TECHNICAL MEMORANDUM

811

Forecasting tropical cyclone landfall using ECMWF's seasonal forecasts from System 4

Dan L. Bergman[1,2], Linus Magnusson[3], Johan Nilsson[2] and Frederic Vitart[3]

Forecast/Research Department

[1]The Third Swedish National Pension Fund, [2]Meteorological Department Stockholm University and [3]ECMWF

October 2017

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/en/research/publications

Contact: library@ecmwf.int

Abstract

The European Centre for Medium-Range Weather Forecast's seasonal forecast system 4 (System 4) has been used to issue basin-wide tropical cyclone forecasts since 2011. This report describes a method developed to forecast seasonal landfall risk using the ensembles of cyclone tracks generated by System 4. The method has been applied to analyze and retrospectively forecast the landfall risk along different segments of the North American coast, with a focus on the U.S. part.

The main result is that the method can be used to forecast landfall for some parts of the coast, but the method's skill is, generally speaking, lower for landfall than for basin-wide forecasts of activity. The rank correlations between forecast and observation are 0.6 for basin-wide storm number, 0.5 for landfall anywhere along the coast, and 0.3 for landfall along the U.S. part of the coast. When we limit the forecast to the peak of the hurricane season (August, September, and October), the correlation increases to 0.6 for the entire coast whereas it remains close to 0.3 for the U.S. part.

The forecast error is substantial in all cases, in part due to model error, in part due to the chaotic dynamics embedded in the climate system. A crude analysis suggests that the forecast error can be reduced by 10 to 25 percent (depending on the forecast) before reaching the ultimate limit set by the chaotic dynamics. In conclusion, the quality of the forecasts is well in line with that obtained using other state-of-the-art methods, and it is sufficient to be of use for organizations, such as reinsurance companies, that plan and operate with a statistical mind-set on a multi-year horizon.

1 Introduction

Hurricanes, or tropical cyclones[1], are among the most destructive yet beautiful geophysical phenomena on earth (Emanuel, 2005a). They can cause extensive damage and result in catastrophic loss of life. In the U.S. alone more than 10,000 lives have been lost since the beginning of the last century and economic values well in excess of $ 200,000,000,000 have been destroyed (Pielke Jr et al., 2008; Blake et al., 2011). Urbanization and migration towards coastal cities and low lying areas continue at the same time as global warming is causing a rise in ocean temperature, sea levels, and possibly also in destructive hurricane activity. To what extent climate change impacts the number and destructiveness of hurricanes is an open area of research (Emanuel, 2005b; Webster et al., 2005; Landsea et al., 2006; Emanuel et al., 2008; IPCC, 2013). Regardless of the impact of climate change on hurricane activity, an improved understanding of hurricanes and an improved ability to forecast activity, in particular landfalling activity, can save lives and guide mitigation efforts more efficiently.

Seasonal forecasting of Atlantic hurricane activity was pioneered by Gray when, in 1984, he demonstrated the link between, on the one hand, El Niño Southern Oscillation (ENSO), the Quasi-Biennial Oscillation of the equatorial zonal wind (QBO) as well as some other climate variables; and, on the other hand, Atlantic hurricane activity. Both the ENSO and the QBO represent slow, irregular oscillations or modes of the climate system, modes that can persist during significant parts of a hurricane season (Gray, 1984). The persistence and predictability of such slow modes in the climate system form an important basis for seasonal hurricane forecasting today.

Most of these modes involve the ocean owing to its relative inertia as compared to much of the atmospheric dynamics. Important modes with links to Atlantic hurricane activity are the ENSO, the QBO, the Atlantic Multidecadal Oscillation (AMO), the Atlantic Meridional Mode (AMM), the Madden-Julien Oscillation (MJO), and the North Atlantic Oscillation (NAO). These modes exhibit dynamics on different time scales ranging from decades in the case of the AMO to months in the case the MJO, to weeks in

[1] We will use the term hurricane to denote any tropical cyclone with a wind speed above 64 knots and the term tropical storm or just storm to denote any tropical cyclone with a wind speed above 34 knots, regardless of the basin.

the case of the fastest dynamics of the NAO. Some of these modes, such as the NAO, exhibit dynamics on multiple time scales ranging from weeks to several years. These modes are interrelated and interact to different degrees. And, to the extent they are related to hurricane activity, they can be used as variables in *statistical forecast models*. Generally, speaking the understanding of the physical mechanisms linking these modes to hurricane genesis, intensification, and movement is limited. See Camargo et al. (2010) for a discussion on natural climate variability and its link to hurricane activity.

The ENSO may be the mode with the strongest impact on Atlantic hurricane activity – according to the estimates of Bove et al. (1998), the probability of two or more U.S. landfalling hurricanes in one season is 0.66 during a La Niña as compared to 0.28 during an El Niño. The main mechanism by which the ENSO is thought to impact Atlantic hurricane activity is via modulation of the wind shear over the Caribbean and the eastern tropical Atlantic. However, although the ENSO can be forecasted with some skill several months ahead, the skill varies with the phase of the ENSO. Generally speaking, the skill is higher and extends to longer lead times once we are in a well-developed El Niño (the situation is less clear in the case of a developed La Niña phase). Because the transition between phases often occurs during the northern hemisphere spring, we often find ourselves in the challenging situation that the expected forecast skill is low just before the onset of the Atlantic hurricane season, see Barnston et al. (2012); Duan and Wei (2013).

Dynamical forecast models differ from statistical models in that they directly simulate the development of the climate system by integration of the dynamical equations. One advantage of the dynamical forecast models is their ability to simultaneously capture several modes of climate variability and their interaction, another is their relative independence of historically established climatological relations (relations that may change[2]), and a third is the possibility to efficiently construct ensembles of forecasts by perturbation of the initial conditions and the dynamical eqations, thereby allowing analysis of the chaotic nature and variability of the climate system. One drawback of the dynamical forecast models is the errors caused by the inevitable approximation and simplification of the true climate system; we can be certain that the model climate differs from the real climate, the question is just to what extent and how it will impact the quality of the forecasts. In Camargo et al. (2010) the authors discuss some of the known limitations of dynamical forecasts models, one example relevant for us being a tendency for some models to forecast too short tracks that move poleward too quickly. Another example is that the forecasted intensity sometimes is low especially when estimated directly from model wind fields. These limitations are partly related to the resolution of the dynamical model used.

The European Centre for Medium-Range Weather Forecasts (ECMWF), The UK MetOffice, and the National Oceanic and Atmospheric Administration (NOAA) among others produce dynamical seasonal forecasts of basin-wide tropical cyclone activity. However, although at least one commercial organization (Tropical Storm Risk) provides seasonal forecasts of landfall risk, neither the ECMWF, nor the UK Metoffice, nor the NOAA, publish such forecasts, despite the great societal interest in preparing and warning for potential damages ashore. This is linked to the difficulty to forecast landfall with sufficient skill. The continuous development of numerical coupled atmospheric-ocean prediction models by ECMWF and others has now reached a stage where it is worth re-examining the possibility to forecast landfall.

Ideally, we would like to forecast storm and hurricane landfall, intensity, size, forward speed as well as other parameters critical for damage and destruction. However, it is natural (and challenging enough) to start by forecasting the landfall itself: after all, unless we have a landfall or a sufficiently close bypass of the eyewall, there will be no or little damage. Moreover, System 4 has a limited resolution and details of

[2] The relation between the QBO and Atlantic hurricane activity provides one example, see Klotzbach (2007).

the pressure and wind fields as well as the tracks are not resolved with sufficient accuracy to allow for some of the more detailed analyses.

The performance of the UK Met Office GloSea5 model with regards to tropical cyclone activity and landfall forecasts were recently assessed by Camp et al. (2015). GloSea5 is a fully coupled atmosphere-ocean global seasonal forecast system like System 4, with a comparable spatio-temporal resolution. The model uses a slightly different tracker and definition of tropical cyclones and the main time period analyzed is somewhat shorter (1992-2003) than the time period that we use (1981-2014). Their main conclusions relevant for comparison with our results are that GloSea5 has skill in forecasting tropical storm 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.

Manganello et al. (2016) have assessed the seasonal forecast skill for basin-wide and regional tropical cyclone activity in three experimental higher atmospheric resolution versions of System 4 as part of a project labeled Minerva. The resolutions used were T319, T639 and T1279, corresponding to horizontal resolutions of 62, 32 and 16 km, in all cases higher than the resolution of the operational version of System 4 (T255, corresponding to a horizontal 80 km near the equator) that forms the basis for this study. The criteria for identification of tropical storms were similar to those that will be used in this report (section2.4), but the minimum wind speed criterion was increased to adjust for the higher resolutions of the models. For the T1279 resolution forecasts they find some, but still limited, skill close to the U.S. coastline.

The present memo reports the results of an investigation into to what extent ECMWF's seasonal forecast system 4 (System 4) can be used to forecast storm 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 robustness and quality of the forecasts have been analyzed. We will use the term LF+ to denote the method developed.

The memo 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 draw some conclusions regarding the method and its applicability. Finally, in the appendices supplementary data and analyses are provided. The method and the results have previously been reported in a Master's thesis, Bergman (2016), 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.

2 Method and data

In this section, the method developed to forecast landfall is described. The method uses System 4's ensembles of storm tracks and historical data from Hurdat (Landsea et al., 2013) to construct the forecasts. We will use the term LF+ to denote the method here developed. (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 (a) the mapping of System 4 cyclone wind speeds to corresponding real wind speeds (Section 2.4), (b) the determination of track extent and landfall (section 2.5), and (c) the calculation of forecasts of different variables, for example the number of expected landfalls along a segment of the coast (Section 2.6).

Before describing the method, we will, however, begin this section by describing the observations and historical data used as well as relevant aspects of System 4 and the ECMWF tracker.

2.1 Observation and historical record

Around the world, meteorological organizations and institutes monitor, collect, and analyze data on tropical cyclones. In the case of the North Atlantic (including the Gulf of Mexico and the Caribbean) the National Oceanic and Atmospheric Administration (NOAA) has compiled and reanalyzed data on all known tropical cyclones dating back to 1851 (Landsea et al., 2013). We have used their most recent database, Hurdat2, for validation and construction of our forecasts. Hurdat2 is one of the most comprehensive and reliable sources of observational tropical cyclone data for this basin. In the following, the Hurdat2 database will be referred to simply as Hurdat.

Briefly, Hurdat contains records of all known tropical cyclones that have formed over the Atlantic between 1851 and 2015. There are data on the maximum wind, the minimum pressure, and the location for systems ranging in strength from tropical depressions to hurricanes. As a general rule, the data are available every six hours during the life of each system. For natural reasons, the accuracy of the data and the completeness of the record are lower for earlier dates. While the record of hurricanes making landfall in the U.S. is generally regarded as reliable from 1900 and onwards, the completeness and accuracy of the data for systems moving over the Atlantic have improved gradually with two notable steps being taken; first, in the forties, with the advent of reconnaissance flights and then, in the sixties, with the launch of weather satellites.

In addition to Hurdat, we have used ERA-interim (Dee et al., 2011) for assessment and construction of the forecasts. Briefly, ERA-interim is a global data set of meteorological fields and variables for the period 1979 to today. ERA-interim reanalysis has been constructed by using historical observations with the ECMWF data assimilation system. Its horizontal resolution[3] is approximately 80 km (which is the same as that used by System 4) and sixty vertical levels from the surface up to 0.1 hPa are used. Due to its limited resolution, ERA cannot resolve finer details of tropical cyclones, such as the eye wall. ERA-interim will be referred to as ERA in the following.

The number of 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 hurricanes that formed (number), the activity (number of hurricane days), and the hurricane landfalls in the U.S. (U.S. landfall) for each year during the period 1900 through 2014. As can be seen there is a considerable interannual variability in the historical record. Increasing trends in the number and activity of hurricanes can also be observed. These trends are largely due to underreporting or missing data before systematic aircraft reconnaissance in the fifties and the launch of weather satellites in the sixties, see for example Villarini et al. (2012). The landfall data is more reliable, and no statistically significant trend can be found. (In fact, there is a weak, non-significant, negative trend.)

In order to design and assess the forecast method, we need climatological estimates of the mean and standard deviation of the relevant variables (number, activity, 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 activity and 1900-2014 for landfall. Table 1 shows the estimated means and standard deviations.

When deciding on these periods we are leaning on the fact that climate change has not left a clear mark

[3] ERA-interim's horizontal spectral resolution is T255 and grid-point calculations are on a reduced Gaussian N128 grid.

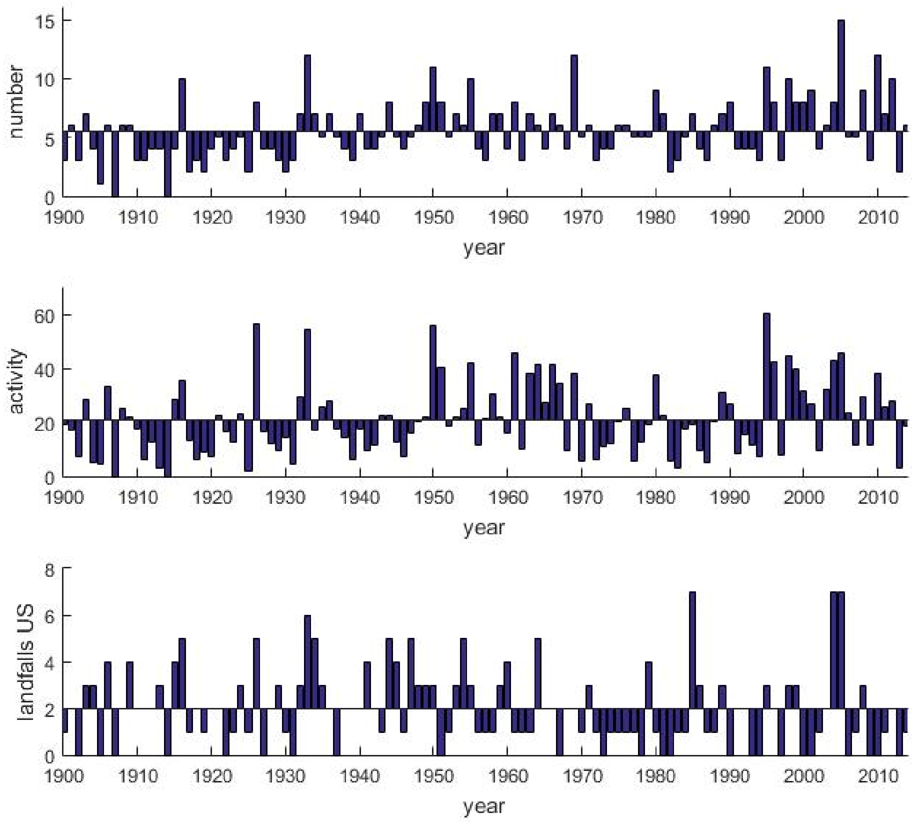

Figure 1: Observed hurricane number, activity, and U.S. landfall. Bars up and down from period means. Data from Hurdat.

in the statistics during these time periods. Also, while the satellites made the record nearly complete post 1970, we regard the basin-wide record since 1950 as sufficiently accurate. Moreover, the better quality of the historical record post 1970 has to be weighed against (a) 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. The choice of the period 1950-2014 for number and, in particular, activity can be discussed; Landsea et al. (2010) argue that underreporting of weak systems in the beginning of the period is a marked drawback.

By taking the view that, for example, the observed storm number for each year during the selected time period represents a random draw from an unknown but constant probability distribution representing the "true climatology", the sample mean represents our statistically best estimate of the "true" mean and the standard deviation our best estimate of the "true" standard deviation. Moreover, we can use bootstrap[4] to estimate confidence intervals for the mean and standard deviation. Table 1 compiles the estimated means and standard deviations for the key storm and hurricane related variables.

Table 1: Mean and standard deviation for storm and hurricane *number*, *activity*, and *U.S. landfall* estimated from Hurdat. *best* denotes the corresponding best estimate, and *lower* and *upper* the bounds of 95-percent, 2-sided, confidence intervals.

	period	mean			standard deviation		
		best	lower	upper	best	lower	upper
STORM							
number	1950-2014	10.8	9.9	12.0	4.2	3.6	5.6
activity	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
activity	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

As can be seen in the table, the best estimate of the expected mean storm number is 10.8 and the corresponding 95-percent confidence interval indicates that the "true" mean should lie somewhere between 9.9 and 12.0. Similarly, for hurricanes the estimate is 6.1 ± 0.7 and for U.S. storm landfall 4.0 ± 0.4.

The standard deviation, which measures the interannual variability, is considerable in relation to the mean for all variables, and in particular for the landfall variables; it is, for example, 39 percent for storm number, 61 percent for U.S. storm landfall, and in the extreme 86 percent for U.S. hurricane landfall.

In summary, analysis of the Atlantic cyclone climatology has enabled us to formulate and use a set of reasonable assumptions to estimate climatological means and standard deviations for validation and

[4] Bootstrap is method whereby a number of alternative data sets are created by resampling of the original set, see Press et al. (1992).

design of the forecast method. (In appendix B, a supplementary comparison of System 4's climatology with the historical climatology as captured by Hurdat and ERA is available.)

2.2 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 has been operational since November 2011 and is the result of extensive research and development over the last three decades at ECMWF. Today, ECMWF is one of the world leaders in extended range and seasonal forecasting of the weather and climate system.

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*[5] and an atmospheric model, *The ECMWF's Integrated Forecast System (IFS) Cycle 36r4*. The two model components have been coupled[6] to capture atmospheric and oceanic interactions which can be important for hurricane forecasts, the 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 1×1 degrees in the mid-latitudes (equatorially refined) and in the vertical 42 levels are used, 18 of which are in the first 200 meters. The spectral resolution of the atmospheric model is T255 in the horizontal, and grid point calculations are done on a reduced Gaussian N128 grid, which corresponds to a resolution of about 0.7 degrees. In the vertical, 91 levels are used with the top level at 0.01 hPa. The time step in the atmospheric model is 45 minutes, the time step in the oceanic 60 minutes, and the models are coupled every three hours. A more thorough description of both models and the coupling can be found in Molteni et al. (2011).

The unperturbed atmospheric initial conditions for the period 1981 to 2010 have been calculated from ERA and for the period 2011 to 2014 data from ECMWF's operational analysis have been used. Similarly, the oceanic initial conditions come from ORA-S4 for the period 1981-2010 and from the operational NEMOVAR runs for the period 2011-2014. In order to generate the ensemble of forecasts, the initial state was perturbed and stochastic parametrization of the dynamical equations was applied (Palmer et al., 2008). Perturbations were applied to the initial ocean and atmospheric states, whereas land conditions were left unperturbed. However, most of the ensemble spread at the seasonal time scale is internally generated by the chaotic nature of the dynamics of the climate system, and the initial perturbations are the most important during the first month of the forecast. See Molteni et al. (2011) for details.

In this study, retrospective forecasts (so-called reforecasts) and forecasts[7] 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 staring on 1 August each spanning seven months ahead have been used.

2.3 The ECMWF tracker

In order to identify and track tropical cyclones from the model fields forecasted by System 4, a so called *objective tracker* has been used. Briefly, the ECMWF tracker at each time step first identifies potential tropical cyclones using the criteria developed by Vitart et al. (1997), then it associates cyclones

[5]NEMO, *Nucleus for European Modeling of the Ocean*, developed in collaboration between the ECMWF and British and French research institutions (Madec, 2008)

[6]A version of the OASIS3 coupler (developed at CERFACS) has been used.

[7]For the last three years, 2011-2014 real-time were used.

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 (a) the presence of a local maximum in the vorticity (b) the presence of a warm core, and (c) a maximum wind speed of the system of at least 13 m/s. Only tracks which meet the above criteria at least twice during the life of the cyclone are kept. The warm core criterion needs only to be fulfilled once during the life of the cyclone, see Vitart et al. (2011) and references therein for further details. The idea is to as accurately as possible track tropical cyclones of at least tropical storm strength (see below). For this study, track data saved every 12 hours was provided from the tracker. The data included the location of the center of the system, the estimated maximum wind speed, and the estimated minimum pressure.

The rotational circulation associated with tropical cyclones typically extends between 100 to 1000 km from the center and can hence be resolved or at least represented with some degree of accuracy by System 4 given that the horizontal resolution of the model is about 0.7 degrees (corresponding to about 80 km in the tropics). It is, however, not possible to resolve for example the eyewall[8]. The limited resolution of the model and its inability to resolve the finer details of the wind and pressure fields, generally results in estimated maximum winds that are lower and minimum pressures that are higher than observed[9]. For this reason, the criterion for inclusion in the track has been set to 13 m/s instead of 17 m/s, the standard definition of tropical storm strength.

Because each track only contains data saved every 12 hours and, because tracking is stopped when the cyclone weakens sufficiently, a number of fast moving rapidly weakening cyclones that should make landfall according to System 4 will have no track point recorded over land. In order to ensure that we capture such cyclones, each track terminating at sea is tentatively extended in the direction given by the last two track points a distance equal to that between the same two points. If that new tentative point is over land it is added to the track, if not it is discarded.

2.4 Intensity calibration

While the 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, as discussed, not sufficient to fully capture the intense winds near the center of a hurricane. This and other System 4 limitations make forecasts of cyclone frequency, landfall, and in particular cyclone intensity challenging. Care must be exercised when using System 4 for this purpose. One key question for this study is how a tracked cyclone's wind speed and pressure may or may not be used to draw conclusions about its intensity, in particular in the case of landfall.

In order to assess the precision of the tracker, we have applied it to ERA for each one of the hurricane seasons from 1989 through 2009 and compared the wind speeds and pressures obtained to those actually observed according to Hurdat at different points along the tracks. From this comparison, it is clear that the wind speed estimated from ERA is, generally speaking, lower than the wind speed according to Hurdat; borderline storm strength in Hurdat corresponds to, on average, 26 knots according to the tracker and ERA. Similarly, boarder line hurricane strength corresponds to 33 knots. Moreover, because our ultimate objective is to relate the intensity of the cyclones generated by System 4 to the intensity of real cyclones as recorded in Hurdat, we face an additional degree of uncertainty arising from System 4 itself.

[8] The radius of maximum winds typically range from 10 to 25 km (Carrasco et al., 2014).

[9] The tracker estimates the maximum wind, the minimum pressure, and the location of the cyclone center by interpolating between grid points.

Considering the considerable uncertainty in the cyclone maximum wind and central pressure as estimated by applying the tracker to System 4, we have used a statistical quantile-quantile (Q-Q) approach[10] to estimate the corresponding "true" intensity. The result is that, the cyclones with a System 4 wind speed of 24 knots or above will be regarded as storms and cyclones with a wind speed with 31 knots or above will be regarded as hurricanes. This compares reasonably with the ERA-Hurdat wind speed relation, however, one should bear in mind that, these thresholds are only reasonable approximations.

2.5 Landfall determination

In order to determine landfall, we have used a global land mask with a 0.05×0.05 degrees resolution (see Figure 2), which corresponds to 5.6 km meridional minimum resolution near the equator. The resolution of the mask is sufficient given the resolution of System 4 and ERA (about 0.7 degrees, that is 80 km near the equator), and the resolution in the Hurdat position data (0.1 degrees resolution). The land mask has been constructed from the Group for High Resolution Sea Surface Temperature's (GHRSST) high resolution land mask which has a resolution of 0.0083×0.0083 degrees near the equator [11]. For each point within 500 km of land we have calculated the shortest distance to land, to detect cyclones that pass land close 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 degrees north or north of 52 degrees north.

To make the landfall statistics as robust and as comparable as possible, the same land mask and criteria for landfall were applied to the forecasts and the historically observed tracks. 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, not sea, as even if the center of the hurricane 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).

For each track, the following method was applied to find any landfall or seafall points[12]:

1. find all landfall and seafall points along the track,
2. test up to 240 tentative linearly interpolated points between each pair of points on the track. Identify any new landfall or seafall points. Add any such points to the track.
3. remove all landfall points for systems that have not been at least 24 hours over water before landfall. This is only applied to the System 4 and ERA tracks, not the observed Hurdat tracks because their wind speed and position data are more reliable.
4. remove all landfall points closer than 300 km to a previous landfall point.
5. depending on the wind speed at landfall, classify the landfall as a tropical storm landfall or a hurricane landfall (see Section 2.4).

Figure 2 shows examples of tracks identified using this algorithm and the land mask described in the previous section. Genesis points are marked by yellow dots, termination points by red dots, and landfall points by circles.

[10] This approach is briefly described in appendix A.
[11] see GHRSST for more information https://www.ghrsst.org/products-and-services/tools/navo-ghrsst-pp-land-sea-mask/
[12] Landfall occurs whenever a point over sea is followed by a point over land and vice versa for seafall.

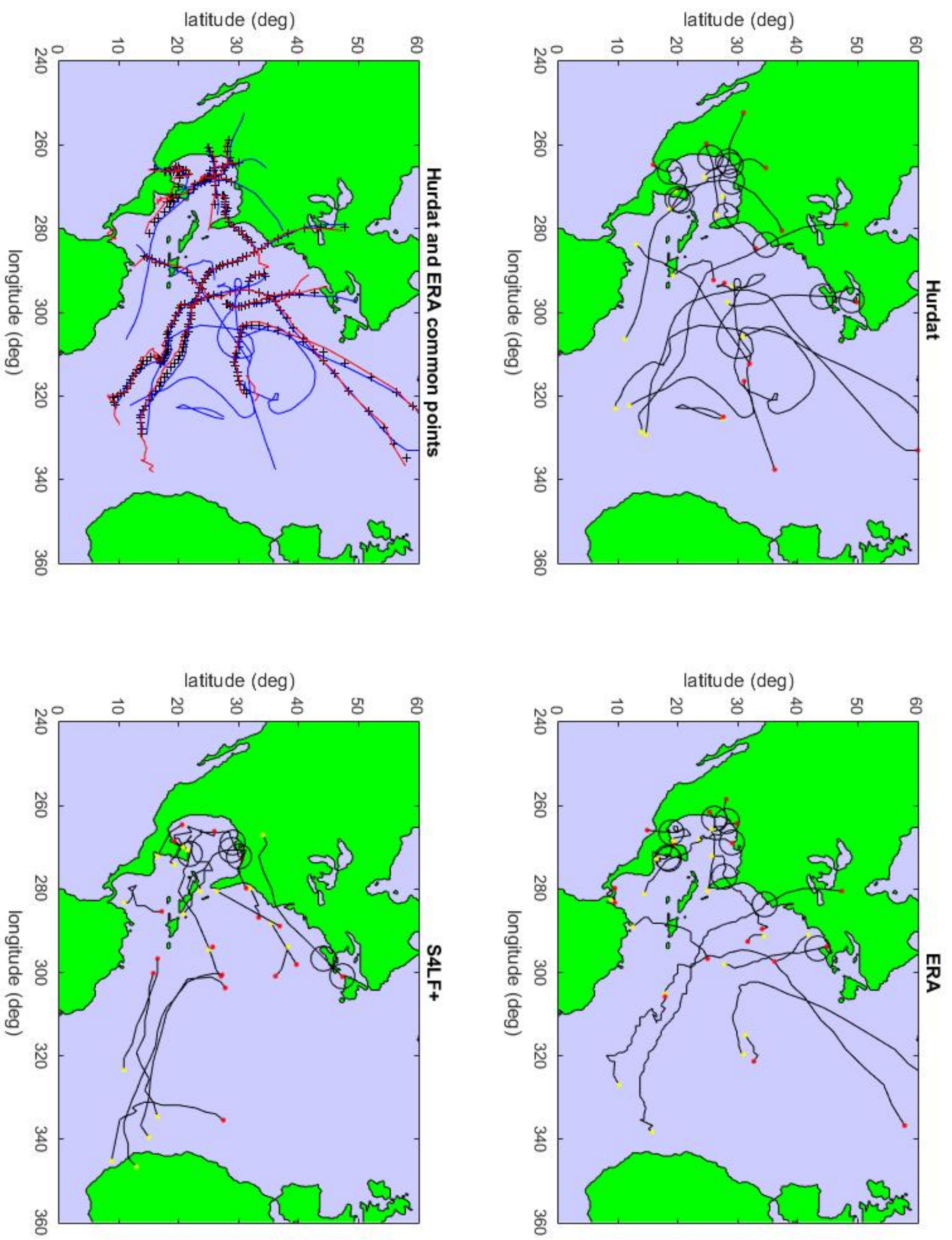

Figure 2: Land mask and Tracks. The Hurdat, ERA, and S4LF+ tracks for the hurricane season 2003. Genesis points are marked by yellow dots, termination points by red dots, and landfall points by circles. In the lower left figure, Hurdat tracks are blue and ERA tracks are red, and common (overlapping) points are marked by +. (For S4LF+ a selection of typical tracks are shown.)

2.6 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 storm and hurricane *number*, *activity*, and *landfall*. Landfall will be predicted for different segments of the North American coast, the main segments being the full coast *N.A. landfall* and the U.S. part of the coast *U.S. landfall*, but other shorter segments will also be considered (for definitions of segments, see Section 3.3). By storm number we simply mean the number of storm strength systems that form during a given year. Similarly, storm activity refers to the number of storm days (If we have two storms in the water at the same time we will count both.). And, storm landfall refers to the number of storm landfalls (thus the same storm can make multiple landfalls, compare Section 2.5). The variables for hurricane strength systems are named analogously. We have chosen the number of storm days to estimate the total basin-wide activity of a given season as opposed to the Accumulated Cyclone Energy (ACE), because the wind speed estimated by the tracker is subject to substantial uncertainty, which is amplified by the quadratic dependence of the ACE on wind speed, thereby making the ACE less robust, compare Section 2.4.

The landfall forecasts for System 4 are calculated for the 34-year 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 Hurdat 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. In order to obtain robust estimates of the means, two different time periods are used: 1950-2014 for the basin-wide variables (number and activity) and 1900-2014 for the landfall variables. The choice of these time periods is discussed in Section 2.1. For every year, there is one Atlantic hurricane season, starting on May 1 and ending on November 30.

In order to describe the forecast calculation, we use some notation: Let t denote the year, $X(t)$ the variable we seek to forecast (for example the number of storms that make landfall), $x(t)$ the observed value of the variable (that is 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 (compare Section 2.1) and even a perfect forecast model will sometimes yield a forecast, $F[X(t)]$, which differs from the subsequently observed outcome, $x(t)$, due to the partly chaotic nature of the climate system. Further, 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 May 1 and for those starting on August 1.

Now, for each season and variable, the basic idea is to construct the forecast $F[X(t)]$ by rescaling ensemble mean of the season using a *factor* such that the average forecast for the full reforecast period equals the historically observed average value of the variable. Formally we may express this as:

$$F[X(t)] = \frac{\bar{x}}{\langle \bar{\hat{x}} \rangle} \langle \hat{x}(t) \rangle, \tag{1}$$

Further, to make the hurricane forecasts more robust, we have used the corresponding storm factors also for the hurricane forecasts. And, similarly, to make the landfall forecasts more robust, we have used the U.S. storm landfall factor for all coastal segments. The reason being that the U.S. coastline is long and its landfall record is reasonably reliable going back to the beginning of the last century (compare section

2.1). This is key, especially when forecasting landfall for shorter segments of the coastline where landfall is infrequent.

By scaling the ensemble means in this fashion, the dispersion among the forecasts increases and becomes more comparable to the historically observed interannual dispersion (see figures 13 and 14 in the appendix). Also, one may note that by choosing this method we accept a systematic error or bias when scored against the observed average for the reforecast period in order to gain robustness. One may also note that the 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.

This approach is similar, but not identical, to the approach used by the ECMWF in the operational forecasts of basin-wide storm and hurricane frequency. The difference being our choice of factors and longer time periods for the historical averages which we expect to yield more robust forecasts going forward (again, compare section 2.1).

3 Results

In this section, the main results are described. First, we look at the forecasts, their precision and quality, and estimate the maximum attainable skill by decomposing the variance and analysis of the model error. Second, we look at the results for different segments of the coast and, finally, at the impact of shortening the forecast lead time.

All the forecasts that we describe in this section have been made using our System 4-based forecast method, LF+, described in Section 2. Further, the forecasts have always been issued on May 1 for the full hurricane season that is the period May 1 through November 30, except in Section 3.4 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.

3.1 Mean error and RMSE

In order to obtain an overview of the forecasts and their quality, one may start by comparing time series of the forecasts to observations (Hurdat). Figure 3 shows such time series for two key basin-wide variables – storm number and storm activity[13] – as well as for two key landfall variables – the number of storm landfalls somewhere along the North American coast (N.A. landfall) and the number of landfalls in the U.S. (U.S. landfall), see section 2.6 for definitions of the variables.

First, we may note that there appears to be a weak increasing trend in the observations, whereas no trend can be spotted in the forecasts. However, the variation in the observations is considerable and the perceived trend may be spurious. Further, the difference between forecast and observation has a random component to it: sometimes the forecast is above and sometimes it is below the observation, sometimes the error is large and sometimes it is small. The historical record contains a number of extreme years with high numbers of systems, high activity and a large number of landfalls. The forecasts are never that extreme. And, similarly, the forecasts are not extreme on the low side either. This is expected because the forecasts are based on ensemble means whereas for each year the observed outcome is a single realization of the partly chaotic climate system. Although less apparent in the figure, there are also, by

[13] Storm activity is estimated as the number of storm days.

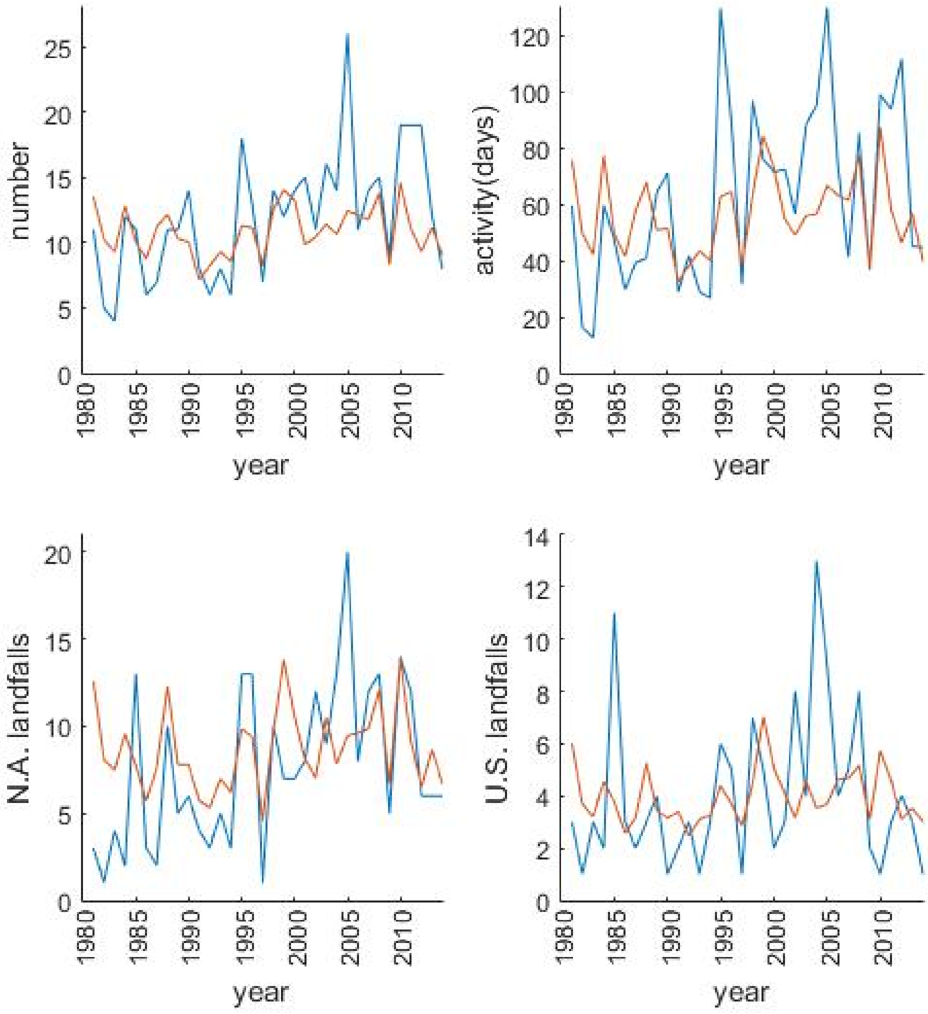

Figure 3: Time series of forecast (orange) and observation (blue) for storm *number*, *activity*, *N.A. landfall*, and *U.S. landfall*.

construction[14], systematic errors in the forecasts when evaluated against observation for the 1981-2014 period. Table 2 shows the mean forecasts and observations as well as the time period systematic (mean) errors and the Root Mean Square Errors (RMSEs) in the forecasts.

Table 2: Mean forecast and observation, their difference (systematic error), and the RMSE for the period 1981-2014 for key storm and hurricane variables.

	STORM				HURRICANE			
variable	forecast	observation	difference	RMSE	forecast	observation	difference	RMSE
number	10.8	11.9	-1.1	4.1	6.5	6.4	0.1	2.6
activity	56	63	-6	26	19	23	-3	12
N.A. landfall	8.7	7.6	1.1	4.1	4.2	3.1	1.1	2.6
U.S. landfall	4.0	4.0	0.0	3.0	2.2	1.7	0.5	2.1

Here we may note that the average number of storms that formed over the Atlantic during the period was 11.9 whereas we forecasted on average 10.8. Since the climate system is in part chaotic we do not expect forecast and observation to coincide, however, we do expect, a zero systematic error to be within the range of what we could have observed. In order to check this, we have estimated a two-sided 95-percent confidence interval[15] for the systematic error. It was found to be $[-2.7, 0.1]$, showing that a zero systematic error is close to the upper bound, but within the statistically permissible range.

Turning to the cyclone activity forecasts, one may note that, on average, S4LF+ forecasts lower cyclone activity than observed. For example, during the period on average 63 storm days were observed each season, whereas only 56 were forecasted. Thus, the systematic error relative to the observed mean is close to 10 percent. Also in this case a zero systematic error is just inside the statistically permissible range as indicated by the corresponding 95-percent confidence interval. Part of the explanation for this lies in a too short average life time of the forecasted cyclones, the average lifetime observed was 5.3 days whereas the average life time forecasted was 4.2 days. Another part of the explanation lies in that the total activity is the product of the average number of storms and their average life time, and, we already know from the above, that the forecasted storm number is on the low side compared to the observed. (The System 4 climatology and life time of cyclones is discussed in appendix B.5.)

For the number of hurricanes, the situation looks better with a systematic error of only 0.1. However, this does not necessarily mean that we are better at forecasting hurricane than storm number – the confidence interval for the systematic error is $[-0.8, 0.9]$, hence the apparent high precision in the hurricane number is likely to be at least in part coincidental (see table 13 in the appendix for further detail and a complete set of confidence intervals). Similarly, the apparent high precision in the period average of the forecasted number of storm landfalls along the U.S. part of the coast is likely to be spurious.

The RMSE provides another view of the precision in the forecasts; in the case of storm 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; moreover, the relative RMSEs (that is RMSEs divided by the means of the corresponding observations) are 35 and 41 percent for storms and hurricanes respectively. The fact

[14] See section 2.6.
[15] All confidence intervals have been estimated using bootstrap.

that the relative RMSEs are substantial is not necessarily indicative of a poor forecast method; the large relative RMSEs may, in part, be a result of the chaotic nature of the climate system, which introduces a considerable degree of unpredictable variability. In Section 3.5 we attempt to quantify the unpredictable part of the variability.

The relative RMSE in the forecast for North American storm landfall is 48 percent, which is comparable to the relative RMSE in the basin-wide forecasts of storm and hurricane number and activity. This is encouraging because the precision in seasonal forecasting of basin-wide activity has been judged sufficient to motivate the regular issuance of forecasts. However, for the other landfall forecasts the relative RMSEs are significantly higher. While discouraging, this is in part expected due to the high degree of variability in the landfall statistics in the observed historical record (for details, see table 1 in the appendix). Again, understanding the chaotic and unpredictable part of the climate system becomes important to assess the quality of the forecasts and what room there may be for improving them. (We will look closer at the unpredictable part of the variability in Section 3.5.)

In summary, the time series of forecast and observation clearly demonstrate that seasonal forecasting is challenging and our forecasts are far from perfect.

3.2 Correlations

In this section we look closer at the forecasts and how we may quantify their usefulness. We will use the term skill as is standard in the forecast community to loosely describe the quality of a forecast. 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 4 shows such plots for the two key basin-wide variables – storm number and storm activity – and the two key landfall variables – N.A. and U.S. storm landfall. (For corresponding hurricane plots see figure 14 in the appendix)

As can be seen in the figure, in all cases except for U.S. landfall, there is a tendency for a high forecast to coincide with a high observation; and, similarly, a low forecast often coincides with a low observation. However, also in the cases where there is an association between forecast and observation, the points are scattered demonstrating the error associated with the forecasts. One natural question to ask is: How much of the observed variability can be explained by the forecasts, and how much remains and has to be attributed either to noise or some other unknown factor?

The main statistic that we have used to quantify the quality of the forecasts, the Spearman rank correlation[16], r, addresses this question; its square can be interpreted as an estimate of the fraction of the variance that is explained by the forecasts. For example, $r = 0.6$, would mean that the forecasts explain 36 percent of the variability among the observations, the remaining 64 percent are either due to noise or some unknown factor. In table 3, rank correlations, the corresponding confidence intervals, and P-values[17] are shown for four of the key variables forecasted. The confidence intervals have been estimated using bootstrap[18].

As can be seen in the table, the rank correlations for storm and hurricane number as well as for activity are close to 0.6 with P-values below 0.001 in all cases. For storms making landfall anywhere along the

[16] The rank correlation is generally speaking similar to, but more robust than, the standard linear correlation. See Press et al. (1992) for a brief description of the rank correlation.

[17] The P-value here is the probability that we would observe the estimated rank correlation of $r = 0.6$ despite a null hypothesis of $r = 0$ being true.

[18] Bootstrap here refers to resampling of forecasts and observations across years and ensemble members.

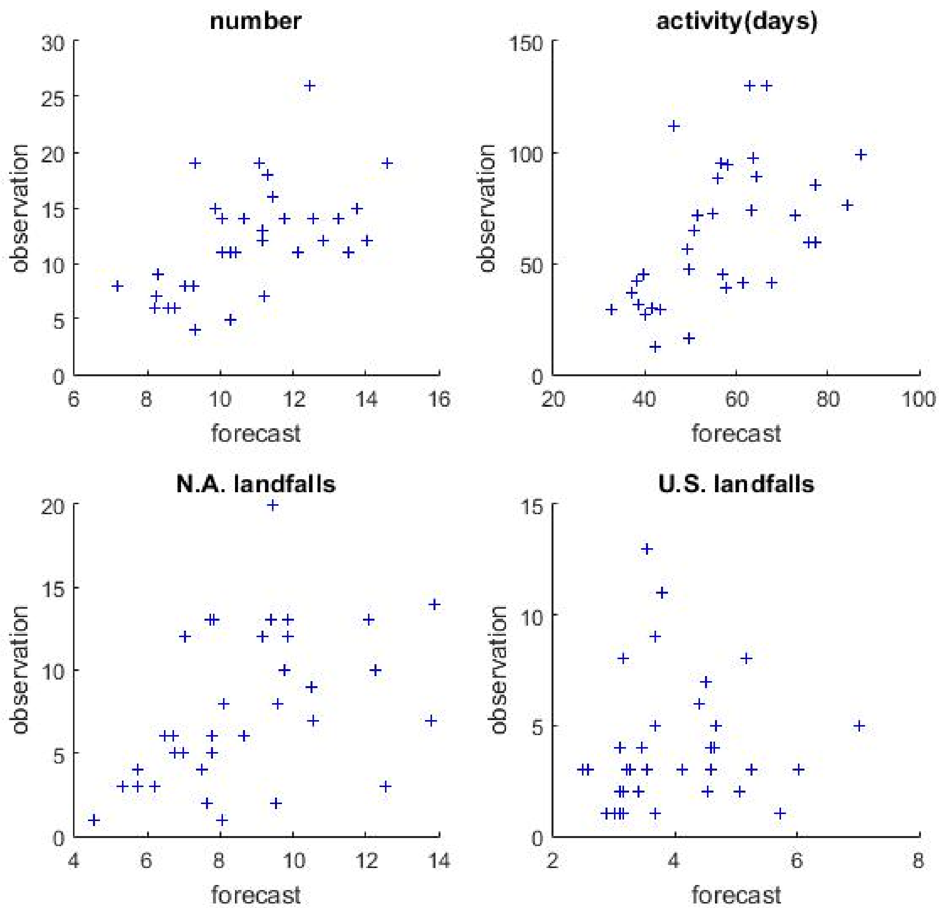

Figure 4: Scatter plots of observation versus forecast for storm number, activity, N.A. landfall, and U.S. landfall. Note the different scales of the x- and y-axes.

North American coast, $r = 0.53$ 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 Spearman rank correlation is only weakly positive (0.29) for U.S. storm landfall and the associated P-value is 0.09, possibly indicating a weak correlation and some associated forecast skill. For hurricane landfall along the U.S. coast, the Spearman rank correlation is even lower and non-significant by any standards. (For landfall statistics on different segments of the coast, see Section 3.3.)

Table 3: Spearman rank correlations, r, for forecasts of key storm and hurricane variables for the period 1981-2014.

	STORM				**HURRICANE**			
variable	r	P-value	lower	upper	r	P-value	lower	upper
number	0.57	0.0004	0.3	0.8	0.55	0.0007	0.3	0.7
activity	0.61	0.0001	0.3	0.8	0.61	0.0001	0.4	0.8
N.A. landfall	0.53	0.001	0.2	0.7	0.29	0.10	-0.1	0.6
U.S. landfall	0.29	0.09	-0.1	0.6	0.11	0.55	-0.2	0.4

Another basic way to quantify the usefulness of the forecasts is by counting the number of years for which the forecast deviate from the mean of the forecasts in the same direction as the observation deviate from mean of the observations. In table 4 we can see that in the one extreme (in the case of storm number and storm landfall in North America) the forecast gets the direction right 76 percent of the time, whereas the in other extreme (the case of landfalling U.S. hurricanes) the forecast direction is right only 56 percent of the time.

Table 4: Fraction of years when the forecast and the observation deviate in the same direction from their respective means.

	Storm	Hurricane
number	0.76	0.59
activity	0.68	0.68
NA landfall	0.76	0.65
US landfall	0.59	0.56

Yet another way to understand the usefulness of the forecasts of, for example, the number of N.A. storm landfalls is by classification of each year either as a *low*, a *normal*, or a *high* year as follows: Out of the total of 34 years considered, the eleven years with the lowest number of N.A. storm landfalls are placed into the low category, the twelve next into the normal category, and the eleven last into the high category. Similarly, the forecasts are ranked from lowest to highest and the eleven with the lowest rank are classified as low, the next twelve as average, and the last eleven as high. Using this classification, we can easily compare the number of correct and incorrect forecasts for each one of the categories.

Table 5 shows forecast and observation for N.A. storm landfall as an example. In total, 14 out of the 34 forecasts were correct and 20 were incorrect. However, only once a low season was 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. storm landfall forecasts, in total 13 out of 34 forecasts are correct, but high and low forecasts, are often completely wrong, see table 5.

Table 5: Forecasts of *low*, *normal*, and *high* seasons for N.A. and U.S. storm landfalls.

N.A. landfall		forecast		
		low	normal	high
obser- vation	low	6	4	1
	normal	4	3	5
	high	1	5	5

U.S. landfall		forecast		
		low	normal	high
low	L	6	2	3
normal	N	3	4	5
high	H	2	6	3

In summary, the data and analyses described show that there is a moderate association between forecast and observation for N.A. storm landfalls comparable in strength to the associations for the basin-wide variables storm number and storm activity; however, no clear association can be spotted in the case of U.S. landfall. For hurricane strength systems, the associations for the basin-wide variables are equally strong as for the storm strength systems, but for the landfall variables, no clear associations are present (the interested reader is referred to figure 14 in appendix C).

3.3 Landfall along different segments

In order to better understand the method's ability to forecast landfall, the North-American coastline was divided into different segments. Figure 5 shows the points along the coastline that were used to define the segments, for example the longest segment (a-h) corresponds to what we call the North American (N.A.) coast. We also consider landfall in two metropolitan areas: Miami and New York.

First, we may consider the observed average numbers of storm and hurricane landfalls along the different segments for the period 1900-2014[19]. The red line in figure 6 shows these averages and the red 95-percent confidence intervals indicate the associated uncertainty. We can see that the uncertainty is the largest in absolute terms for N.A. storm landfall (about plus/minus one storm), however, on a relative basis, the uncertainty is larger for the shorter segments and especially for the metropolitan regions where landfall is

[19] The landfall record as captured by Hurdat is regarded as reliable back to 1900 at least for the U.S. part of the coastline.

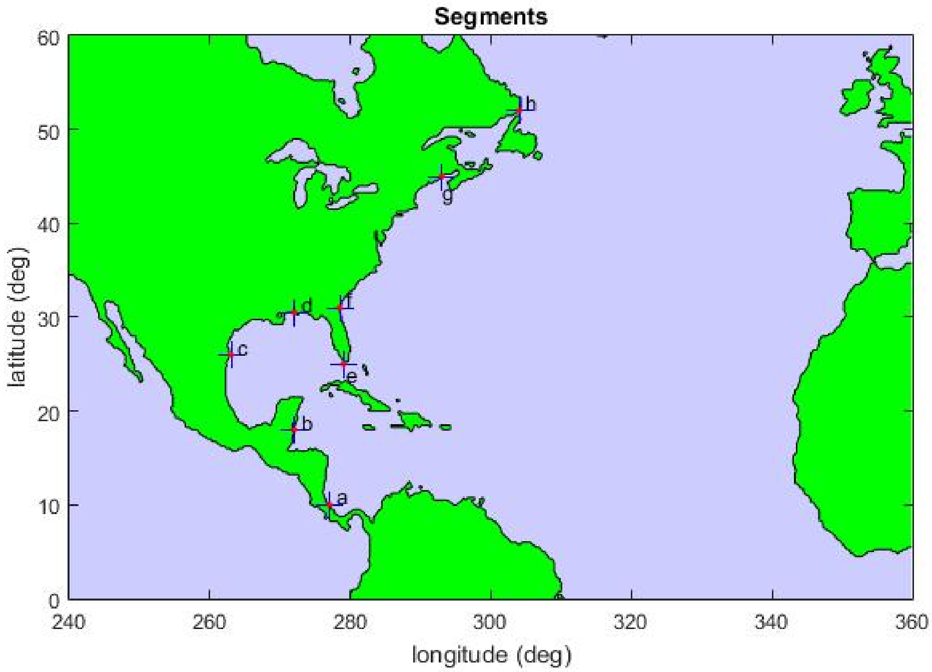

Figure 5: Points showing the division of the North American coastline into segments.

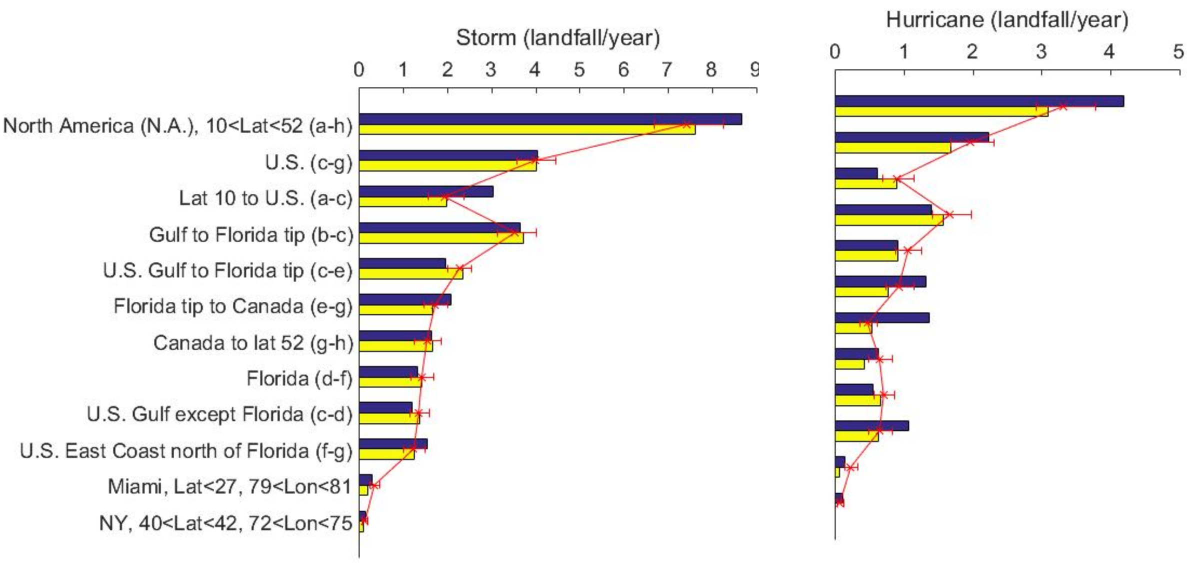

Figure 6: Average number of landfalls per year along different segments of the North American coastline. Forecasted landfalls (blue bars) and observed landfalls (yellow bars). Data for the forecast period 1981-2014. The red line and error bars show the average annual number of observed landfalls for the period 1900-2014 and the associated 95-percent confidence intervals. See figure 5 for the locations of the start and end points of the segments.

infrequent and the statistics become poor. One may also note that the average number of storm landfalls during the 1981-2014 period (yellow bars) are within the confidence intervals except for the shortest segments (Miami and N.Y.), where statistics are less reliable. The average number of hurricane landfalls for the 1981-2014 period are inside the intervals for some segments and outside for some.

Second, we may compare the forecasts to observation. In the case of storms, the forecasts (blue bars in the figure) are, with a few exceptions, within the confidence intervals. The main exceptions are for N.A. landfall and landfall south of the U.S. where the forecasts are significantly higher than observations. This may be due to the formation of a number of unphysical low pressure systems just north of South America early in the year in the System 4 forecast model as discussed by Manganello et al. (2016). It may also be due to underreporting of weak storm landfall outside the U.S. during the early parts of the previous century, which would push down the corresponding estimates of the confidence intervals.

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 misclassification of storm strength systems as hurricanes and it may also be due to model misclassification of a number of non-tropical storm systems moving off the U.S. main land and curving back to make landfall thereby inflating the landfall count in this region (see System 4 model climatology in the appendix B).

Moving on to the skill and error associated with the forecasts, one may first note, as expected, that the RMSEs are substantial 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 3.2 that the skill is higher for 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 coast south of the U.S. (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). (Table 15 in the appendix contains RMSEs and correlations for reference.)

In summary, the average forecasts for the different segments are in reasonable agreement with observation. Forecast skill for Gulf landfall is high in contrast to U.S. East Coast landfall, where no skill is observed.

3.4 The impact of lead time on forecast skill

To gain insight into how lead time impacts forecast skill, forecasts with different start dates (May 1 and August 1) were compared with each other. The forecasts were made for the peak of the hurricane season only, that is for the period August, September, and October (ASO). ASO is the most active part of the Atlantic hurricane season accounting for about 75 percent of the total activity. (The distribution of storm activity over the year is shown in figure 9 in the appendix.)

Table 6 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 in the U.S., which increases to about 0.3 (with the p-value of being non-zero dropping just below 0.1). The RMSE in the forecasts also improves by 10 to 15 percent for storm and hurricane number and activity. The improvement is smaller (about 5 percent) for landfall variables, except in the case of N.A. hurricane landfall where the improvement is close to 10 percent.

The highest skill is obtained for the basin-wide forecasts, where we reach correlation levels close to 0.7. For the hurricane number forecast we reach a correlation near 0.8. Dividing the seasons into low, normal, and high as in Section 3.2 shows that the August hurricane number forecast is correct in 19 cases,

Table 6: 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 May 1 and August 1 respectively. The period forecasted is August, September, and October.

	Spearman rank correlation		RMSE	
	May	August	May	August
STORM				
number	0.62	0.66	3.1	2.8
activity	0.61	0.72	22	19
N.A. landfall	0.58	0.59	3.2	3.1
U.S. landfall	0.25	0.31	2.8	2.7
HURRICANE				
number	0.62	0.77	2.3	1.9
activity	0.60	0.75	12	10
N.A. landfall	0.35	0.50	2.1	1.9
U.S. landfall	0.08	0.30	1.7	1.7

a near miss in 14, and a total miss in one case. Interestingly, both the May and August forecasts for storm number as well as the May forecast for hurricane number are better according to this classification metric: they are correct in 21 cases, a near miss in 12, and a total miss in one case only. Higher correlation coefficients are thus not necessarily reflected in a better classification score.

In conclusion, the total improvement of the forecast skill obtained by focusing on the ASO period and shortening the lead time by three months is significant. Especially, for the basin-wide forecasts (for example hurricane number) where we reach correlation levels close to 0.7. Also the RMSEs are reduced by ten to fifteen percent for these forecasts. The improvements are smaller for landfall. Moreover, considering the improvements in the forecast skill when narrowing the forecast period to ASO (compare the correlations in tables 3 and 6), and the fact that most of the destructive hurricane activity occurs during this period, one may consider issuing operational forecast for this period separately.

3.5 Chaotic part of variability

The climate system is in part chaotic, which limits the skill attainable by any forecast method. One strength of our method is its ability to lay bare this chaotic behavior by using the ensemble of forecasts. We have devised an admittedly crude procedure to use this strength to estimate the impact of the chaotic dynamics and explore the limits of the maximum attainable forecast skill.

First, in order to estimate the impact of the chaotic dynamics on the variability of the forecasted variables, we have used the 51 ensemble members to estimate the *intra-annual* variance for each year; and, then we have computed the average of these variance estimates across the 34-year period to obtain an estimate of the chaotic, unpredictable, part of the variability. Next, we have estimated the predictable part of the variance as the total variance minus the intra-annual variance[20]. Table 7 shows that the fraction of the total variance that is potentially predictable according to this analysis. The fraction is below 40 percent for all variables; and, for the landfall variables, the fraction drops below 15 percent. This indicates that the chaotic dynamics severely limits the maximum attainable skill, in particular for landfall forecasts.

Table 7: Decomposition of the total variability in the forecast variables into an interannual (potentially forecastable) and an intra-annual (chaotic) component. Definitions in the text.

variable	fraction of total variance	
	inter	intra
STORM		
number	0.26	0.74
activity	0.34	0.66
NA landfall	0.08	0.92
US landfall	0.14	0.86
HURRICANE		
number	0.20	0.80
activity	0.26	0.74
NA landfall	0.07	0.93
US landfall	0.09	0.91

If we assume that the real climate system has a similar relation between the intra-annual and total variability as the forecasts do, then any attempt to score the forecasts against observations of the climate system (for example Hurdat) using the RMSE will show limited skill. Moreover, under this assumption, we can estimate the magnitude of the chaotic part of the variability of the real climate system and hence obtain an estimate of the "smallest possible" forecast error[21]. Table 8 shows the smallest possible forecast error as well as the actual. For all variables, the smallest error is 75 percent or more of the actual, indicating limited room for improving the forecasts. We will examine the forecast error and the possibility to reduce it from another angle in the following section.

[20] We have also estimated the predictable part of the variance directly as the variance in the *interannual* variance in the annual ensemble mean over the 34-year period. That alternative approach yields very similar results.

[21] The "smallest possible" forecast error is estimated as the square root of the product of the total observed variance in the real climate system and the ratio between the intra-annual and total variance estimated using the forecasts.

Table 8: Estimated *smallest* and *actual* forecast error relative to observation (Hurdat). The smallest possible forecast error has been estimated as described in the text. Both errors are RMSEs.

variable	**forecast error**			
	Storm		Hurricane	
	smallest	actual	smallest	actual
number	3.7	4.1	2.4	2.6
activity	22.3	26.4	11.7	12.3
N.A. landfall	3.9	4.1	2.3	2.6
U.S. landfall	2.3	3.0	1.6	2.1

3.6 Perfect-model vs. actual predictability

The climate system is not only in part chaotic (as discussed above), but it is also incredibly complex and all models to date are simplifications of the true system. Our forecast method LF+ builds on System 4 and its modeling of the climate system. Hence, when we score our forecasts against Hurdat, the total forecast error will be the result of (a) the inherent chaos of the dynamical system and (b) forecast model errors. If we instead score our forecasts (which are ensemble mean forecasts) against each one of the 51 ensemble members, then both the forecast and the outcome will be subject to the same underlying model errors. For this reason, one may hypothesize that the total forecast error should be lower and the skill higher when forecasting the ensemble members than when forecasting Hurdat observations. In table 9 the forecast errors are shown in both cases.

First, we may note that when forecasting the ensemble members, the systematic forecast errors disappear by construction. The RMSEs, on the other hand, remain because they are the result of the stochastic and chaotic nature of the model. Second, we may note that RMSEs in storm number and activity drop as expected. The storm number drops from 4.1 to 3.1 and the activity from 26 to 20, which indicate that the forecasts can be improved by reducing the model error. Similarly, there appears to be room for improvement in the corresponding hurricane forecasts. However, the RMSEs in the landfall forecasts increase. This indicates that the forecasts are overdispersed, meaning that the variability between the ensemble members is too large, at least with respect to the landfall variables. It may well be that the forecasts are overdispersed (but to a lesser degree) also with respect to the number and activity variables.

We have also compared the rank correlations obtained when forecasting the ensemble members with those previously obtained when forecasting observation. Briefly, the correlations obtained when forecasting the ensemble members are in all cases well inside[22] the corresponding confidence intervals obtained when forecasting observation, which indicates that there is limited room to improve this metric by reducing the model error. Part of the explanation for this may lie in that correlations are insensitive to systematic biases (as are linear correlations for that matter) and they are also relatively robust with respect to extreme observations and outliers in contrast to the RMSE. (For reference, a full set of confidence intervals and correlations can be found in table 13 in the appendix.)

[22] In fact, the correlations obtained when forecasting the ensemble members are somewhat lower than those obtained when forecasting observation.

Table 9: Forecasting ensemble members versus forecasting observations. The table shows the RMSEs obtained when scoring the forecasts against the ensemble members (*ensemble*) and when scoring against Hurdat (*observed*). *est* denotes the estimated RMSE (mean estimate), *lower* and *upper* the bounds of the associated 95-percent confidence intervals. Table 13 in the appendix contains a more complete data set.

variable	**Storm**				**Hurricane**			
	ensemble	observed			ensemble	observed		
		est	lower	upper		est	lower	upper
number	3.1	4.1	3.2	5.7	2.4	2.6	2.1	3.6
activity	20	26	21	34	10	12	10	16
N.A. landfall	5.7	4.1	3.3	5.6	3.9	2.6	1.9	3.9
U.S. landfall	3.5	3.0	2.2	4.3	2.6	2.1	1.5	2.9

In conclusion, we have made an admittedly rather crude analysis of the impact of the model error which indicates that (a) the forecasts are overdispersed with respect to the landfall variables and possibly also with respect to the basin-wide variables, and (b) elimination of the model error may reduce the RMSE for the basin-wide variables by up to 25 percent. The model error's impact on the rank correlations is small and the analysis does not suggest any potential to improve that metric.

3.7 Comparison with observed correlations

Here we compare and discuss the Hurdat-Hurdat correlations (OO-correlations, where O stands for observation), S4LF+-S4LF+ correlations (FF-correlations, where F stands for forecast), and the S4LF+-Hurdat cross correlations (FO-correlations). Table 10 shows these correlations for storm strength systems (data on hurricane strength systems can be found in table 14 in the appendix).

First, one may note that the OO-correlation between number and landfall is higher for the full North American coast (N.A. landfall) than for the U.S. part (U.S. landfall). The correlations are 0.7 and 0.4 respectively. This is not surprising because the sensitivity to the track is reduced when we allow landfall anywhere along the coast. Similarly, the correlation between forecasted number and observed landfall is higher for the entire North American coast than the U.S. part alone (0.5 and 0.3 respectively).

Second, the OO-correlations between the basin-wide variables and the landfall variables are in the range 0.4 to 0.7, which is significantly lower than the corresponding FF-correlations which are close to 0.9 in all cases. A similar pattern is visible for the corresponding hurricane correlations (see table 14 in the appendix). This is consistent with the fact that the forecasts are based on ensemble means, which suppresses the random, chaotic components associated with storm formation and movement. Recalculation of the correlations between the basin-wide and the landfall variables using each reforecast in the ensemble instead of the ensemble means, yields correlations in the line with those empirically observed.

Third, one may note that, the forecasts of the basin-wide variables correlate as strongly with the observed outcomes for the landfall variables, as does the "direct" forecasts of the landfall variables. Thus, according to this correlation metric, an "indirect" forecast of landfall using the basin-wide variables is associated with the same skill as the direct forecasts, which suggests that the forecast skill of S4LF+

Table 10: Spearman rank correlations between different storm variables for the period 1981-2014. In the orange box Hurdat-Hurdat correlations, in the yellow S4LF+-S4LF+ correlations, and in the green Hurdat-S4LF+ cross correlations.

		Hurdat				**S4LF+**			
		Number	Activity	N.A. Landfall	U.S. Landfall	Number	Activity	N.A. Landfall	U.S. Landfall
Hurdat	Number	1.0	0.9	0.7	0.4	0.6	0.6	0.6	0.5
	Activity	0.9	1.0	0.7	0.5	0.6	0.6	0.6	0.5
	N.A. Landfall	0.7	0.7	1.0	0.6	0.5	0.5	0.5	0.5
	U.S. Landfall	0.4	0.5	0.6	1.0	0.3	0.3	0.3	0.3
S4LF+	Number					1.0	1.0	0.9	0.9
	Activity					1.0	1.0	0.9	0.9
	N.A. Landfall					0.9	0.9	1.0	0.9
	U.S. Landfall					0.9	0.9	0.9	1.0

predominantly arises from its ability to forecast storm number and to a smaller degree from its ability to forecast tracks.

Fourth, as previously noted in Section 3.2, the correlation between forecast and observation for N.A. storm landfall is 0.5, whereas the correlation between forecast and observation for basin-wide storm number is 0.6. This suggests that landfall, for the full North American coast, can be forecasted with almost the same skill as basin-wide storm number. (In both cases, the statistical uncertainty associated with the forecasts is considerable, in part due to model error, in part due to the chaotic dynamics embedded in the climate system.)

4 Conclusion

We have developed a method, LF+, to calculate landfall risk using the ensembles of storm tracks generated by System 4 (Section 2). Briefly, a global land mask of suitable resolution was constructed and algorithms to identify landfalling cyclones were developed. Using cyclone tracks from Hurdat, ERA, and System 4, a mapping between observed and modeled maximum wind speeds was created, enabling classification of System 4 model cyclones as either storms or hurricanes. Finally, 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 their 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 storm than hurricane landfall, part of the explanation being the more robust nature of the storm statistics (see Section 2.4). Second, one may note that the forecast skill measured by the rank correlation, r, varies between different segments of the coast.

The highest correlations ($r = 0.5$) were found for landfall along three segments: (i) the entire North American coast, (ii) the coast south of the U.S., 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 U.S., 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 lifetime and a too quick steering towards 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 to November) to the peak of the season (ASO), the correlation increases to 0.6 for storm 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. When we, in addition, shorten the lead time by three months using the forecasts issued on August 1 instead of those issued on May 1, the storm landfall skill increases only marginally. Turning to the basin-wide variables (number and activity), we have the opposite situation: 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 storms and to 0.8 for hurricanes. These correlations are high in comparison to those that we and others (Camp et al., 2015; Manganello et al., 2016) obtain for the full season. Moreover, the forecast RMSEs drop by between 10 and 20 percent when shortening the lead time. Considering that about 75 percent of the destructive hurricane landfall activity occurs during ASO and that the forecast skill is higher for this period, for basin-wide as well as some landfall variables, it may be worth issuing forecasts specifically for ASO.

In order to assess the quality of the forecasts and gauge the potential for improvement, 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 is such that it limits the usefulness of the forecasts (especially when forecasting landfall) 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 storm 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 (a) model error and (b) the inherently chaotic nature of the climate and weather system. The latter will ultimately set a limit to the skill attainable by any model. In order to estimate the size of this inevitable chaotic part of the error, we have decomposed our model's variability in an intra- and an inter-seasonal part, where the former was used as a proxy for chaotic part of the variability. Taking this, admittedly rather crude approach, the chaotic part was found to dominate, leaving only an estimated 10 to 25 percent room for RMSE reduction, depending on the variable forecasted. We have also made a crude analysis of the impact of the model error using a prefect-model approach, which indicates that (a) the forecasts are overdispersed with respect to the landfall variables and possibly also with respect to the basin-wide variables, and (b) elimination of the model error may reduce the RMSE for the basin-wide variables significantly (on the order of 25 percent).

For society, seasonal forecasting of landfall risk is more important than forecasting of basin-wide cyclone activity, the simple reason being that the largest loss of life and economical damage occur onshore. Nevertheless, leading forecasting organizations such as ECMWF and NOAA, only issue forecasts of basin-wide Atlantic cyclone activity, one main reason being the difficulty to forecast landfall on a seasonal timescale with sufficient skill. However, as demonstrated in this memo, for the North American coast and some of its segments, the skill in forecasting landfall is now approaching that of the forecasts for basin-wide Atlantic cyclone activity.

Currently, forecasts of basin-wide cyclone activity are used by, for example, reinsurance companies to indirectly estimate landfall risk and price reinsurance contracts. Our analysis suggests that the direct forecasts of landfall risk that can be obtained from S4LF+ are as good as these indirect forecasts, at least for the full North American coast. 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.

In summary, the main conclusion of this memo is that System 4 with LF+ attached can be used to issue seasonal forecasts of landfall risk for the entire North American coast and for some of its segments. However, the forecasts are subject to substantial uncertainty, in part due to model error, in part due to the chaotic dynamics embedded in the climate system. This limits the usefulness of the forecasts mainly to organizations that plan and operate with a statistical mind-set on a multi-year horizon.

Acknowledgements

The authors are grateful to Prof. Erland Källén, Director of Research at ECMWF, for suggesting the topic, to Prof. Lars Bergman and Dr Suzana Camargo for providing helpful perspectives, to Mr David Richardson for sharp and constructive comments on the manuscript.

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? *Bulletin of the American Meteorological Society*, **93 (5)**, 631–651.

Bergman, D. L., 2016: Seasonal forecasting of storm and hurricane landfall using ECMWF's System 4, master's thesis at Stockholm University.

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/National Weather Service, National Centers for Environmental Prediction, National Hurricane Center.

Bove, M. C., J. J. O'Brien, J. B. Eisner, C. W. Landsea, and X. Niu, 1998: Effect of El Niño on US landfalling hurricanes, revisited. *Bulletin of the American Meteorological Society*, **79 (11)**, 2477–2482.

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*, **4**, 325.

Camp, J., 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. *Quarterly Journal of the Royal Meteorological Society*, **141 (691)**, 2206–2219.

Carrasco, C. A., C. W. Landsea, and Y.-L. Lin, 2014: The influence of tropical cyclone size on its intensification. *Weather and Forecasting*, **29 (3)**, 582–590.

Dee, D., et al., 2011: The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137 (656)**, 553–597.

Duan, W. and C. Wei, 2013: The "spring predictability barrier" for enso predictions and its possible mechanism: results from a fully coupled model. *International Journal of Climatology*, **33 (5)**, 1280–1292.

Emanuel, K., 2005a: *Divine wind: The history and science of hurricanes*. Oxford University Press.

Emanuel, K., 2005b: Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, **436 (7051)**, 686–688.

Emanuel, K., R. Sundararajan, and J. Williams, 2008: Hurricanes and global warming: Results from downscaling ipcc ar4 simulations. *Bulletin of the American Meteorological Society*, **89 (3)**, 347.

Gray, W. M., 1984: Atlantic seasonal hurricane frequency. part i: El Nino and 30 mb quasi-biennial oscillation influences. *Monthly Weather Review*, **112 (9)**, 1649–1668.

IPCC, 2013: Summary for policymakers. in: Climate change 2013: The physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Klotzbach, P. J., 2007: Revised prediction of seasonal Atlantic basin tropical cyclone activity from 1 August. *Weather and Forecasting*, **22 (5)**, 937–949.

Landsea, C., J. Franklin, and J. Beven, 2013: The revised Atlantic hurricane database (hurdat2). *United States National Oceanic and Atmospheric Administration*.

Landsea, C. W., B. A. Harper, K. Hoarau, and J. A. Knaff, 2006: Can we detect trends in extreme tropical cyclones? *Science*, **313 (5786)**, 452–454.

Landsea, C. W., G. A. Vecchi, L. Bengtsson, and T. R. Knutson, 2010: Impact of duration thresholds on Atlantic tropical cyclone counts. *Journal of Climate*, **23 (10)**, 2508–2519.

Madec, G., 2008: Nemo reference manual, ocean dynamics component: Nemo-opa. note du pole de modelisation. reference manual 27, Institut Pierre-Simon Laplace (IPSL), France.

Manganello, J. V., et al., 2016: Seasonal forecasts of tropical cyclone activity in a high-atmospheric-resolution coupled prediction system. *Journal of Climate*, **29 (3)**, 1179–1200.

Molteni, F., et al., 2011: The new ecmwf seasonal forecast system (System 4). Technical memorandum 656, internal report, European Centre for Medium-Range Weather Forecasts (ECMWF).

Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer, 2008: Stochastic parametrization and model uncertainty. Technical memorandum 598, internal report, European Centre for Medium-Range Weather Forecasts (ECMWF).

Pielke Jr, R. A., J. Gratz, C. W. Landsea, D. Collins, M. A. Saunders, and R. Musulin, 2008: Normalized hurricane damage in the United States: 1900–2005. *Natural Hazards Review*, **9 (1)**, 29–42.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: *Numerical recipes in C*. 2d ed., Cambridge university press Cambridge.

Van der Grijn, G., J.-E. Paulsen, F. Lalaurette, and M. Leutbecher, 2005: Early medium-range forecasts of tropical cyclones. Ecmwf newsletter no. 102, European Centre for Medium-Range Weather Forecasts (ECMWF), 7–14 pp.

Villarini, G., G. A. Vecchi, and J. A. Smith, 2012: Us landfalling and North Atlantic hurricanes: Statistical modeling of their frequencies and ratios. *Monthly Weather Review*, **140 (1)**, 44–65.

Vitart, F., J. Anderson, and W. Stern, 1997: Simulation of interannual variability of tropical storm frequency in an ensemble of GCM integrations. *Journal of Climate*, **10 (4)**, 745–760.

Vitart, F., F. Prates, A. Bonet, and C. Sahin, 2011: New tropical cyclone products on the web. Ecmwf newsletter no. 130, European Centre for Medium-Range Weather Forecasts (ECMWF), 17–23 pp.

Webster, P. J., G. J. Holland, J. A. Curry, and H.-R. Chang, 2005: Changes in tropical cyclone number, duration, and intensity in a warming environment. *Science*, **309 (5742)**, 1844–1846.

A Appendix A - Intensity calibration

While the 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, as discussed, not sufficient to fully capture the intense winds near the center of a hurricane. This and other System 4 limitations make forecasts of cyclone frequency, landfall, and in particular cyclone intensity challenging. Care must be exercised when using System 4 for this purpose. One key question for this study is how a tracked cyclone's wind speed and pressure may or may not be used to draw conclusions about its intensity, in particular in the case of landfall.

In order to assess the precision of the tracker, we have applied it to ERA for each one of the hurricane seasons from 1989 through 2009 and compared the wind speeds and pressures obtained to those actually observed according to Hurdat[23]. Figure 7 (left column) shows the relation between Hurdat and ERA wind speed for those points on the tracks where the ERA and Hurdat tracks overlap sufficiently. A track point in Hurdat is considered to overlap with a point in ERA provided that they are not more than 300 km apart at a given point in time. Examples of Hurdat and ERA tracks and the degree of overlap are shown in figure 2.

From figure 7 it is clear that the wind speed estimated from ERA is, generally speaking, lower than the wind speed according to Hurdat. It is also clear that the relation is subject to considerable variability, that is the same wind speed in Hurdat corresponds to a rather broad range of wind speeds in ERA as estimated by the tracker. (Similarly, the pressure in Hurdat corresponds to a rather broad range of pressures in ERA.) If we calculate the average wind speed estimated from ERA for all points of borderline storm strength[24] in Hurdat we get an average of 26 knots and the corresponding average for borderline hurricanes is 33 knots.

In the present study, we are particularly interested in the possibility to forecast landfalling hurricanes and the damage related to them. We would therefore like to use either the wind speed or the pressure estimated by the tracker to draw conclusions regarding the "true" intensity of the cyclones. However, as can be seen in the figure, no clear one-to-one mapping between Hurdat and ERA wind speeds exist and the same is true for the pressure. Moreover, because our ultimate objective is to relate the intensity of the cyclones generated by System 4 to the intensity of real cyclones as recorded in Hurdat, we face an additional degree of uncertainty arising from System 4 itself.

In order to relate the intensity of the cyclones forecasted by System 4 to the intensities of real tropical storms and hurricanes, a statistical quantile-quantile (Q-Q) approach has been used. In figure 8 the Q-Q plot[25] of the wind speeds of System 4 and the wind speeds in Hurdat is shown. Here one can see that a

[23] see Section 2.1 for a brief description of ERA and the Hurdat data set.

[24] Hurdat record wind speed in 5 knot intervals, thus the 35 knots interval is the closest to borderline tropical storm strength and 65 knots the closest to borderline hurricane strength.

[25] A Q-Q plot maps the quantiles of two distributions against each other.

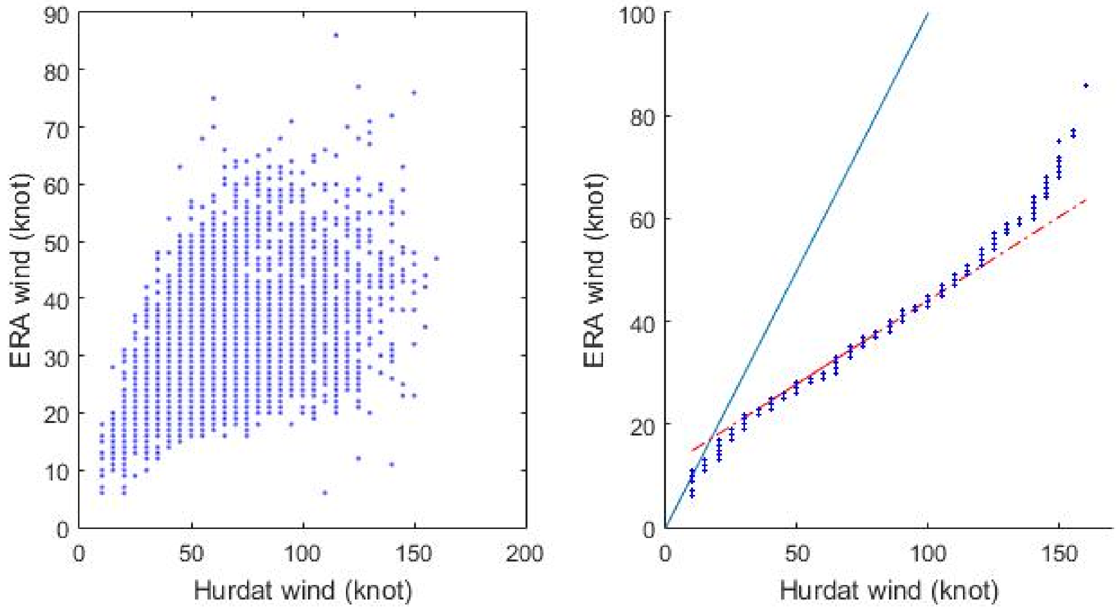

Figure 7: The relation between Hurdat and ERA wind speed along the tracks (left) and the corresponding Q-Q plot (right). Red line fitted between the first and third quantile and extrapolated. Blue line shows how a one-one relation would look.

System 4 wind speed of 24 knots corresponds to storm strength (34 knots) and a System 4 wind speed of 31 knots corresponds to hurricane strength (64 knots). We will denote System 4 cyclones with wind speeds of at least 24 knots as *tropical storms* or just *storms* and cyclones with wind speeds of 31 knots or higher as *hurricanes*. One should bear in mind that this mapping is only a reasonable approximation. A similar mapping would be obtained if we were to instead use the ERA-Hurdat wind-wind relations in figure 7.

In conclusion, the cyclones with a System 4 wind speed of 24 knots or above will be regarded as storms. Cyclones with a System 4 wind speed with 31 knots or above will be regarded as hurricanes (and storms). Due to the lack of a clear one-to-one mapping between System 4 and Hurdat wind speeds, these thresholds are only reasonable approximations.

B Appendix B - Comparison of System 4, Hurdat, and ERA climatology

In this appendix, the System 4 climatology is compared the historical as captured by Hurdat and ERA. The focus is on aspects relevant for the development and validation of our System 4 based tropical cyclone forecast method.

B.1 Tracks

As described previously, the Hurdat database contains storm tracks for all known tropical storms and hurricanes since 1851 for the Atlantic basin. The database can be regarded as complete since about 1970 following the launch of weather satellites. The ERA data set, constructed by reanalysis, contains global meteorological fields from 1979 to today (see Section 2.1).

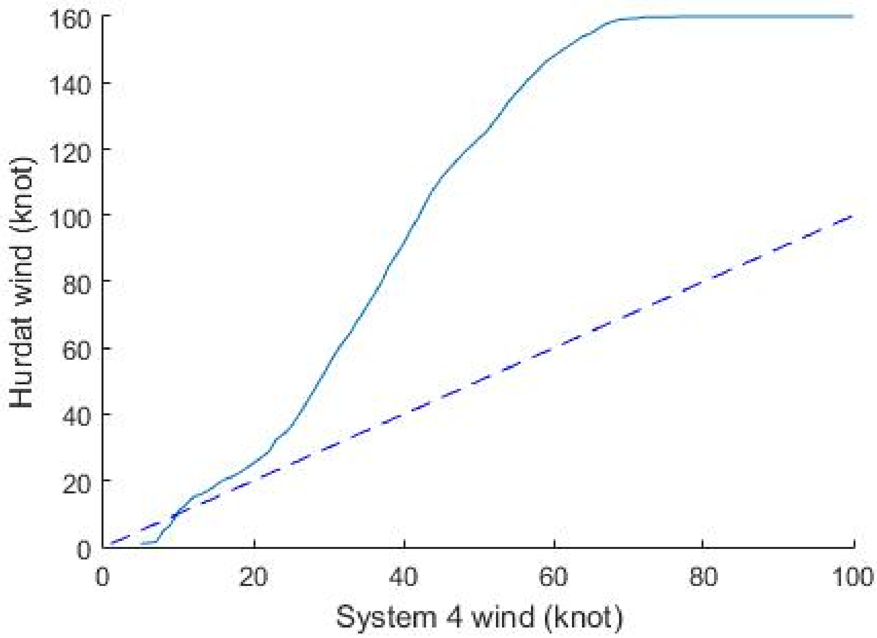

Figure 8: Q-Q plot of the System 4 wind distribution versus the Hurdat wind distribution. Each point on the curve identifies a pair of wind speeds (one System 4 and one Hurdat wind speed) that corresponds to the same quantiles in their respective distributions. The dashed line shows how a one-one relation would look

ECMWF has applied their tracker both to the System 4 reforecasts and the ERA fields for each season during the period 1989 through 2009 and has thereby generating two sets of tracks that can be compared with each other and the Hurdat tracks. We have taken all three sets of original tracks (Hurdat, ERA, and System 4 tracks) and processed them using the LF+ method in order to make them as comparable as possible. The tracks we discuss in this appendix are the processed tracks and all the statistics that we discuss are based on these processed tracks. As an example, figure 2 in Section 2.5 shows the ERA, the System 4, the Hurdat tracks for the hurricane season 2003.

First we may compare the Hurdat and the ERA tracks and their degree of overlap (the overlapping points[26] are marked with + in the figure). On average (over the full 1989-2009 period) the ERA tracks captures 57 percent of the Hurdat track points (over land and water) and 61 percent of the Hurdat track points that correspond to landfall. Thus the tracker applied to ERA correctly identifies about 60 percent of the cyclone track points, but, conversely, it misses about 40 percent of the points. In addition, the tracker, when applied to the ERA fields, sometimes identifies a cyclone before Hurdat and sometimes misclassifies non-tropical systems as tropical yielding extra ERA track points that not present in Hurdat. The number of points in the ERA tracks not present in the Hurdat tracks divided by the total number of Hurdat points is about 20 percent, for landfall points the ratio is about 16 percent.

Second, using the 2003 season as an example, we may compare the System 4 tracks to those recorded in Hurdat (again see figure 2). We may note that (a) System 4 landfall is less common than Hurdat landfall, (b) System 4 detects some depressions or cyclones moving off Africa before Hurdat, and (c) System 4 shows one storm system forming over the U.S. and four forming outside the U.S. east coast, whereas

[26] A track point in Hurdat is considered to overlap with a point in ERA provided that they are not more than 300 km apart at a given point in time.

Hurdat shows zero such systems. Although, 2003 is just one of the 34 hurricane seasons studied, the observations discussed here are typical and impact the forecasts and statistics.

B.2 Distributions

B.3 Temporal distribution of activity

In order to compare and assess the quality of the System 4 climatology we have estimated the temporal distribution of storm days over the season for Hurdat, ERA, and System 4, using data for the period 1989-2009. The 95-percent, two-sided confidence intervals for the distributions were also estimated using bootstrap across seasons and ensemble members. Figure 9 shows the distributions and confidence intervals.

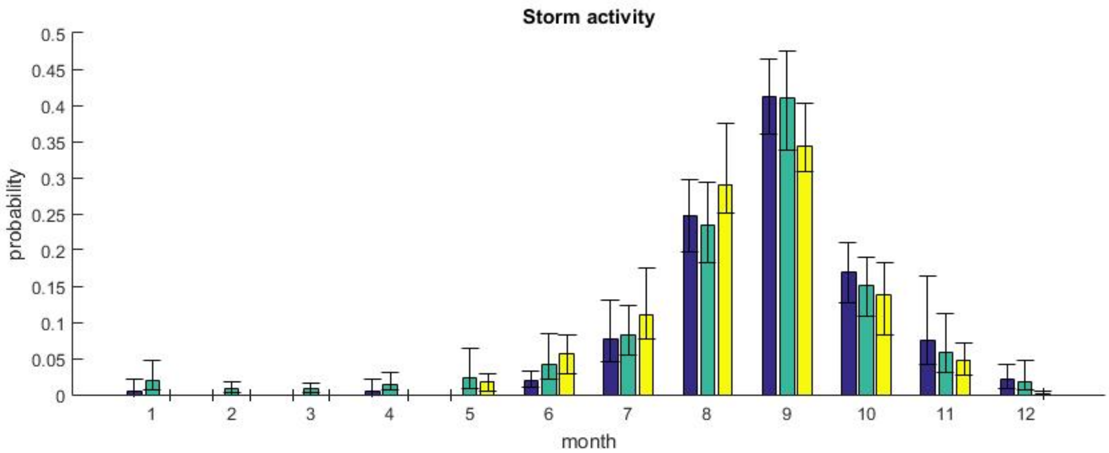

Figure 9: The intraseasonal distribution of storm activity, measured as storm days over sea for Hurdat (blue), ERA(green), and the model (yellow). Note that the model distribution is non-zero only for the period May through November (the forecast period); therefore, for ease of comparison, the probability distributions have all been normalized to unity using only the May-November period.

Overall, the model is in reasonable agreement with Hurdat and ERA, and the agreement between Hurdat and ERA is even better. For example, the peak month is September followed by August, October, and June in all three cases. One difference that one may note is that the model shows more activity than Hurdat during May-August and less during September-November. This may partly be due to the formation of a number of unphysical low pressure systems just north of South America early in the year in the model as discussed by Manganello et al. (2016). The fact that the ERA distribution shows more tropical storms between January and May, indicates that the tracker incorrectly misclassifies some non-tropical weather systems as tropical. Since the same tracker has been applied to the model, it is not unlikely that we also in this case pick up some non-tropical systems.

B.4 Spatial distribution

In order understand the quality of the model's cyclone tracks and how that may impact the forecasts we have looked at spatial distributions of certain types of points along the tracks. The points we have considered are: genesis location, activity, termination location, west max location, and peak wind location. These are defined as follows: For each track the genesis location is the first point over water with storm strength wind speed[27], the termination location is the last point on the track, the west max location is the most westerly point on the track over water, the peak wind location is the point with the highest wind speed, and the activity distribution has been calculated using all points over water along the track between the genesis and the termination point. Figure 10 shows these points and figure 11 shows the corresponding one-dimensional distributions.

As can be seen in figure 10 and 11. The model has relatively more genesis events close to the African cost and close to the U.S. east coast north of Florida and fewer in the middle of the Atlantic than historically observed. Also, the model has fewer genesis events in the Gulf of Mexico. The larger number of systems close to the African coast can, at least partly, be an artifact created by the tracker as systems are detected early. Similarly, the spike in genesis events at about 290 degrees east is likely to in part be early detection and misclassification by the tracker of weather systems moving east, off the North-American continent.

One may also note that the model has more termination events in the middle of the Atlantic than does Hurdat and ERA. Thus, of the storms forming in the eastern parts of the MDR fewer makes it across the Atlantic. As can be seen in figure 11 it is relatively more common for the model tracks to terminate at lower latitudes. One may also note that Hurdat has a higher number of tracks where the west max point occurs at about 30 degrees North.

The peak wind distribution of the model (see figure 11) has a relatively wide spread, one possible cause of this is the model's limited ability to assign a reliable wind speed to the cyclones, compare Section 2.3.

Over all, the model distributions resemble the Hurdat distributions with several notable exceptions, some of which are likely to impact the model's ability to forecast landfall as discussed in the main text.

[27] That is a wind speed in excess of 34 knots for Hurdat and 17 knots for ERA and the model

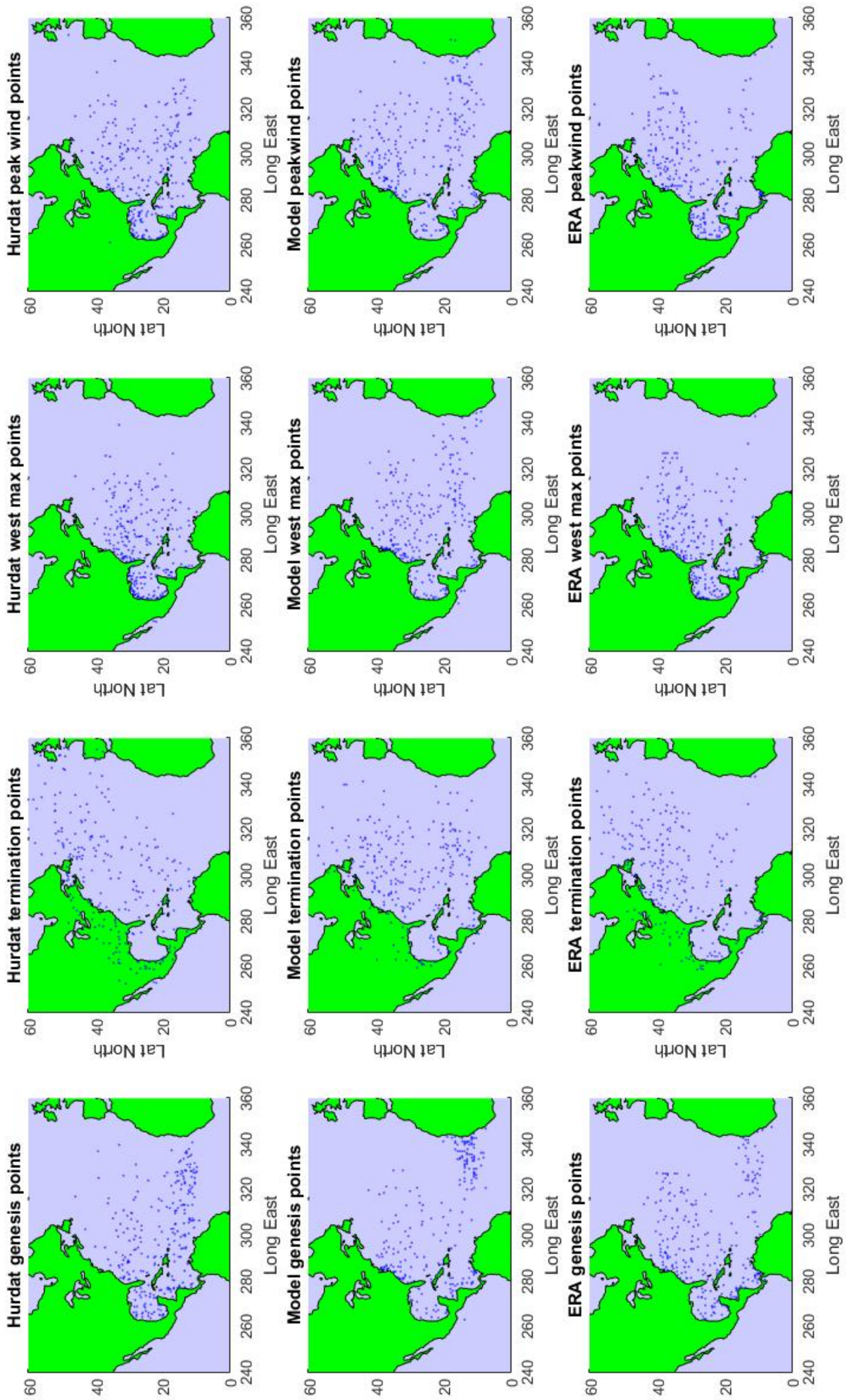

Figure 10: The latitude-longitude distributions of genesis, termination, west-max and peak-wind points for Hurdat (top row), the model (middle row), and ERA (bottom row). Distributions based on data from the period 1989-2009. All track points during the period have been used for Hurdat and ERA. For the model, a number of points equal to the number of Hurdat points have been selected randomly.

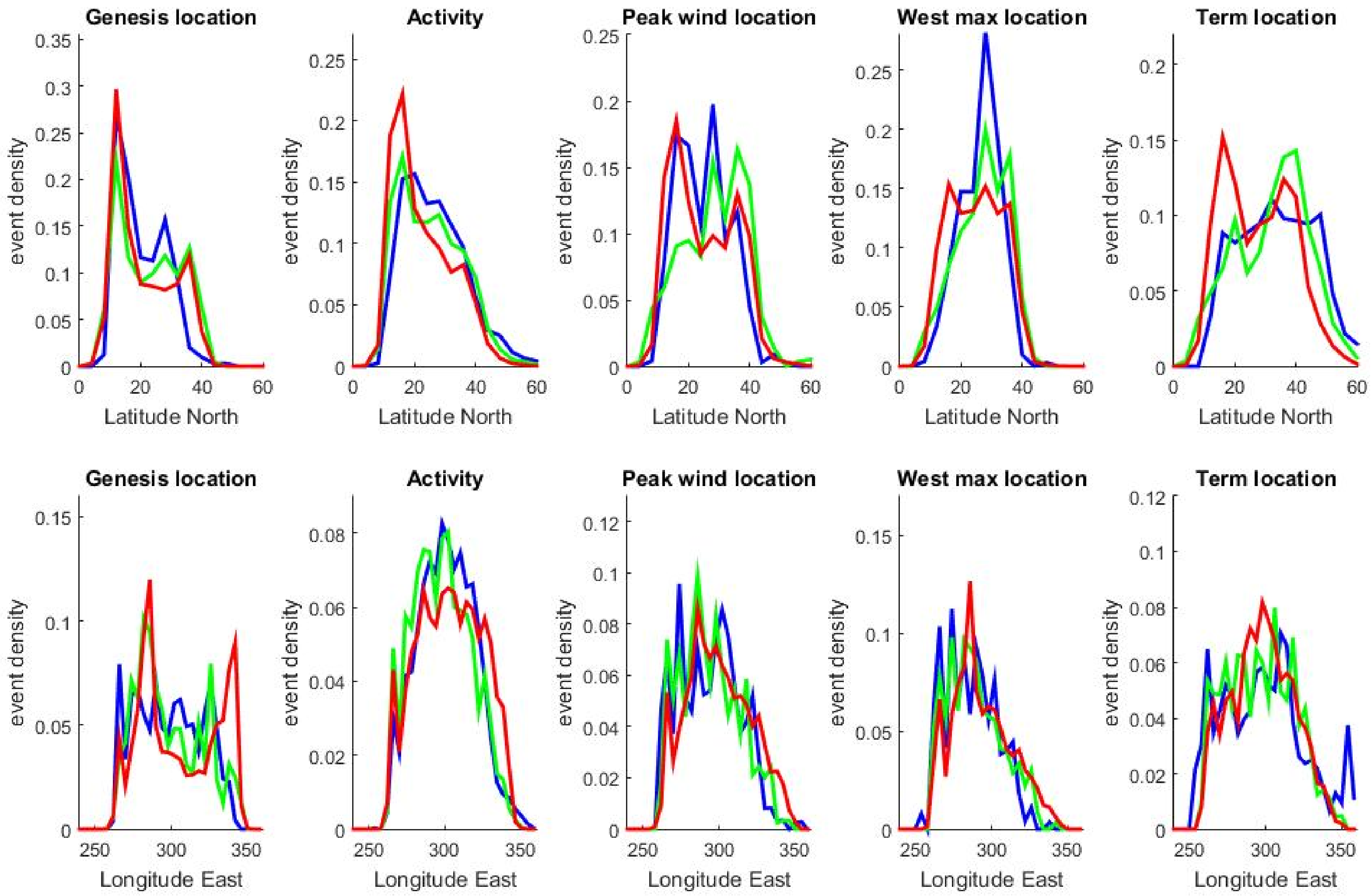

Figure 11: Meridional and latitudinal distributions of genesis location, activity, termination location, west max location and peak wind location for Hurdat (blue), ERA (green), and the model (red). The distributions are based on data from the period 1989-2009.

B.5 Number, activity, and landfall

In order to design and assess our forecast model we need estimates of the average number and variability of storms and hurricanes as well as their activity and landfall rates. This is discussed in Section 2.1 in the main text. Here we merely provide supplementary data (see table 16) and compare model and Hurdat statistics for the 1981-2014 period.

By focusing on the 1981-2014 period, that is the full time period for which reforecasts were made, we may compare the model to the Hurdat statistics for each year. Figure 12 shows the evolution of the storm and hurricane number, the activity, and the number of landfalls; and table 11 the corresponding averages and standard deviations for this period.

First, one may note that the average number of systems (storms and hurricanes) as well as the variability (standard deviation) are in reasonable agreement. Second, one may note two significant differences (a) the average numbers of landfalling systems are substantially lower both for the entire North American coast and the U.S. part, and (b) despite that the average number of systems are higher for the model, the average activity is lower.

The low average activity is a consequence of a shorter average life time of the model systems, which in its turn is likely to, at least in part, be the result of both too many termination events in the mid-Atlantic and the inclusion (misdetection by the tracker) of short-lived non-tropical systems among the tropical.

For the low number of model landfalls, a similar set of factors are among the likely causes: one being model's lack of genesis events in the Gulf, another being the model's high number of termination events in the Mid-Atlantic, and a third being the misclassification of a number non-tropical systems as tropical by the model (provided that these systems are less likely to make landfall than tropical systems). These factors have been discussed in Section B.4, see for example figure 10 for the distributions of termination and genesis points and figure 2 for examples of tracks.

For these reasons, it is likely that the mean number of "real" cyclones that form per year in the model climatology is lower than in reality as captured by Hurdat; and that too many model cyclones terminate before they reach the U.S. coast. It is likely that this will lower the skill of the forecasts.

C Appendix C - Supplementary data

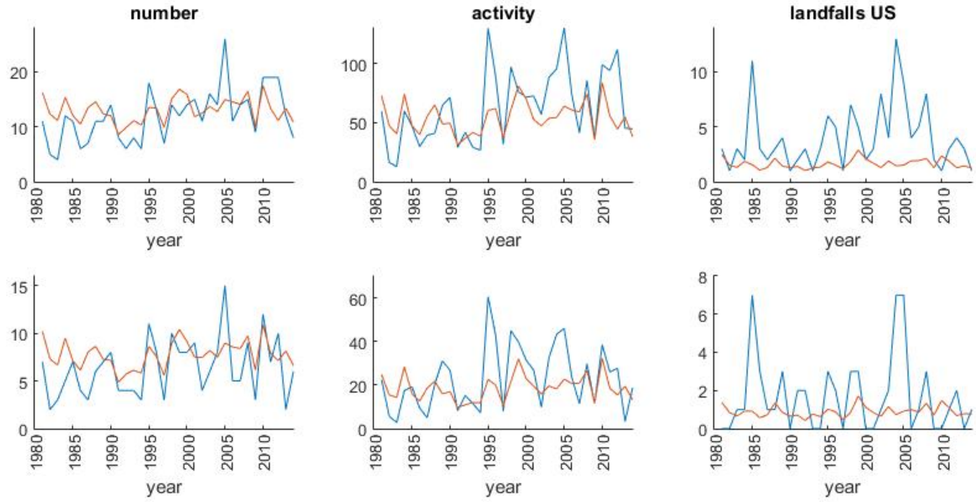

Figure 12: Model (orange) and Hurdat (blue) storm number, activity, and U.S. landfall (top row) for the period 1981-2014. The bottom row shows same variables for Hurricanes. Note that the model time series correspond the ensemble means. Note that this raw model data does not match the forecasts.

Table 11: Storm and hurricane number, activity, N.A. landfall, and U.S. landfall. Period 1981-2014. Note that the standard deviation has been estimated across years for Hurdat, and across years and ensemble members for the Model. Note that this raw model data does not match the forecasts, see the text.

STORM	mean		standard deviation	
	Model	Hurdat	Model	Hurdat
number	13.0	11.9	4.4	4.8
activity	54.0	62.9	23.2	31.2
N.A. landfall	3.6	7.6	2.5	4.6
U.S. landfall	1.6	4.0	1.5	2.9
HURRICANE	mean		std	
	Model	Hurdat	Model	Hurdat
number	7.8	6.4	3.2	3.1
activity	18.6	22.9	11.3	14.6
N.A. landfall	1.7	3.1	1.7	2.6
U.S. landfall	0.9	1.7	1.1	2.0

Table 12: Statistics on cyclone number, activity, N.A. landfall, and U.S. landfall for Hurdat, ERA and the model for different time periods. Mean, denotes the period average, std, the standard deviation, SEM, the standard deviation of the mean, and lower and upper the bounds of 95-percent confidence intervals.

STORM data set	period	number mean	number std	number SEM	mean lower	mean upper	std lower	std upper
Hurdat	1900-2014	9.9	4.3	0.4	9.2	10.8	3.8	5.1
	1950-2014	10.8	4.2	0.5	9.9	12.0	3.6	5.6
	1900-1980	9.1	3.7	0.4	8.3	9.9	3.3	4.4
	1981-2014	11.9	4.8	0.8	10.4	13.7	3.9	6.7
	1989-2009	12.5	4.6	1.0	10.8	14.7	3.2	7.2
ERA	1989-2009	16.5	6.3	1.4	14.3	19.6	4.4	8.5
Model	1981-2014	13.0	4.4	0.7	12.8	13.2	4.2	4.5
		activity						
Hurdat	1900-2014	52	27	2	48	57	23	31
	1950-2014	56	27	3	50	64	24	33
	1900-1980	48	23	3	43	53	20	28
	1981-2014	63	31	5	53	74	26	39
	1989-2009	68	31	7	56	82	24	40
ERA	1989-2009	68	34	7	56	85	25	45
Model	1981-2014	54	23	4	53	55	22	24
		N.A. landfall						
Hurdat	1900-2014	7.4	4.2	0.4	6.7	8.2	3.7	5.2
	1950-2014	7.1	4.1	0.5	6.2	8.1	3.5	5.1
	1900-1980	7.3	4.1	0.5	6.5	8.3	3.4	5.4
	1981-2014	7.6	4.6	0.8	6.1	9.3	3.9	6.1
	1989-2009	8.4	4.6	1.0	6.7	10.5	3.6	6.6
ERA	1989-2009	7.5	4.2	0.9	5.9	9.4	3.2	5.4
Model	1981-2014	3.6	2.5	0.4	3.4	3.7	2.4	2.6
		U.S. landfall						
Hurdat	1900-2014	4.0	2.4	0.2	3.6	4.5	2.1	2.9
	1950-2014	3.7	2.6	0.3	3.2	4.4	2.1	3.4
	1900-1980	4.0	2.2	0.2	3.5	4.5	1.9	2.7
	1981-2014	4.0	2.9	0.5	3.2	5.2	2.2	4.1
	1989-2009	4.6	3.1	0.7	3.5	6.1	2.2	4.6
ERA	1989-2009	3.9	2.5	0.6	2.9	5.0	2.1	3.1
Model	1981-2014	1.6	1.5	0.3	1.6	1.7	1.4	1.6

HURRICANE data set	period	number mean	number std	number SEM	mean lower	mean upper	std lower	std upper
Hurdat	1900-2014	5.5	2.7	0.2	5.0	6.0	2.3	3.2
	1950-2014	6.1	2.7	0.3	5.5	6.8	2.2	3.4
	1900-1980	5.1	2.4	0.3	4.7	5.7	2.0	2.9
	1981-2014	6.4	3.1	0.5	5.4	7.5	2.5	4.1
	1989-2009	6.8	3.1	0.7	5.6	8.3	2.4	4.6
ERA	1989-2009	12.3	5.1	1.1	10.5	14.9	3.5	6.9
Model	1981-2014	7.8	3.2	0.5	7.6	7.9	3.1	3.3
		activity						
Hurdat	1900-2014	21	14	1	19	23	12	16
	1950-2014	23	14	2	20	27	12	16
	1900-1980	20	13	1	17	23	11	16
	1981-2014	23	15	3	18	28	12	19
	1989-2009	27	16	3	21	34	13	20
ERA	1989-2009	28	15	3	22	35	12	20
Model	1981-2014	19	11	2	18	19	11	12
		N.A. landfall						
Hurdat	1900-2014	3.3	2.4	0.2	2.9	3.8	2.0	3.0
	1950-2014	3.2	2.3	0.3	2.7	3.8	1.9	3.1
	1900-1980	3.4	2.3	0.3	3.0	4.0	1.9	3.1
	1981-2014	3.1	2.6	0.4	2.4	4.1	2.0	3.9
	1989-2009	3.5	2.7	0.6	2.6	5.0	1.8	4.5
ERA	1989-2009	4.6	3.2	0.7	3.5	6.1	2.3	4.3
Model	1981-2014	1.7	1.7	0.3	1.6	1.8	1.6	1.8
		U.S. landfall						
Hurdat	1900-2014	2.0	1.7	0.2	1.7	2.3	1.5	2.0
	1950-2014	1.7	1.7	0.2	1.3	2.2	1.4	2.2
	1900-1980	2.1	1.5	0.2	1.8	2.4	1.4	1.8
	1981-2014	1.7	2.0	0.3	1.1	2.5	1.4	2.7
	1989-2009	1.9	2.1	0.5	1.1	2.9	1.3	3.0
ERA	1989-2009	2.2	2.2	0.5	1.4	3.3	1.7	2.8
Model	1981-2014	0.9	1.1	0.2	0.8	0.9	1.0	1.2

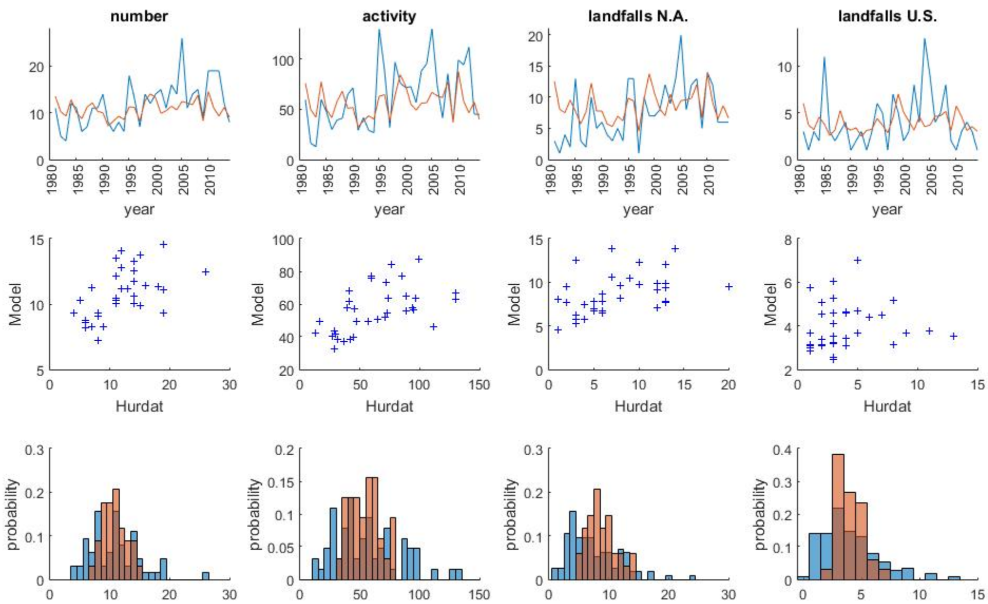

Figure 13: Forecast and observation for: storm *number*, *activity* (measured as storm days), *landfalls N.A.* and *landfalls U.S.* for the period 1981-2014. Top row, time series of forecasts (orange) and observations (blue); middle row, forecast as a function of observation; and, bottom row, probability distributions for the forecasts(orange) and observation (blue). In the case of landfall, the probability distributions for observations have been estimated using the full 1900-2014 period, in the case of number and activity, using the period 1950-2014. All bars in the bottom row of the figure start from the horizontal axis, when bars overlap, the color is dark orange.

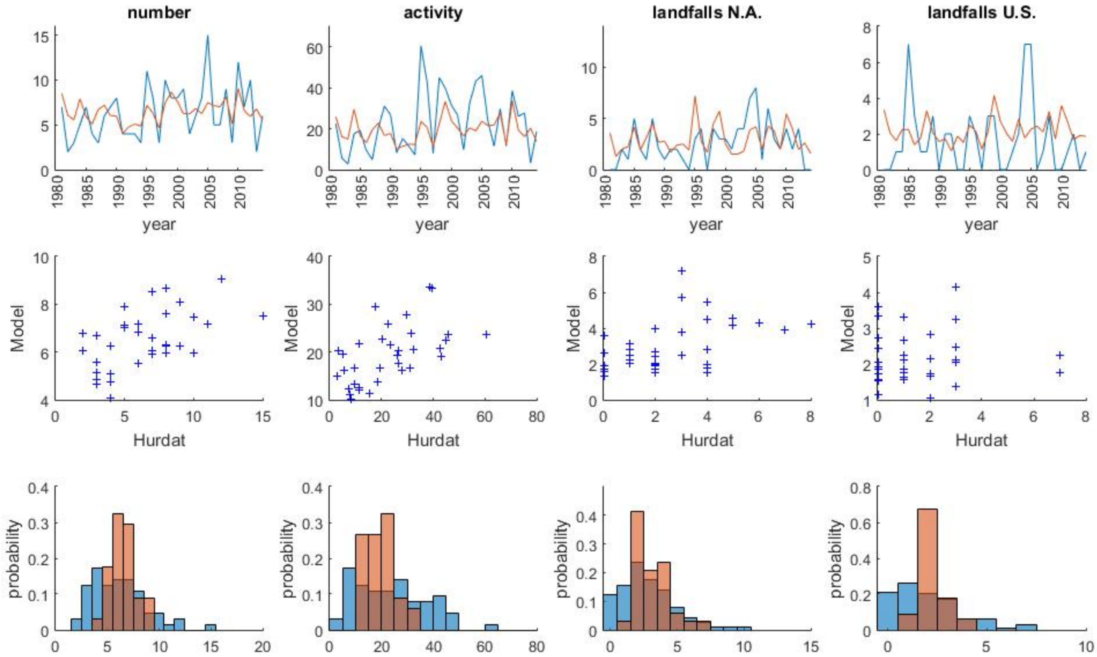

Figure 14: Forecast and observation for: hurricane *number*, *activity* (measured as hurricane days), *landfalls N.A.* and *landfalls U.S.* for the period 1981-2014. Top row, time series of forecasts (orange) and observations (blue); middle row, forecast as a function of observation; and, bottom row, probability distributions for the forecasts(orange) and observation (blue). In the case of landfall, the probability distributions for observations have been estimated using the full 1900-2014 period, in the case of number and activity, using the period 1950-2014. All bars in the bottom row of the figure start from the horizontal axis, when bars overlap, the color is dark orange.

Table 13: Comparison of the S4LF+ forecasts when scored against the observed outcome according to Hurdat (*observed*) and when scored against the different ensemble members (*ensemble*) for the period 1981-2014. The *forecasted* and the *observed* means are shown. For the *difference* between forecast and observation, the mean error (that is the systematic error) and the RMSE are shown. Also Spearman rank correlations r and associated P-value are shown. Lower and upper bounds on bootstrapped confidence intervals for r, the mean, and the RMSE are also shown. Other variables and metrics as described in the main text.

		mean		systematic error			RMSE			Spearman rank correlation			
		forecasted	observed	est	lower	upper	est	lower	upper	est	P-value	lower	upper
STORM													
number	observed	10.8	11.9	-1.1	-2.7	0.1	4.1	3.2	5.8	0.57	0.0004	0.3	0.8
	ensemble	10.8	10.8	0.0	-0.1	0.1	3.1	3.0	3.3	0.51	0.0000	0.5	0.5
activity	observed	56.4	62.9	-6.5	-16.2	1.3	26.4	20.9	33.6	0.61	0.0001	0.3	0.8
	ensemble	56.4	56.4	0.0	-0.9	0.9	19.7	19.0	20.4	0.58	0.0000	0.5	0.6
N.A. landfall	observed	8.7	7.6	1.1	-0.4	2.3	4.1	3.3	5.6	0.53	0.001	0.2	0.8
	ensemble	8.7	8.7	0.0	-0.3	0.3	5.7	5.5	6.0	0.36	0.0000	0.3	0.4
U.S. landfall	observed	4.0	4.0	0.0	-1.2	0.8	3.0	2.2	4.3	0.29	0.09	-0.1	0.6
	ensemble	4.0	4.0	0.0	-0.2	0.2	3.5	3.3	3.6	0.25	0.0000	0.2	0.3
HURRICANE													
number	observed	6.5	6.4	0.1	-0.8	0.9	2.6	2.1	3.6	0.55	0.0007	0.3	0.7
	ensemble	6.5	6.5	0.0	-0.1	0.1	2.4	2.3	2.5	0.45	0.0000	0.4	0.5
activity	observed	19.4	22.9	-3.5	-7.8	0.2	12.3	9.8	16.3	0.61	0.0001	0.4	0.8
	ensemble	19.4	19.4	0.0	-0.5	0.5	10.1	9.7	10.6	0.50	0.0000	0.5	0.5
N.A. landfall	observed	4.2	3.1	1.1	0.1	1.9	2.6	2.0	3.9	0.29	0.10	-0.1	0.6
	ensemble	4.2	4.2	0.0	-0.2	0.2	3.9	3.8	4.2	0.27	0.0000	0.2	0.3
U.S. landfall	observed	2.2	1.7	0.5	-0.3	1.1	2.1	1.5	2.8	0.11	0.5	-0.2	0.4
	ensemble	2.2	2.2	0.0	-0.1	0.1	2.6	2.5	2.7	0.22	0.0000	0.2	0.3

Table 14: Spearman rank correlations between different storm and hurricane variables. In the orange box Hurdat-Hurdat correlations, in the yellow model-model correlations, and in the green model-Hurdat cross correlations.

			HURDAT								S4LF+							
			Storm				**Hurricane**				**Storm**				**Hurricane**			
			Number	Activity	N.A. Landfall	U.S. Landfall	Number	Activity	N.A. Landfall	U.S. Landfall	Number	Activity	N.A. Landfall	U.S. Landfall	Number	Activity	N.A. Landfall	U.S. Landfall
HURDAT	**Storm**	Number	1.0	0.9	0.7	0.4	0.8	0.8	0.6	0.2	0.6	0.6	0.6	0.5	0.6	0.6	0.6	0.6
		Activity	0.9	1.0	0.7	0.5	0.9	0.9	0.6	0.4	0.6	0.6	0.6	0.5	0.6	0.6	0.5	0.5
		N.A. Landfall	0.7	0.7	1.0	0.6	0.7	0.7	0.8	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
		U.S. Landfall	0.4	0.5	0.6	1.0	0.4	0.5	0.7	0.7	0.3	0.3	0.3	0.3	0.3	0.3	0.2	0.2
	Hurricane	Number	0.8	0.9	0.7	0.4	1.0	0.9	0.6	0.4	0.5	0.6	0.5	0.5	0.6	0.6	0.5	0.5
		Activity	0.8	0.9	0.7	0.5	0.9	1.0	0.7	0.5	0.5	0.6	0.6	0.5	0.6	0.6	0.5	0.6
		N.A. Landfall	0.6	0.6	0.8	0.7	0.6	0.7	1.0	0.7	0.3	0.3	0.4	0.3	0.3	0.4	0.3	0.3
		U.S. Landfall	0.2	0.4	0.4	0.7	0.4	0.5	0.7	1.0	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.1
S4LF+	**Storm**	Number									1.0	1.0	0.9	0.9	1.0	1.0	0.9	0.9
		Activity									1.0	1.0	0.9	0.9	1.0	1.0	0.9	0.8
		N.A. Landfall									0.9	0.9	1.0	0.9	0.9	0.9	0.9	0.9
		U.S. Landfall									0.9	0.9	0.9	1.0	0.9	0.9	0.9	0.9
	Hurricane	Number									1.0	1.0	0.9	0.9	1.0	1.0	0.9	0.9
		Activity									1.0	1.0	0.9	0.9	1.0	1.0	0.9	0.9
		N.A. Landfall									0.9	0.9	0.9	0.9	0.9	0.9	1.0	0.9
		U.S. Landfall									0.9	0.8	0.9	0.9	0.9	0.9	0.9	1.0

Table 15: Forecasts of the number of storm and hurricane landfalls for the period 1981-2014 along different segments of the North American coast. The start and end points of the segments are shown in figure 5.

Segment points	Segment name	forecasted	observed	difference best	difference lower	difference upper	RMSE best	RMSE lower	RMSE upper	Spearman r	Spearman P-value	Spearman lower	Spearman upper
STORM													
a-h	North America (N.A.), 10<Lat<52	8.7	7.6	1.1	-0.4	2.3	4.1	3.2	5.6	0.53	0.001	0.21	0.74
c-g	U.S.	4.0	4.0	0.0	-1.2	0.9	3.0	2.2	4.3	0.27	0.13	-0.09	0.54
a-c	Lat 10 up to U.S.	3.0	2.0	1.1	0.3	1.6	2.0	1.5	2.7	0.48	0.004	0.15	0.72
b-e	Gulf to Florida tip	3.6	3.7	-0.1	-1.1	0.7	2.7	2.0	4.2	0.45	0.01	0.10	0.71
c-e	U.S. part of Gulf to Florida tip	1.9	2.4	-0.4	-1.1	0.2	2.0	1.6	2.4	0.24	0.17	-0.14	0.54
e-g	Florida tip to Canada	2.1	1.6	0.4	-0.2	0.9	1.6	1.2	2.4	0.11	0.53	-0.18	0.42
g-h	Canada to lat 52	1.6	1.6	0.0	-0.5	0.4	1.4	1.2	1.8	0.31	0.07	-0.02	0.58
d-f	Florida	1.3	1.4	-0.1	-0.7	0.3	1.4	1.0	1.8	0.29	0.10	-0.07	0.55
c-d	U.S. part of Gulf except Florida	1.2	1.4	-0.2	-0.7	0.3	1.6	1.3	2.0	-0.18	0.30	-0.50	0.16
f-g	U.S. East coast north of Florida	1.5	1.2	0.3	-0.2	0.7	1.3	0.9	1.8	0.21	0.23	-0.08	0.48
Miami	Miami, Lat<27, 79<Lon<81	0.3	0.2	0.1	0.0	0.2	0.4	0.3	0.5	0.31	0.08	-0.11	0.62
NY	NY, 40<Lat<42, 72<Lon<75	0.1	0.1	0.1	-0.1	0.1	0.3	0.2	0.4	0.09	0.63	-0.09	0.29
HURRICANE													
a-h	North America (N.A.), 10<Lat<52	4.2	3.1	1.1	0.1	1.9	2.6	1.9	3.9	0.29	0.10	-0.07	0.56
c-g	U.S.	2.2	1.7	0.5	-0.3	1.1	2.1	1.5	2.9	0.10	0.56	-0.24	0.42
a-c	Lat 10 up to U.S.	0.6	0.9	-0.3	-0.8	0.1	1.3	0.9	1.7	0.36	0.03	0.02	0.64
b-e	Gulf to Florida tip	1.4	1.6	-0.2	-1.0	0.3	1.8	1.1	3.1	0.34	0.05	0.01	0.60
c-e	U.S. part of Gulf to Florida tip	0.9	0.9	0.0	-0.5	0.3	1.3	0.9	1.9	0.15	0.38	-0.17	0.46
e-g	Florida tip to Canada	1.3	0.8	0.6	0.0	0.9	1.2	0.8	2.0	0.04	0.82	-0.29	0.38
g-h	Canada to lat 52	1.4	0.5	0.8	0.6	1.0	0.7	0.5	0.9	0.36	0.04	-0.01	0.63
d-f	Florida	0.6	0.4	0.2	-0.2	0.4	0.9	0.6	1.3	0.14	0.44	-0.22	0.45
c-d	U.S. part of Gulf except Florida	0.5	0.6	-0.1	-0.5	0.2	0.9	0.7	1.2	0.17	0.34	-0.17	0.47
f-g	U.S. East coast north of Florida	1.1	0.6	0.4	0.1	0.7	0.9	0.7	1.3	0.10	0.57	-0.26	0.42
Miami	Miami, Lat<27, 79<Lon<81	0.1	0.1	0.1	-0.1	0.1	0.3	0.1	0.4	-0.09	0.61	-0.30	0.02
NY	NY, 40<Lat<42, 72<Lon<75	0.1	0.0	0.1	0.1	0.1	0.1	0.1	0.1	-	-	-	-

Table 16: Comparison of rolling 10-year average climatological forecasts (clim10y) to the S4LF+ forecasts for the time period 1981-2014. The *forecasted* and the *observed* means are shown. For the systematic (mean) error, the RMSE, and the rank correlation *r*; the best estimates *est* and *lower* and *upper* bounds on the corresponding 95-percent confidence intervals are shown. For the rank correlation the associated P-value is also shown. Other variables and metrics as described in the main text.

		mean		systematic error			RMSE			Spearman rank correlation			
		forecasted	observed	est	lower	upper	est	lower	upper	est	P-value	lower	upper
STORM													
number	S4LF+	10.8	11.9	-1.1	-2.7	0.1	4.1	3.2	5.8	0.57	0.0004	0.27	0.75
	clim10y	11.0	11.9	-0.9	-2.4	0.5	4.3	3.4	5.8	0.51	0.002	0.19	0.74
activity	S4LF+	56	63	-6	-16	1	26	21	34	0.61	0.0001	0.33	0.76
	clim10y	58	63	-5	-15	4	29	23	39	0.43	0.01	0.13	0.67
N.A. landfall	S4LF+	8.7	7.6176	1.1	-0.4	2.3	4.1	3.3	5.6	0.53	0.001	0.21	0.75
	clim10y	7.0	7.6	-0.6	-2.2	0.7	4.3	3.5	5.4	0.4	0.01	0.1	0.68
U.S. landfall	S4LF+	4.0	4.0	0.0	-1.2	0.8	3.0	2.2	4.3	0.29	0.09	-0.1	0.6
	clim10y	3.9	4.0	-0.1	-1.3	0.8	3.1	2.3	4.3	0.11	0.55	-0.27	0.44
HURRICANE													
number	S4LF+	6.5	6.4	0.1	-0.8	0.9	2.6	2.1	3.6	0.55	0.0007	0.27	0.74
	clim10y	6.2	6.4	-0.2	-1.2	0.8	3.0	2.5	3.9	0.15	0.4	-0.25	0.47
activity	S4LF+	19	23	-3	-8	0	12	10	16	0.61	0.0001	0.39	0.76
	clim10y	23	23	0	-5	4	14	11	21	0.27	0.13	-0.11	0.56
N.A. landfall	S4LF+	4.2	3.1	1.1	0.1	1.9	2.6	2.0	3.9	0.29	0.1	-0.1	0.6
	clim10y	3.1	3.1	0.0	-1.0	0.8	2.7	2.1	3.8	0.0	0.9	-0.4	0.32
U.S. landfall	S4LF+	2.2	1.7	0.5	-0.3	1.1	2.1	1.5	2.8	0.11	0.55	-0.24	0.43
	clim10y	1.7	1.7	0.1	-0.8	0.6	2.1	1.5	2.9	-0.14	0.42	-0.46	0.20

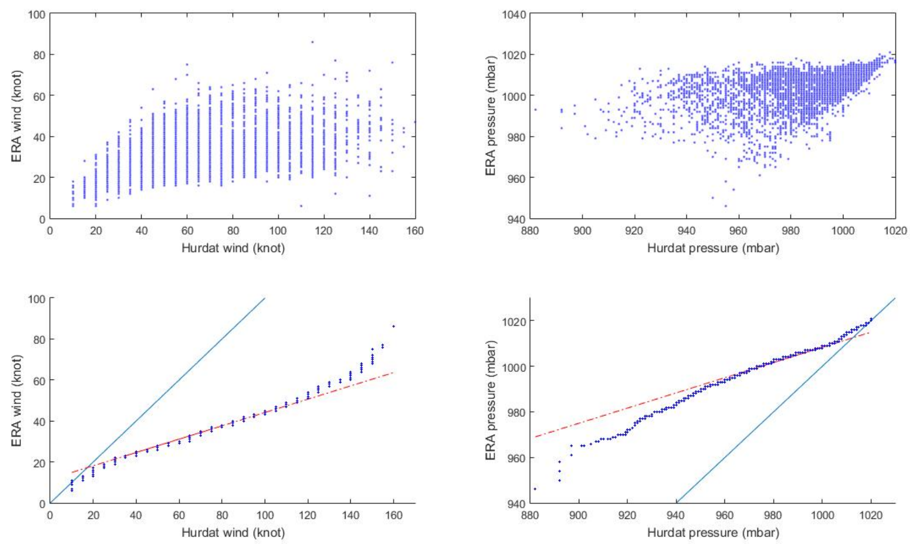

Figure 15: The wind (top left) and pressure (top right) relations between Hurdat and ERA along the tracks; and the Hurdat-ERA Q-Q plots for the wind (bottom left) and pressure (bottom right). Red lines fitted between the first and third quantile and extrapolated. Blue lines show how a one-one relation would look.