south of the United States (segment a–c), and the Gulf (segment b–e). The correlation is not significant for the East Coast (segment e–g). This is consistent with a stronger link between the ENSO phase and the number of landfalls in the tropical regions. Also, the correlation between Florida landfall and ENSO is not significant and cannot

TABLE 6. Spearman rank correlations ρ between the Niño-3.4 index (July–November mean) and the TC landfall number along different segments. Correlations between the observed index and the observed number of landfalls as well as System 4 internal correlations between the forecasted index and the forecasted number of landfalls are shown. System 4 internal correlations are based on a larger sample size ($N = 34 \times 51$) than correlations involving observations ($N = 34$). The start and end-points of the segments are shown in Fig. 2. Boldface font denotes significance above the 99% level. The remaining correlations have p values of 0.2 or worse.

Landfall segment		ρ	
Points	Name	Observed	System 4
a–h	North America, $10^\circ\text{N} < \text{lat} < 52^\circ\text{N}$	−0.46	−0.28
c–g	United States	−0.24	−0.19
a–c	Lat 10°N up to United States	−0.53	−0.26
b–e	Gulf to Florida tip	−0.43	−0.22
e–g	Florida tip to Canada	−0.15	−0.13
d–f	Florida	−0.22	−0.11

be used to support or refute the link found by Smith et al. (2007).

System 4 predicts the Niño-3.4 SST with an anomaly correlation of 0.8 up to 6 months ahead starting from 1 May. Furthermore, as shown in Table 6, the System 4 Niño-3.4 index forecasts and its landfall forecasts are negatively correlated, which is in line with the corresponding observed correlations. Also, the correlations between the System 4 forecasts and the observed Niño-3.4 index range between -0.7 and -0.5 basinwide, as well as for the landfall variables, and are significant at the 99% level or above, suggesting that part of the landfall forecast skill stems from System 4's ability to forecast ENSO phases correctly.

f. The impact of lead time on forecast skill

To gain insight into how lead time impacts forecast skill, forecasts with different start dates (1 May and 1 August) were compared with each other. The forecasts were made for the peak of the hurricane season only, that is, for the period of August, September, and October (ASO). ASO is the most active part of the Atlantic hurricane season accounting for about 85% of the total activity (Landsea 2019).

Table 7 shows the impact of shortening the lead time on the forecast skill and error as measured by the rank correlation and the RMSE. As can be seen, the rank correlation generally increases as the lead time is shortened, indicating an increase in skill. The largest increases are for hurricanes and, in particular, for hurricane landfall over the United States, which increases to about 0.3 (with the two-sided p value of being nonzero dropping just below 0.1). The RMSE in the forecasts also improves by 10%–15% for TC

TABLE 7. The impact of lead time on forecast skill and error. The table shows the Spearman rank correlation and the RMSE in the forecast relative to observation. Forecast start dates are 1 May and 1 Aug, respectively. The period forecasted is August, September, and October. Boldface font denotes significance above 95%. The one-sided p values for U.S. TC and hurricane landfall are 0.08 and 0.32, respectively.

	Spearman rank correlation		RMSE	
	May	Aug	May	Aug
TC				
Number	0.62	0.66	3.1	2.8
Days	0.61	0.72	22	19
North American landfalls	0.58	0.59	3.2	3.1
U.S. landfalls	0.25	0.31	2.8	2.7
Hurricane				
Number	0.62	0.77	2.3	1.9
Days	0.60	0.75	12	10
North American landfalls	0.35	0.50	2.1	1.9
U.S. landfalls	0.08	0.30	1.7	1.7

and hurricane number and days. The improvement is smaller for landfall variables (about 5% for TC landfall and 10% for North American hurricane landfall, whereas no improvement is observed in the case of U.S. hurricane landfall).

The highest skill is obtained for the basinwide forecasts, where we reach correlation levels close to 0.7 for the August forecasts. For the hurricane number forecast we even reach a correlation near 0.8. Dividing the seasons into low, normal, and high as in section 3c shows that the August hurricane number forecast is correct in 19 cases, a near miss in 14, and a total miss in 1 case. Table 8 shows the performance of the TC number and North American landfall forecasts for both lead times. In these two cases one can see that the number of correct forecasts increases as the lead time is shortened, especially in the case of North American landfall. The corresponding skill scores S increase from 0.29 to 0.34 for TC number and from 0.21 to 0.47 for TC North American landfall.

In conclusion, the forecast skill improves when the forecast period is limited to ASO and the lead time is shortened by three months, especially for the basinwide forecasts (e.g., hurricane number) where we reach correlation levels close to 0.7. Also the RMSEs are reduced by 10%–15% for these forecasts. The improvements are smaller for landfall.

4. Discussion and conclusions

We have developed a method, LF+, to calculate landfall risk using the ensembles of storm tracks generated

TABLE 8. Forecasts vs observation of low, normal, and high seasons at two different lead times (1 May and 1 Aug) for TC number and North American TC landfall.

		May forecast			Aug forecast		
		Low	Normal	High	Low	Normal	High
		Number					
Observed	Low	8	2	1	7	3	1
	Normal	2	5	5	4	5	3
	High	1	5	5	0	4	7
		North American landfall					
Observed	Low	6	3	2	7	3	1
	Normal	5	4	3	3	7	2
	High	0	5	6	1	2	8

by System 4 (see [section 2](#)). Briefly, a global land mask of suitable resolution was constructed and algorithms to identify landfalling cyclones were developed. Using observed and System 4 cyclone tracks, a mapping between observed and modeled maximum wind speeds was created, enabling classification of System 4 model cyclones as either TCs or hurricanes. To calculate the values of the forecast variables, a simple method to rescale the System 4 ensemble averages to match more robust, historically observed averages was developed. This method is similar to that used by ECMWF in operational forecasts.

The LF+ forecast method has been applied to the System 4 tracks, to analyze and retrospectively forecast landfall risk along different segments of the North American coast ([section 3](#)). First, one may note that the forecast skill of LF+ is, generally speaking, higher for TC than hurricane landfall, with part of the explanation being the larger number TC landfalls and hence the more robust nature of the TC statistics. Second, one may note that the forecast skill measured by the rank correlation ρ varies between different segments of the coast. The highest correlations ($\rho = 0.5$) were found for landfall along three segments: (i) the entire North American coast, (ii) the coast south of the United States, and (iii) the Gulf coast. Interestingly, the correlation for the relatively long U.S. part of the coast is only 0.3; and, if we isolate the East Coast of the United States, the correlation is not even significantly different from zero. The contrast between the forecast skill for Gulf landfall and for East Coast landfall is striking. One possible explanation for the lack of skill in forecasting East Coast landfall is the too short average TC lifetime and a too quick steering toward the north of model cyclones forming in the eastern parts of the main development region.

When reducing the forecast period from the full hurricane season (May–November) to the peak of the season (ASO), the correlation increases to 0.6 for TC landfall

anywhere along the North American coast, whereas it remains low, that is close to 0.3, for landfall along the U.S. part of the coast (cf. [Tables 3](#) and [7](#)). When we, in addition, shorten the lead time by three months using the forecasts issued on 1 August instead of those issued on 1 May, the TC landfall skill increases only marginally. Turning to the basinwide variables (number and days), the impact of reducing the forecast period to ASO is small and instead the impact of shortening the lead time is considerable—the correlations increase to 0.7 for TCs and to 0.8 for hurricanes. These correlations are high in comparison to those that we and others obtain for the full season using forecasts issued on 1 May ([Camp et al. 2015](#); [Manganello et al. 2016](#); [Murakami et al. 2016a,b](#)). Moreover, the forecast RMSEs drop by between 10% and 20% when shortening the lead time. Considering that about 75% of the destructive hurricane landfall activity occurs during ASO and that the forecast skill is higher for this period, for basinwide as well as some landfall variables, it may be worth issuing forecasts specifically for ASO.

To assess the expected quality of the forecasts going forward, we have analyzed the systematic error and the RMSE. First, one may note that the RMSE is significantly larger than the systematic error for almost all our forecasts. The magnitude of the RMSE of the forecasts (especially the landfall forecasts) is substantial, and it becomes important to consider the causes of the error and to what extent we may reduce them. For example, the RMSE in the forecast of TC landfall anywhere along the North American coastline is 4.1 whereas the mean number of landfalls is 8.7 (the systematic error is 1.1). The two main causes of the forecast error are 1) model error and 2) the inherently chaotic nature of the climate and weather system. The latter will ultimately set a limit to the skill attainable by any model. We discuss this further and attempt to quantify the unpredictable part of the variability in [Bergman et al. \(2017\)](#).

To place our results in context, they were compared with those obtained from two recently developed models, GloSea5 of the Met Office ([Camp et al. 2015](#)) and the high-resolution GFDL FLOR Coupled Model (GFDL model in the following) of the Geophysical Fluid Dynamics Laboratory ([Murakami et al. 2016a](#)). Both models are dynamical fully coupled atmosphere–ocean global seasonal forecast systems like System 4, with comparable spatiotemporal resolutions. The models use different trackers and definitions of tropical cyclones and landfall. Moreover, the reforecast periods differ: 1992–2013 for GloSea5 and 1980–2014 for the GFDL model.

TABLE 9. Comparison of the forecast skill of S4LF+, GloSea5, and GFDL. S4LF+ and GFDL have the same forecast date, 1 May, whereas GloSea5 initiates the forecasts using observational data until 8 May. The GFDL model data are from figures of limited resolution in [Murakami et al. \(2016a\)](#).

Correlation								
Forecast parameters				Spearman		Pearson		RMSE
Model	Period	Variable	Season	ρ	p value	r	p value	
S4LF+	1981–2014	TC number	May–Nov	0.57	0.0004	0.53	0.001	4.1
	1992–2013	TC number	May–Nov	0.54	0.01	0.55	0.01	—
	1992–2013	TC number	Aug–Oct	0.58	0.005	0.62	0.002	—
GloSea5	1992–2013	TC number	Jun–Nov	—	—	0.59	0.002	—
GFDL	1980–2014	TC number	Jul–Nov	0.6	—	—	—	5.4
S4LF+	1981–2014	U.S. landfall	May–Nov	0.29	0.09	0.09	0.61	3.0
	1992–2013	U.S. landfall	May–Nov	0.21	0.4	0.04	0.9	—
	1981–2014	North American landfall	May–Nov	0.53	0.001	0.45	0.008	4.1
	1981–2014	North American landfall	Aug–Oct	0.58	0.0004	0.53	0.001	3.2
	1992–2013	North American landfall	Aug–Oct	0.66	0.0007	0.61	0.003	—
GloSea5	1992–2013	U.S. landfall	Jun–Nov	—	—	0.22	0.164	—
	1992–2013	Caribbean landfall	Jun–Nov	—	—	0.69	0.000	—
GFDL	1980–2014	U.S. landfall	July–Nov	0.25	—	—	—	2.2

The main conclusions drawn by [Camp et al. \(2015\)](#) relevant for comparison with our results are that GloSea5 has skill in forecasting tropical TC number and activity (measured as ACE) for the Atlantic basin. It also has skill in forecasting landfall in the Caribbean, but no skill in forecasting U.S. landfall. Their estimates of the linear ρ correlations between the forecasted and observed variables are shown in [Table 9](#). For ease of comparison we have calculated comparable quantities when possible using S4LF+. When comparing GloSea5 with S4LF+, we find similar skill levels for TC number and days (the latter variable is not shown in the table). Neither model shows significant skill in forecasting U.S. landfall when using the linear correlation, however S4LF+ shows a weak 0.29 Spearman rank correlation for U.S. landfall for the period 1981–2014 (p value of 0.09). The corresponding metric is not available for GloSea5. The S4LF+ correlation for landfall anywhere along the North American coast (north of 10°N and south of 52°N, excluding all islands except the Canadian islands) is higher than the GloSea5 correlation for the U.S. coastline alone, but lower than the GloSea5 correlation for the Caribbean. One may note that GloSea5 includes Nova Scotia in their definition of the United States and the South and Central American coast up to the United States in their definition of the Caribbean.

When we compare with the results obtained by [Murakami et al. \(2016a\)](#) using their dynamical model and a 2-month lead time so that their forecasts are initiated at the same date as ours, we note that they obtain a Spearman correlation of ~ 0.25 (we obtain 0.29) and a RMSE of ~ 2.2 (we obtain 3.0) for U.S. TC landfall. Similarly, the Spearman correlations they obtain for TC number are in line with ours (see

[Table 9](#)). One may note that [Murakami et al. \(2016a\)](#) obtain a higher correlation, ~ 0.4 , and a lower RMSE, ~ 2.0 , for U.S. TC landfall, when using a hybrid dynamical-statistical model. [Murakami et al. \(2016b\)](#) have also investigated the possibility to use their Hi-FLOR dynamical model to forecast landfall (horizontal resolution of 25 km, reforecast period 1980–2015). Using a buffer zone of 300 km they forecast landfall in the United States and Caribbean, and report Spearman correlations ranging from 0.3 to 0.6 for lead times between 0 and 3 months.

Leading forecasting organizations such as ECMWF, NOAA, and the Met Office only disseminate forecasts of basinwide Atlantic cyclone activity, with one reason being the difficulty in forecasting landfall on a seasonal time scale with sufficient skill. However, as shown in this paper, for the North American coast and some of its segments, the skill in forecasting TC landfall is now between 0.4 and 0.5 compared with 0.6 for basinwide Atlantic activity. Moreover, it is reasonable to assume that further refinement of S4LF+, or development of other forecast methods, will tilt the balance more in favor of issuing direct landfall forecasts. One area where landfall risk forecasts may provide economic value in the future is for insurance companies which, year after year, have to decide on the amount of reinsurance to be bought for the upcoming hurricane season. Such statistically minded long-term organizations may adjust their annual reinsurance purchase based on seasonal forecasts, provided that the forecasts are associated with sufficient skill ([Emanuel et al. 2012](#)).

In summary, our main conclusion is that System 4 with LF+ attached can be used to issue seasonal forecasts of landfall risk for the entire North American

coast, the Gulf coast, and the coast south of the United States. However, the forecasts are subject to substantial uncertainty, in part due to model error, and in part due to the chaotic dynamics embedded in the climate system. The quality of the forecasts is in line with that obtained by others.

Acknowledgments. The authors are grateful to Prof. Erland Källén for suggesting the topic, to Prof. Lars Bergman and Dr. Suzana Camargo for providing helpful perspectives, and to Mr. David Richardson for sharp and constructive comments. We are also grateful for the extensive and helpful suggestions provided by the reviewers. The method and most of the results have previously been reported in a Master's thesis (Bergman 2016) and an ECMWF techmemo (Bergman et al. 2017) and are the results of a collaboration between the European Centre for Medium-Range Weather Forecast, the Department of Meteorology at Stockholm University, and the Third Swedish National Pension Fund.

REFERENCES

- Barnston, A. G., M. K. Tippett, M. L. L'Heureux, S. Li, and D. G. DeWitt, 2012: Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bull. Amer. Meteor. Soc.*, **93**, 631–651, <https://doi.org/10.1175/BAMS-D-11-00111.1>.
- Bergman, D. L., 2016: Seasonal forecasting of storm and hurricane landfall using ECMWF's System 4. M.S. thesis, Department of Meteorology, Stockholm University, 72 pp.
- , L. Magnusson, J. Nilsson, and F. Vitart, 2017: Forecasting tropical cyclone landfall using ECMWF's seasonal forecasts from system 4. Tech. Memo. 811, European Centre for Medium-Range Weather Forecasts, 48 pp., <https://doi.org/10.21957/lg4nj4d1o>.
- Blake, E. S., C. W. Landsea, and E. J. Gibney, 2011: The deadliest, costliest, and most intense United States tropical cyclones from 1851 to 2010 (and other frequently requested hurricane facts). NOAA Tech. Memo. NWS NHC-6, NOAA/National Weather Service, National Centers for Environmental Prediction, National Hurricane Center, 49 pp., <https://www.nhc.noaa.gov/pdf/nws-nhc-6.pdf>.
- Boudreault, M., L.-P. Caron, and S. J. Camargo, 2017: Reanalysis of climate influences on Atlantic tropical cyclone activity using cluster analysis. *J. Geophys. Res. Atmos.*, **122**, 4258–4280, <https://doi.org/10.1002/2016JD026103>.
- Bove, M. C., J. J. O'Brien, J. B. Eisner, C. W. Landsea, and X. Niu, 1998: Effect of El Niño on U.S. landfalling hurricanes, revisited. *Bull. Amer. Meteor. Soc.*, **79**, 2477–2482, [https://doi.org/10.1175/1520-0477\(1998\)079<2477:EOENOO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1998)079<2477:EOENOO>2.0.CO;2).
- Camargo, S. J., A. H. Sobel, A. G. Barnston, and P. J. Klotzbach, 2010: The influence of natural climate variability on tropical cyclones, and seasonal forecasts of tropical cyclone activity. *Global Perspectives on Tropical Cyclones: From Science to Mitigation*, J. C. L. Chan and J. D. Kepert, Eds., Vol. 4, World Scientific, 325–360, https://doi.org/10.1142/9789814293488_0011.
- Camp, J., and L.-P. Caron, 2017: Analysis of Atlantic tropical cyclone landfall forecasts in coupled GCMs on seasonal and decadal timescales. *Hurricanes and Climate Change*, J. Collins and K. Walsh, Eds., Vol. 3, Springer, 213–241.
- , M. Roberts, C. MacLachlan, E. Wallace, L. Hermanson, A. Brookshaw, A. Arribas, and A. Scaife, 2015: Seasonal forecasting of tropical storms using the Met Office GloSea5 seasonal forecast system. *Quart. J. Roy. Meteor. Soc.*, **141**, 2206–2219, <https://doi.org/10.1002/qj.2516>.
- Duan, W., and C. Wei, 2013: The “spring predictability barrier” for ENSO predictions and its possible mechanism: Results from a fully coupled model. *Int. J. Climatol.*, **33**, 1280–1292, <https://doi.org/10.1002/joc.3513>.
- Emanuel, K., 2005: *Divine Wind: The History and Science of Hurricanes*. Oxford University Press, 285 pp.
- , F. Fondriest, and J. Kossin, 2012: Potential economic value of seasonal hurricane forecasts. *Wea. Climate Soc.*, **4**, 110–117, <https://doi.org/10.1175/WCAS-D-11-00017.1>.
- Gray, W. M., 1984: Atlantic seasonal hurricane frequency. Part I: El Niño and 30 mb quasi-biennial oscillation influences. *Mon. Wea. Rev.*, **112**, 1649–1668, [https://doi.org/10.1175/1520-0493\(1984\)112<1649:ASHFPI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1984)112<1649:ASHFPI>2.0.CO;2).
- Klotzbach, P. J., 2011: El Niño–Southern Oscillation's impact on Atlantic basin hurricanes and U.S. landfalls. *J. Climate*, **24**, 1252–1263, <https://doi.org/10.1175/2010JCLI3799.1>.
- , and W. M. Gray, 2009: Twenty-five years of Atlantic basin seasonal hurricane forecasts (1984–2008). *Geophys. Res. Lett.*, **36**, L09711, <https://doi.org/10.1029/2009GL037580>.
- Kossin, J. P., S. J. Camargo, and M. Sitkowski, 2010: Climate modulation of North Atlantic hurricane tracks. *J. Climate*, **23**, 3057–3076, <https://doi.org/10.1175/2010JCLI3497.1>.
- Landsea, C. W., 2019: How many hurricanes have there been in each month? NOAA, accessed 26 June 2019, <https://www.aoml.noaa.gov/hrd/tcfaq/E17.html>.
- , and J. L. Franklin, 2013: Atlantic hurricane database uncertainty and presentation of a new database format. *Mon. Wea. Rev.*, **141**, 3576–3592, <https://doi.org/10.1175/MWR-D-12-00254.1>.
- , R. A. Pielke, A. M. Mestas-Nunez, and J. A. Knaff, 1999: Atlantic basin hurricanes: Indices of climatic changes. *Climatic Change*, **42**, 89–129, <https://doi.org/10.1023/A:1005416332322>.
- , G. A. Vecchi, L. Bengtsson, and T. R. Knutson, 2010: Impact of duration thresholds on Atlantic tropical cyclone counts. *J. Climate*, **23**, 2508–2519, <https://doi.org/10.1175/2009JCLI3034.1>.
- MacLachlan, C., and Coauthors, 2015: Global seasonal forecast system version 5 (GloSea5): A high-resolution seasonal forecast system. *Quart. J. Roy. Meteor. Soc.*, **141**, 1072–1084, <https://doi.org/10.1002/qj.2396>.
- Manganello, J. V., and Coauthors, 2016: Seasonal forecasts of tropical cyclone activity in a high-atmospheric-resolution coupled prediction system. *J. Climate*, **29**, 1179–1200, <https://doi.org/10.1175/JCLI-D-15-0531.1>.
- Molteni, F., and Coauthors, 2011: The new ECMWF seasonal forecast system (System 4). Tech. Memo. 656, Internal Rep., European Centre for Medium-Range Weather Forecasts, 51 pp., <https://www.ecmwf.int/sites/default/files/elibrary/2011/11209-new-ecmwf-seasonal-forecast-system-system-4.pdf>.
- Murakami, H., G. Villarini, G. A. Vecchi, W. Zhang, and R. Gudgel, 2016a: Statistical–dynamical seasonal forecast of North Atlantic and U.S. landfalling tropical cyclones using the

- high-resolution GFDL FLOR coupled model. *Mon. Wea. Rev.*, **144**, 2101–2123, <https://doi.org/10.1175/MWR-D-15-0308.1>.
- , and Coauthors, 2016b: Seasonal forecasts of major hurricanes and landfalling tropical cyclones using a high-resolution GFDL coupled climate model. *J. Climate*, **29**, 7977–7989, <https://doi.org/10.1175/JCLI-D-16-0233.1>.
- Nicholls, N., 1979: A possible method for predicting seasonal tropical cyclone activity in the Australian region. *Mon. Wea. Rev.*, **107**, 1221–1224, [https://doi.org/10.1175/1520-0493\(1979\)107<1221:APMFPS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1979)107<1221:APMFPS>2.0.CO;2).
- Panofsky, H. A., and G. W. Brier, 1958: Some applications of statistics to meteorology. Mineral Industries Extension Services, College of Mineral Industries, Pennsylvania State University, 224 pp.
- Pielke, R. A., Jr., J. Gratz, C. W. Landsea, D. Collins, M. A. Saunders, and R. Musulin, 2008: Normalized hurricane damage in the United States: 1900–2005. *Nat. Hazards Rev.*, **9**, 29–42, [https://doi.org/10.1061/\(ASCE\)1527-6988\(2008\)9:1\(29\)](https://doi.org/10.1061/(ASCE)1527-6988(2008)9:1(29)).
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: *Numerical Recipes in C: The Art of Scientific Computing*. 2nd ed. Cambridge University Press, 994 pp.
- Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, <https://doi.org/10.1175/JCLI-D-12-00823.1>.
- Smith, S. R., J. Brolley, J. J. O'Brien, and C. A. Tartaglione, 2007: ENSO's impact on regional U.S. hurricane activity. *J. Climate*, **20**, 1404–1414, <https://doi.org/10.1175/JCLI4063.1>.
- Van der Grijn, G., J.-E. Paulsen, F. Lalaurette, and M. Leutbecher, 2005: Early medium-range forecasts of tropical cyclones. *ECMWF Newsletter*, No. 102, ECMWF, Reading, United Kingdom, 7–14.
- Vecchi, G. A., and T. R. Knutson, 2011: Estimating annual numbers of Atlantic hurricanes missing from the HURDAT database (1878–1965) using ship track density. *J. Climate*, **24**, 1736–1746, <https://doi.org/10.1175/2010JCLI3810.1>.
- , and Coauthors, 2014: On the seasonal forecasting of regional tropical cyclone activity. *J. Climate*, **27**, 7994–8016, <https://doi.org/10.1175/JCLI-D-14-00158.1>.
- Villarini, G., G. A. Vecchi, and J. A. Smith, 2012: U.S. landfalling and North Atlantic hurricanes: Statistical modeling of their frequencies and ratios. *Mon. Wea. Rev.*, **140**, 44–65, <https://doi.org/10.1175/MWR-D-11-00063.1>.
- Vitart, F., and T. N. Stockdale, 2001: Seasonal forecasting of tropical storms using coupled GCM integrations. *Mon. Wea. Rev.*, **129**, 2521–2537, [https://doi.org/10.1175/1520-0493\(2001\)129<2521:SFOTSU>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2521:SFOTSU>2.0.CO;2).
- , J. Anderson, and W. Stern, 1997: Simulation of interannual variability of tropical storm frequency in an ensemble of GCM integrations. *J. Climate*, **10**, 745–760, [https://doi.org/10.1175/1520-0442\(1997\)010<0745:SOIVOT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<0745:SOIVOT>2.0.CO;2).
- , F. Prates, A. Bonet, and C. Sahin, 2011: New tropical cyclone products on the web. *ECMWF Newsletter*, No. 130, ECMWF, Reading, United Kingdom, 17–23.